Sensitive Data Donation
Reframing Intimate Data Practices through Design

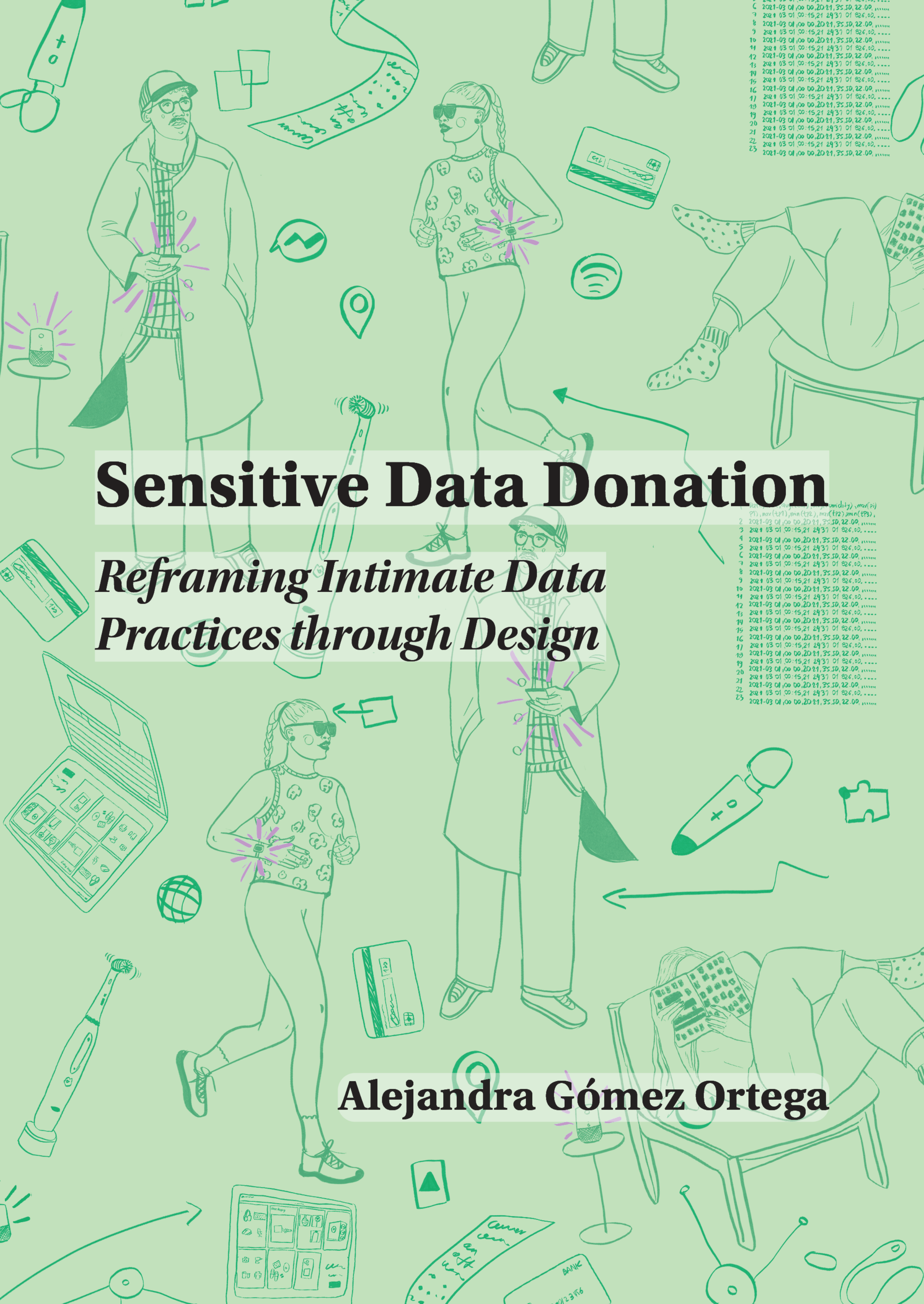
Gomez Ortega, A.

DOI
10.4233/uuid:94933b7f-9106-498d-aaac-ff6f0a41e46c

Publication date
2024

Document Version
Final published version

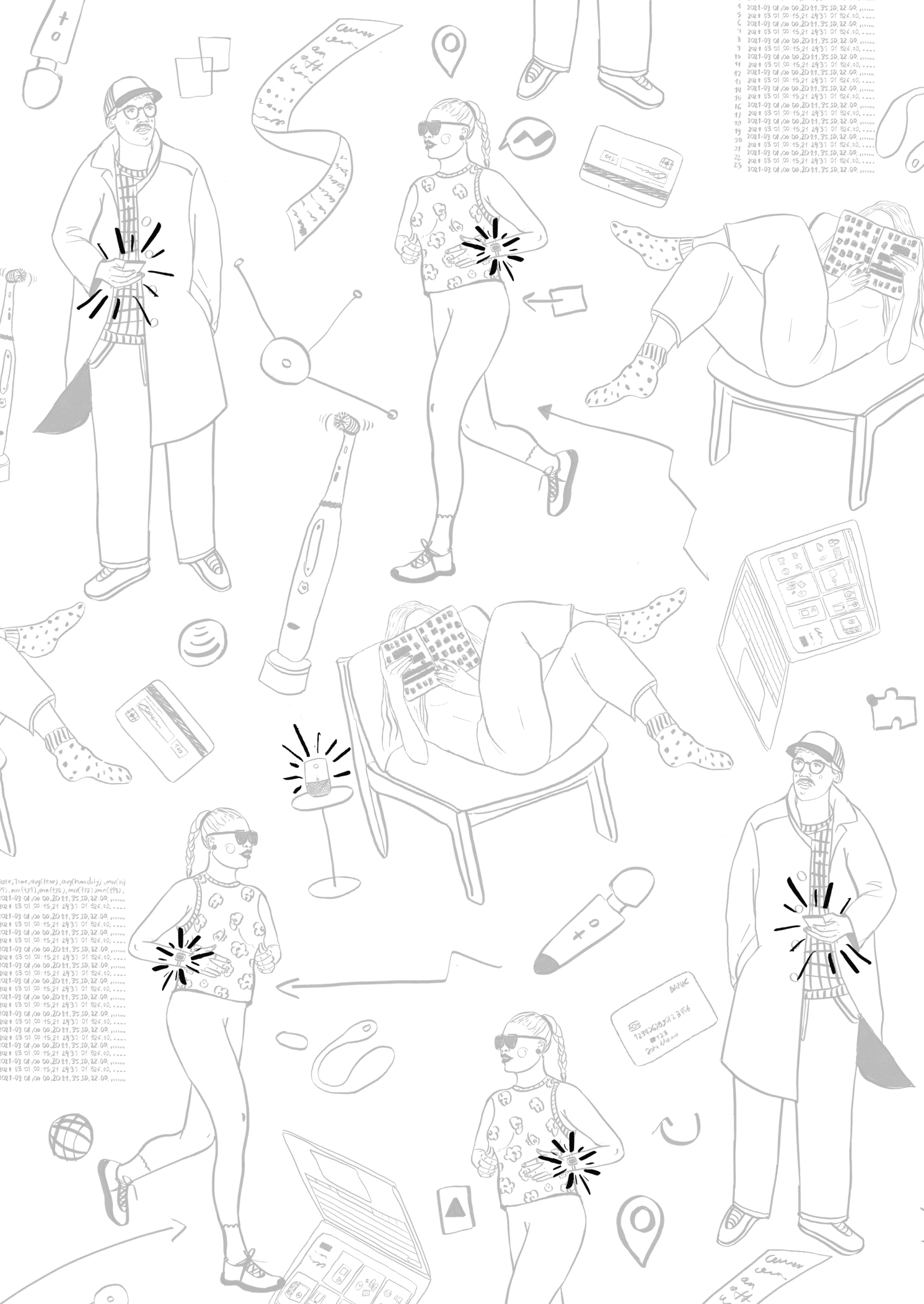
Citation (APA)
Gomez Ortega, A. (2024). *Sensitive Data Donation: Reframing Intimate Data Practices through Design*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:94933b7f-9106-498d-aaac-ff6f0a41e46c



Sensitive Data Donation

Reframing Intimate Data Practices through Design

Alejandra Gómez Ortega


BANK

Sensitive Data Donation

Reframing Intimate Data Practices through Design

Sensitive Data Donation

Reframing Intimate Data Practices through Design

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Monday, 8 July 2024 at 12:30 o'clock

by

Alejandra GÓMEZ ORTEGA

Master of Science in Biomedical Engineering, Politecnico di Milano, Italy
born in Medellín, Colombia

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. G.W. Kortuem,	Delft University of Technology, *promotor*
Dr. J. Bourgeois,	Delft University of Technology, *copromotor*
Independent members:	
Dr. A.G. Avila Bernal	Universidad de los Andes, Colombia
Prof.dr. M. Balaam	KTH Royal Institute of Technology, Sweden
Prof.dr. F.M.T. Brazier	Delft University of Technology, Netherlands
Prof.dr. J.A.C. Rietjens	Delft University of Technology, Netherlands
Prof.dr.ir. M.S. Kleinsmann	Delft University of Technology, Netherlands

Dr. J. Love of Delft University of Technology has contributed greatly to the preparation of this dissertation.

Keywords: Data Donation, Data Feminism, Data Governance, Personal Data, Sensitive Data, Participatory Design, Research through Design.

Cover by: Luisa Castellanos Ramirez

ISBN 978-94-6366-852-1

La teoría me la sé, pero cómo me la meto en el cuerpo.

Gabriela Wiener

Contents

Summary ix

Acknowledgements xv

Preface xxiii

1 Introduction 3
- 1.1 Context and Motivation . . . 4
- 1.2 Thesis Position in Research Landscape . . . 6
- 1.3 Research Aim . . . 9
- 1.4 Research Approach . . . 10
- 1.5 Research Contributions . . . 14
- 1.6 Thesis Structure . . . 15

2 Interrogating Recent Approaches to Data Donation 17
- 2.1 Introduction . . . 18
- 2.2 Conceptual Framework Analysis . . . 18
- 2.3 Critique: Data Donation through a Data Feminism Lens . . . 23
- 2.4 Sensitive Data Donation . . . 25
- 2.5 Chapter Takeaways . . . 30

3 Integrating Multiple Knowledge(s) into Data Donation 33
- 3.1 Introduction . . . 35
- 3.2 Intimate Context: Menstrual Tracking . . . 36
- 3.3 Data Donation Process . . . 36
- 3.4 Applying and Reflecting on Data Donationn . . . 41
- 3.5 Results . . . 42
- 3.6 Discussion . . . 47
- 3.7 Chapter Takeaways . . . 52

4 Incorporating Dynamic Consent into Data Donation 55
- 4.1 Introduction . . . 57
- 4.2 Intimate Context: Voice Assistants in the Home . . . 58
- 4.3 Data Donation Process . . . 59
- 4.4 Applying and Reflecting on Data Donation . . . 65
- 4.5 Results . . . 66
- 4.6 Discussion . . . 73
- 4.7 Chapter Takeaways . . . 77

5 Conceptualizing Sensitivity and Intimacy through Data Donation 79
5.1 Introduction 81
5.2 Intimate Context: Misalignments in the Home 82
5.3 Methodology 83
5.4 Results 88
5.5 Discussion 98
5.6 Chapter Takeaways 102

6 Interlude: Dataslip 105
6.1 Introduction 106
6.2 Creepy Personal Data 108
6.3 Futuring and Design Fiction 108
6.4 Method 109
6.5 Speculative Scenarios 113
6.6 Discussion 124
6.7 Notes 125
6.8 Chapter Takeaways 125

7 Promoting Different Degrees of Participation in Data Donation 129
7.1 Introduction 131
7.2 Intimate Context: Sports and the Menstrual Cycle 132
7.3 Data Donation Process 133
7.4 Applying and Reflecting on Data Donation 137
7.5 Results 138
7.6 Discusion 146
7.7 Chapter Takeaways 149

8 Discussion and Conclusion 151
8.1 Overview 152
8.2 Summary of Answers to Research Questions 152
8.3 Implications 158
8.4 Limitations 160
8.5 Future Challenges 160
8.6 Concluding Remarks 161

Epilogue: Unforeseen Challenges and Lessons Learned 163

Data and Code 171

References 173

List of Publications 189

Curriculum Vitæ 191

Summary

Motivation

Our routine interactions with connected products and services result in the collection and indefinite storage of personal digital-trace data. These data are increasingly entangled with our lives. What we experience is scattered around multiple data points: our browsing history can account for our interests and worries, our messaging logs can account for our social relationships, and our purchase history can account for our dietary preferences. Digital-trace data is increasingly valuable for scientific research as it can offer insights into specific aspects of our daily experiences. Researchers across various disciplines have been developing methods and tools to access these data. One of these, and the focus of this thesis, is **data donation**.

Objective

This thesis aims to develop an **alternative data donation approach informed by the principles of Data Feminism** that empowers donors to know and care for their data and promotes different forms of knowledge and participation. The alternative approach, called Sensitive Data Donation, is directed at intimate research contexts, such as the home and the body. These contexts are inherently private and situated spaces where it is crucial to explore and negotiate people's relationships with their data.

Outline and Results

Chapter 1 motivates the need for an alternative approach to data donation informed by the principles of Data Feminism. The chapter states the aim of the thesis and outlines the research questions and the methodological choices – describing the participatory Research through Design approach.

Chapter 2 describes the key concepts and the current practices in the data donation literature, synthesizes these into a conceptual framework, and introduces a critique through the principles of Data Feminism. The critique informs a first approximation of the Sensitive Data Donation principles and the Sensitive Data Donation approach.

Chapter 3 investigates how data donation can integrate multiple forms of knowledge other than digital-trace data, specifically people's embodied and situated knowledge. The chapter introduces the first iteration of an empirical data donation process. It focuses on value creation, data curation, and data interpretation. It is applied in a research project aiming to support people who track their menstrual cycle in interacting with the data they produce. The chapter concludes with a reflection on the actors involved in data donation and how their roles are influenced by data interpretation.

Chapter 4 investigates how data donation can foster an incremental understanding of data that invites donors to (re)assess their participation. The chapter introduces the second iteration of an empirical data donation process. It focuses on incrementally supporting donors in knowing their data before deciding to donate it and through their voluntary participation in subsequent activities. It is applied in a research project aiming to investigate people's perceptions of their speech records – data collected by connected voice assistants. The chapter concludes with a reflection on ongoing and dynamic consent practices in data donation.

Chapter 5 delves into the data and its perceived characteristics. The chapter investigates how donors perceive the 'sensitivity' and 'intimacy' of their data, specifically their Google Assistant speech records, when faced with a comprehensive overview. It proposes a conceptualization of 'sensitive' and 'intimate' data that accounts for their nuance and contextual properties.

Chapter **6** is an ***interlude***. The chapter introduces dataslip. It is a provocative artifact that materializes the personal data trail in the form of a receipt. It aims to raise awareness over personal data collection and elicit creepiness. Further, the chapter describes the use of dataslip to foster a conversation with community members around the challenges of personal data collection and envisioning alternative futures.

Chapter 7 investigates how data donation can promote different degrees of participation, specifically contributors, collaborators, and co-creators. The chapter introduces the third iteration of an empirical data donation process. It focuses on incorporating dedicated activities for each of the three degrees of participation. It is applied in a research project aiming to investigate athletes' perceptions of the impact of their menstrual cycle in sports through their physical activity data. The chapter concludes with a reflection on how different degrees of participation lead to various choices and impact both the data donation process and the data.

Chapter 8 presents a concluding overview of the research. The chapter discusses the findings, highlights the research contributions, discusses the implications, reflects on the limitations, and provides recommendations for future work.

Resumen

Motivación

Nuestras interacciones diarias con productos y servicios conectados resultan en la recopilacion y el almacenamiento indefinido de datos personales. Estos datos están cada vez más entrelazados con nosotros mismos. Lo que experimentamos hace parte de múltiples datos: nuestro historial de navegación da cuenta de nuestros intereses y preocupaciones, nuestro historical de mensajes da cuenta de nuestras relaciones sociales y nuestro historial de compras da cuenta de nuestras preferencias dietéticas. Los datos son cada vez más valiosos para la investigación científica, ya que pueden ofrecer información sobre aspectos específicos de nuestras experiencias diarias. Investigadoras de diversas disciplinas han estado desarrollando métodos y herramientas para acceder a estos datos. Uno de ellos, y el foco de esta tesis, es la **donación de datos**.

Objetivo

Esta tesis tiene como objetivo desarrollar un **enfoque alternativo de donación de datos informado por los principios del Feminismo de Datos** que invite a las donantes a conocer y cuidar sus datos y promueva diferentes formas de conocimiento y participación. El enfoque alternativo, llamado Donación de Datos Sensibles, está dirigido a contextos de investigación íntimos, como el hogar y el cuerpo. Estos contextos son espacios inherentemente privados y situados donde es crucial explorar y negociar las relaciones entre las personas y sus datos.

Estructura y Resultados

Capítulo 1 motiva la necesidad de un enfoque alternativo para la donación de datos informado por los principios del Feminismo de Datos. Establece el objetivo de la tesis y describe las preguntas de investigación y las decisiones metodológicas, profundizando en el enfoque de Investigación por medio del Diseño.

Capítulo 2 interroga los conceptos y prácticas de donación de datos a través de una revisión de la literatura. Estos se sintetizan en un marco conceptual de donación de datos y se critican a través de los principios del Feminismo de Datos. Finalmente, el capítulo consolida las ideas del marco conceptual y la crítica en los cinco principios de la Donación de Datos Sensibles. Estos principios se plasman en un enfoque de cinco fases que se explorará y aplicará en capítulos posteriores.

Capítulo 3 investiga cómo la donación de datos puede integrar múltiples formas de conocimiento, más allá de los datos, específicamente el conocimiento que viene de las personas. El capítulo presenta la primera iteración de un proceso empírico de donación de datos que se centra en la creación de valor y la interpretación de los datos. Se aplica en un proyecto de investigación que tiene como objetivo ayudar a las personas que realizan un seguimiento de su ciclo menstrual a interactuar con los datos que producen. El capítulo concluye con una reflexión sobre los actores involucrados en la donación de datos y cómo sus roles se ven influenciados por la interpretación de los datos.

Capítulo 4 investiga cómo la donación de datos puede fomentar una comprensión incremental de los datos que invite a las donantes a (re)evaluar su participación. El capítulo presenta la segunda iteración de un proceso empírico de donación de datos, centrada en apoyar progresivamente a las donantes para que conozcan sus datos antes de decidir donarlos. Se aplica en un proyecto que tiene como objetivo investigar las percepciones de las personas sobre sus registros de voz. El capítulo concluye con una reflexión sobre las prácticas de consentimiento dinámicas en la donación de datos.

Capítulo 5 profundiza en los datos y sus características percibidas. El capítulo investiga cómo las donantes perciben la “sensibilidad” y la “intimidad” de sus datos, específicamente sus registros de voz del Asistente de Google, cuando se enfrentan a ellos a profundidad. Propone una conceptualización de datos “sensibles” e “íntimos” que tiene en cuenta sus matices y propiedades contextuales.

Capítulo 6 es un breve ***interludio***. El capítulo presenta a dataslip. Se trata de un artefacto provocador que materializa el rastro de datos personales en un recibo. Su objetivo es crear conciencia sobre la recopilación de datos personales y provocar incomodidad. Además, el capítulo describe el uso de dataslip para fomentar una conversación con personas de todas las edades sobre los desafíos de la recopilación de datos personales y la visión de futuros alternativos.

Capítulo 7 investiga cómo la donación de datos puede promover diferentes grados de participación, específicamente contribuyentes, colaboradores y cocreadores. El capítulo presenta la tercera iteración de un proceso empírico de donación de datos, centrada en incorporar actividades dedicadas para cada uno de los tres grados de participación. Se aplica en un proyecto que tiene como objetivo investigar las percepciones de atletas sobre el impacto de su ciclo menstrual en su rendimiento deportivo. El capítulo concluye con una reflexión sobre cómo los diferentes grados de participación conducen a diversas opciones e impactan tanto en el proceso de donación de datos como en los datos.

Capítulo 8 presenta una visión general concluyente de la investigación. El capítulo analiza los hallazgos, destaca las contribuciones de la investigación, analiza las implicaciones, reflexiona sobre las limitaciones y proporciona recomendaciones para investigar en el futuro.

Samenvatting

Motivatie

Onze dagelijkse interactie met verbonden producten en diensten resulteert in de verzameling en eindeloze opslag van digitale sporen: persoonlijke data. Deze data raken meer en meer met onze levens vervlochten. Onze ervaringen raken verspreid over een variëteit aan datapunten: onze zoekgeschiedenis op het internet toont onze interesses en zorgen, berichten laten onze sociale relaties met anderen zien en onze aankoopgeschiedenis bevat onze dieetwensen. Digitale sporen krijgen steeds meer waarde voor wetenschappelijk onderzoek aangezien ze inzicht kunnen bieden in aspecten van ons dagelijks leven. Onderzoekers in verschillende disciplines hebben methodes en hulpmiddelen ontwikkeld om toegang tot deze data te krijgen. Eén daarvan, de focus van dit proefschrift, is **data donatie**.

Doel

Het doel van dit proefschrift is om een **alternatieve benadering voor data donatie te ontwikkelen, gebaseerd op principes van het datafeminisme**, die donoren in staat stelt hun data te leren kennen en om zorgvuldig met hun data om te gaan en die verschillende vormen van kennis en participatie bevordert. Deze alternatieve benadering, genaamd Gevoelige Data Donatie (SDD), richt zich op intieme onderzoekscontexten, zoals de woonomgeving en het lichaam. Deze contexten zijn inherent privé en contextafhankelijk, wat het cruciaal maakt om de relatie tussen mensen en hun data samen te verkennen en vorm te geven.

Overzicht en Resultaten

Hoofdstuk 1 motiveert waarom er behoefte is aan een alternatieve benadering voor data donatie, voortkomend uit de principes van het datafeminisme. Dit hoofdstuk presenteert de doelstelling van dit proefschrift, introduceert de onderzoeksvragen en methodologische keuzes en beschrijft de participatieve Onderzoek door Ontwerp (RtD) aanpak.

Hoofdstuk 2 beschrijft de centrale concepten en huidige praktijken rondom data donatie zoals beschreven in de literatuur. Het integreert deze concepten en praktijken in een conceptueel raamwerk en bekritiseert ze vanuit het datafeminisme [52]. Deze kritiek vormt het startpunt voor een eerste aanzet tot de principes en aanpak van Gevoelige Data Donatie (Sensitive Data Donation).

Hoofdstuk 3 onderzoekt hoe data donatie verschillende vormen van data, anders dan digitale sporen, zoals de belichaamde en gesitueerde kennis van mensen, kan integreren. Dit hoofdstuk introduceert een eerste iteratie van een empirisch data

donatieproces. Het focust op het creëren van waarde en het cureren en interpreteren van data. Het wordt toegepast in een onderzoeksproject met als doel om mensen die hun menstruatiecyclus bijhouden te ondersteunen in hun interacties met de data die zij genereren. Het hoofdstuk wordt afgesloten met een reflectie op de actoren die betrokken zijn bij data donatie en hoe hun interpretatie van data hun rol beïnvloedt.

Hoofdstuk 4 onderzoekt hoe data donatie het begrip van data kan vergroten, om daarmee donoren in staat te stellen hun deelname (opnieuw) te overwegen. Dit hoofdstuk introduceert een tweede iteratie van het empirische data donatieproces. Het focust op stapsgewijze ondersteuning van donoren om hun data te leren kennen, voordat ze besluiten deze data te doneren, tijdens vrijwillige deelname aan een serie activiteiten. Het proces wordt geïmplementeerd in een onderzoeksproject met als doel om de menselijke perceptie van spraakopnamen te onderzoeken - data verzameld door digitale stemassistenten. Het hoofdstuk wordt afgesloten met een reflectie op het voortdurende en dynamische proces om toestemming te geven voor data donatie in de praktijk.

Hoofdstuk 5 duikt in de data en hoe hun eigenschappen worden ervaren. Het hoofdstuk onderzoekt hoe donoren de ‘sensitiviteit’ en ‘intimiteit’ van hun data, in het bijzonder spraakopnamen van Google Assistent, ervaren wanneer ze geconfronteerd worden met een overzicht daarvan. Het introduceert de concepten ‘sensitieve’ data en ‘intieme’ data, die rekening houden met de nuances en contextuele eigenschappen van data.

Hoofdstuk 6 is een ***intermezzo***. In dit hoofdstuk wordt dataslip geïntroduceerd. Het is een provocerend artefact dat de persoonlijke digitale sporen van mensen materialiseert in de vorm van een kassabon, met als doel om bewustzijn over de verzameling van persoonlijke data te vergroten en een ongemakkelijk gevoel uit te lokken. Verder beschrijft dit hoofdstuk het gebruik van dataslip om het gesprek met mensen aan te gaan over de uitdagingen rondom de verzameling van persoonlijke data en om alternatieve toekomstbeelden te verkennen.

Hoofdstuk 7 onderzoekt hoe data donatie verschillende niveaus van participatie kan bevorderen, namelijk bijdragen, samenwerken en co-creëren. Dit hoofdstuk introduceert een derde iteratie van het empirische data donatieproces. Het richt zich erop om de specifieke activiteiten voor elk van de drie participatieniveaus vorm te geven in het proces. Het proces wordt toegepast in een onderzoeksproject met als doel de ervaring van atleten wat betreft de impact van hun menstruatiecyclus op hun sport te onderzoeken met behulp van data over hun fysieke activiteiten. Het hoofdstuk wordt afgesloten met een reflectie op hoe verschillende niveaus van participatie tot andere keuzes leiden en zodoende het data donatieproces én de data beïnvloeden.

Hoofdstuk 8 sluit af met een overzicht van het onderzoek. Het hoofdstuk presenteert de bevindingen, benadrukt de bijdragen van het onderzoek, bespreekt de implicaties, reflecteert op de beperkingen en geeft aanbevelingen voor toekomstig onderzoek.

Acknowledgements

Without the care, support, and availability (to gossip) of many lovely people, this thesis would not have been here in its physical or digital manifestation. I'm extremely grateful to all of them.

First, I want to acknowledge the generous contribution of many people who participated in the studies that constitute this thesis by donating their (sensitive) data. Thank you for trusting me with your data, your patience when the data donation process was not as smooth as you would have wanted, your time, and your awesome ideas. I hope this thesis honors our collaboration! I'm also grateful to the many individuals and organizations who connected with my research, invited me to share it with their communities, and helped me invite potential participants.

Gerd Kortuem, you challenged me to be critical about my research and encouraged me to find and embrace my voice. I learned a lot from you and I'm very proud we eventually found a way to collaborate. Thank you for always being open to having difficult conversations. **Jacky Bourgeois**, you allowed me the freedom and flexibility to pursue the research that excited me. Thank you for inviting me to this PhD and supporting me in many ways – the most fun ones being helping me move an office chair in your *bakfiets* and fitting dataslip into a tiny elevator in Utrecht.

Jeff Love, you motivated me not to settle, reminded me to be proud of myself and my research, and even stepped in as my unofficial co-promotor during the last months of my PhD. Thank you for giving me thoughtful critical feedback and guiding me through the practicalities of collecting (sensitive) personal data. I appreciate how you made time for everything: from gossiping over coffee and recommending books to helping me with many data stewardship questions and reading my thesis. **Carlo van der Valk**, you helped me with the data donation platform every time I needed it. Without you, it would have taken me much longer to complete my PhD. Thank you for your help and patience and for attending scary meetings with the privacy officers with me. Also, thank you for bringing the best energy and costume to the parties!

My dear office mates, **Di Yan**, **Hosana Morales**, **Roos Teeuwen**, and **Vasilis Milias**, from Zoom coffees during the pandemic to many many in-person hugs, beers, and tears, sharing this PhD journey with you was wonderful! I am extremely proud of everything we overcame and all the ways we were there for each other. I can't wait to celebrate with all of you during your defenses. I love you! **Di Yan**, sharing supervisors and adjacent research topics brought us close, and your sweetness brought us closer. I admire your strength and perseverance. **Hosana Morales**, we became friends instantly. I appreciate our friendship, deep conversations, and similar but different perspectives. Thank you for the daily check-ins, constant encouragement, and for challenging me when necessary. I owe you a lot. **Roos Teeuwen**, I will never forget the first day we met and had a pizza on the street during the pandemic. Thank you for helping me adjust to The Netherlands and reminding me to slow down and take a break. **Vasilis Milias**, the office nerd, you were the first to go through every process. Thank you for sharing your experience with laughter and facing the demons of thermal printing with me.

To the KInD-est colleagues: **Achilleas Psyllidis**, **Alessandro Bozzon**, **Andrea Mauri**, **Céline Offerman**, **Denis Bulygin**, **Evangelos Niforatos**, **Francesca Mauri**, **Garoa Gomez Beldarrain**, **Himanshu Verma**, **James Broadhead**, **Kars Alfrink**, **Mireia Yurrita Semperena**, **Ruben Dekker**, **Samuel Kernan Freire**, **Sara Colombo**, **Shatha Degachi**, **Tianhao He**, **Tilman Dingler**, **Uğur Genç**, **Wilfred van der Vegte**, **Wo Meijer** and **Yunzhong Zhou**, thank you for the fun lunch chats that went places we could never have anticipated. Many thanks to **Himanshu Verma**, who was a big part of making the lunch chats go crazy, for his many tips, and helping me navigate the academic job market.

I'm particularly grateful to all the KInD PhDs and postdocs. I appreciated how sharing our experiences made me feel less alone with my struggles. I enjoyed all the peer-review sessions we organized and the fantastic parties. **James Broadhead**, the days we spent together at the PMB building the dataslip family were among the best ones of my PhD journey. You are a joy to be around, your optimism is inspiring, and you are an amazing performer. Thank you for all your help and motivation! **Wo Meijer**, I loved all our gossip sessions and our Estonian adventure. **Andrea Mauri** and **Uğur Genç**, thank you for showing me the joys and struggles of life after the PhD. It was always fun spending time with you!

Thanks to the IDE students I was lucky enough to advise on their graduation projects: **Eline Muijters**, **Alina Wundsam**, and **Julia Rademaker**. You were a pleasure to work with and have inspired me with your enthusiasm and curiosity. I'm thrilled to see you shine in your chosen paths. I can't wait to see all the amazing things you do next.

I'm incredibly grateful to the young researchers with whom I had the pleasure to collaborate. **Wiebke Hutiri** in you, I always found enthusiasm, encouragement, and the most incredible energy. The many conversations, coffees, and runs we shared always motivated me and reminded me of the fun side of doing research. **Renee Noortman**, I'm so happy we met. In you, I found a fellow feminist data-design researcher with passion, drive, and curiosity. You inspire me to see the positive – even when lost in a resort – and not only imagine but create a better future. **Anna Dobrosovestnova**, I admire your sharpness and critical thinking. Thank you for the early feedback on my work and for inviting me to your lab in Vienna. Thank you also to **Emiel Rijshouwer** and **Els Leclercq**, for inviting me into their project and showing me the value of doing participatory research and activism with care.

Thanks to the emerging Data and Design community for many inspiring conversations and events: **Janne van Kollenburg**, **Yvette Shen**, **Dave Murray-Rust**, **Dajana Nedić**, **Juan Carlos Jimenez**, **Wo Meijer**, **Pranshu Kumar Kumara Chaudhary**, **Peter Lovei**, **Renee Noortman**, **Romain Toebosch**, **Albrecht Kurze**, **Mathias Funk**, **Sandy Gould**, and **Samuel Huron**. I'm particularly grateful to **Sandy Gould** for his mentorship and career advice.

Thank you to **Alba Ávila**, **Madeline Balaam**, **Frances Brazier**, **Judith Rietjens**, and **Maaike Kleinsmann** for agreeing to assess my thesis. Your research has inspired me deeply and I feel honored to discuss my work with such an amazing group of women. **Alba Ávila**, you motivated me to challenge gender stereotypes and create space for myself in a male-dominated field. In you, I've found motivation and inspiration for over 10 years. **Madeline Balaam**, thank you for your kindness and for welcoming me into your *girl computing* lab for a few days while I was wrapping up this thesis.

To my friends abroad, **Alexandra Urbano**, **Ava Ghafari**, **Marine Rérolle**, and **Laura Santos**, thank you for never asking about my thesis. **Daniela Moreno**, the marathons and triathlons we planned together over the past four years were a constant reminder to set limits on my PhD and make time for other things. Thank you for motivating me and challenging me to do more. **Angelica Cortés**, thank you for helping me train for the crazy races and events I decided to participate in and for your friendship. **Laura Thomas**, you believed in my ability to do a PhD from the beginning. Thank you for your love and encouragement. **Veronica Herrera**, thank you for being one message away, and always being there for me. **Cristina Navarrete**, **Cristina Pelaez** and **Natalia Alvarez**, thank you for welcoming me as if I had never left.

To my friends in The Netherlands, you all contributed greatly to this thesis by keeping me away from it. **Antonio Manesco** and **Marilia Silva**, you became our (latin) family away from home. Thank you for showing up to my races, and dataslip exhibitions, for letting me sleep on top of you, and for teaching me the important difference between "pao de queijo" and "pão de queijo". I hope we can continue having brunches, wild adventures, and fighting Latin-American stereotypes together. **Felipe Bucci Ancapi**, and **Kostas Vilkelis**, you made the difficult early pandemic days way more enjoyable. **Carolina Carrillo**, I fondly remember our walks during the pandemic. It was a leaf!

To my parents, thank you for your constant support, for accompanying me and advising me during the most difficult days of the PhD, and for teaching me (and reminding me) that I can do anything. I love and admire you infinitely! **Mami**, you taught me by example to be a strong, brave, and determined woman. You also always invited me to be kind and care for others. **Papi**, you showed me a thousand worlds through reading and taught me how exciting it is to learn. **Pauli**, when I'm with you I feel at home. You, more than anyone, inspire me to forge my path and break it all when necessary. I'm incredibly proud to be your sister and now a super cool aunt to **Aloisia**. To my family in Colombia, thank you for loving and supporting me from a distance and celebrating me every time I'm there.

Gabi, you moved to a cold and windy country in the middle of a pandemic to be with me. You also married me so we could face the uncertainties of life after PhD with a little more ease. I'm constantly impressed by you. You are a super good source of gossip and advice, the first to challenge me, and the kindest person I know. You have contributed greatly to this thesis in multiple ways, and I value them all – but it is still *my* thesis and not *our* thesis. Thank you for loving and caring for me – even through this PhD that made me not as fun to be around some days – and for making breakfast. I know whatever comes next will be okay because you are with me. Con sentimiento para ti!

Agradecimientos

Esta tesis no existiría – en su manifestación física o digital – sin el cuidado, apoyo y disponibilidad (para echar chisme) de muchas personas encantadoras. Estoy muy agradecida con todas ellas.

Primero, quiero reconocer la generosa contribución de todas las personas que participaron en los estudios que constituyen esta tesis donando sus datos (sensibles). Gracias por confiarme sus datos, su paciencia cuando el proceso de donación de datos no fue tan sencillo como hubieran deseado, su tiempo y sus increíbles ideas. ¡Espero que esta tesis honre nuestra colaboración! También estoy agradecida con las muchas personas y organizaciones que se conectaron con mi investigación y me ayudaron a invitar a posibles participantes.

Gerd Kortuem, me desafiaste a ser crítica con mi investigación y me animaste a encontrar mi voz. Aprendí mucho de ti y estoy muy orgullosa de que finalmente hayamos encontrado una manera de colaborar. Gracias por estar siempre abierto a tener conversaciones difíciles. **Jacky Bourgeois**, me permitiste la libertad y flexibilidad para realizar la investigación que me entusiasmaba. Gracias por invitarme a este doctorado y apoyarme de muchas maneras; las más divertidas fueron ayudarme a llevar una silla de oficina en tu *bakfiets* y meter a dataslip en un pequeño ascensor en Utrecht.

Jeff Love, me motivaste a no conformarme, me recordaste que debía estar orgullosa de mí y de mi investigación, e incluso interviniste como mi copromotor no oficial durante los últimos meses de mi doctorado. Gracias por brindarme comentarios críticos y reflexivos y guiarme a través de los aspectos prácticos de la recopilación de datos personales (sensibles). Aprecio mucho el tiempo que me dedicaste para varias actividades, incluyendo chismosear, tomar café, recomendarme libros, responder a mis preguntas de manejo de datos y darme retroalimentación. **Carlo van der Valk**, me ayudaste con la plataforma de donación de datos cada vez que lo necesité. Sin ti, me habría demorado mucho más terminando el doctorado. Gracias por tu ayuda y paciencia y por asistir conmigo a reuniones aterradoras con los funcionarios de privacidad. Además, ¡gracias por traer la mejor energía y disfraces a las fiestas!

A mis queridos compañeros de oficina, **Di Yan**, **Hosana Morales**, **Roos Teeuwen** y **Vasilis Milias**, desde los cafés por Zoom durante la pandemia hasta muchos, muchos abrazos, cervezas, y lágrimas, ¡compartir este viaje de doctorado con ustedes fue maravilloso! Estoy extremadamente orgullosa de todo lo que superamos y de todas las formas en que estuvimos ahí el uno para el otro. No puedo esperar para celebrar con todos ustedes durante sus defensas. ¡Los quiero! **Di Yan**, compartir supervisores y temas de investigación adyacentes nos acercó, y tu dulzura nos acercó todavía más. Admiro tu fuerza y perseverancia. **Hosana Morales**, nos hicimos amigas al instante. Aprecio nuestra amistad, nuestras interminables conversaciones sobre todo tipo de temas, y nuestras perspectivas tan similares pero diferentes. Gracias por los mensajes diarios, por darme ánimo, y por retarme cuando lo necesité. Te debo mucho. **Roos Teeuwen**, nunca olvidaré el primer día que nos conocimos y comimos pizza en la calle durante la pandemia. Gracias por ayudarme a adaptarme a los Países Bajos y recordarme que está bien ir más despacio y descansar. **Vasilis Milias**, el nerd de la oficina, fuiste el primero en pasar por cada paso del doctorado. Gracias por compartir tu experiencia entre risas y enfrentar conmigo los demonios de la impresión térmica.

A los demás colegas de KInD: **Achilleas Psyllidis**, **Alessandro Bozzon**, **Andrea Mauri**, **Céline Offerman**, **Denis Bulygin**, **Evangelos Niforatos**, **Francesca Mauri**, **Garoa Gomez Beldarrain**, **Himanshu Verma**, **James Broadhead**, **Kars Alfrink**, **Mireia Yurrita Semperena**, **Ruben Dekker**, **Samuel Kernan Freire** , **Sara Colombo**, **Shatha Degachi**, **Tianhao He**, **Tilman Dingler**, **Uğur Genç**, **Wilfred van der Vegte**, **Wo Meijer** y **Yunzhong Zhou**, gracias por las divertidas charlas durante el almuerzo que llegaron a lugares que nunca hubiéramos anticipado. Muchas gracias a **Himanshu Verma**, quien contribuyó en gran medida a que las charlas del almuerzo se alocaran, por sus muchos consejos y por ayudarme a navegar en el mercado laboral académico.

Estoy especialmente agradecida con todos los PhDs y postdocs de KInD. Compartir nuestras experiencias me hizo sentir menos sola con mis dificultades. Gracias por todas las sesiones de revisión y retroalimentación que organizamos y las fantásticas fiestas. **James Broadhead**, los días que pasamos juntos en el PMB construyendo la familia dataslip fueron de los mejores de mi doctorado. Es un placer estar cerca de ti, tu optimismo es inspirador y eres un artista increíble. ¡Gracias por toda tu ayuda y motivación! **Wo Meijer**, me encantaron todas nuestras sesiones de chismes y nuestra aventura en Estonia. **Andrea Mauri** y **Uğur Genç**, gracias por mostrarme las alegrías y dificultades de la vida después del doctorado. ¡Siempre fue divertido pasar tiempo con ustedes!

Gracias a las estudiantes de IDE que tuve la suerte de asesorar en sus proyectos de graduación: **Eline Muijters**, **Alina Wundsam** y **Julia Rademaker**. Fue un placer trabajar con ustedes. Su entusiasmo y curiosidad son inspiradores. Me emociona verlas brillar en los caminos que eligieron. No puedo esperar a ver todas las cosas increíbles que harán a continuación.

Estoy increíblemente agradecida con investigadoras jóvenes con quienes tuve el placer de colaborar. **Wiebke Hutiri**, en ti siempre encontré entusiasmo, aliento y la energía más increíble. Las muchas conversaciones, cafés, y carreras que compartimos me motivaron y siempre me recordaron que hacer investigación puede ser divertido. **Renee Noortman**, estoy muy feliz de habernos conocido. En ti encontré a una investigadora feminista en el mismo campo con pasión, ambición, y curiosidad. Me inspiras a ver el lado positivo – incluso al perderse en un campo de golf – y a no solo imaginar sino crear un futuro mejor. **Anna Dobrosovestnova**, admiro tu agudeza y tu pensamiento crítico. Gracias por los primeros comentarios sobre mi trabajo y por invitarme a visitar tu laboratorio en Viena. Gracias también a **Emiel Rijshouwer** y **Els Leclercq**, por invitarme a su proyecto y mostrarme el valor de hacer investigación participativa y activismo con cuidado.

Gracias a la comunidad emergente de Datos y Diseño por tantas conversaciones y eventos inspiradores: **Janne van Kollenburg**, **Yvette Shen**, **Dave Murray-Rust**, **Dajana Nedić**, **Juan Carlos Jiménez**, **Wo Meijer**, **Pranshu Kumar Kumara Chaudhary**, **Peter Lovei**, **Renee Noortman**, **Romain Toebosch**, **Albrecht Kurze**, **Mathias Funk**, **Sandy Gould** y **Samuel Huron**. Estoy especialmente agradecida con **Sandy Gould** por su mentoría y sus consejos.

Gracias a **Alba Ávila**, **Madeline Balaam**, **Frances Brazier**, **Judith Rietjens** y **Maaike Kleinsmann** por aceptar evaluar mi tesis. Su investigación me ha inspirado profundamente y me siento honrada de poder hablar de mi trabajo con un grupo de mujeres tan increíble. **Alba Ávila**, me motivaste a desafiar los estereotipos de género y crear un espacio para mí en un campo dominado por hombres. En ti he encontrado motivación e inspiración durante más de 10 años. **Madeline Balaam**, gracias por tu amabilidad y por darme la bienvenida a tu laboratorio de *girl computing* durante unos días mientras terminaba esta tesis.

A mis amigas en el extranjero, **Alexandra Urbano**, **Ava Ghafari**, **Marine Rérolle** y **Laura Santos**, gracias por nunca preguntar por mi tesis. **Daniela Moreno**, los maratones y triatlones que planeamos juntas durante los últimos cuatro años fueron un recordatorio constante de poner límites a mi doctorado y sacar tiempo para otras cosas. Gracias por motivarme y desafiarme a hacer más. **Angelica Cortés**, gracias por ayudarme a entrenar para las carreras y eventos en los que decidí participar y por tu amistad. **Laura Thomas**, creíste en mi capacidad para hacer un doctorado desde el principio. Gracias por tu amor y aliento. **Veronica Herrera**, gracias por estar a un mensaje de distancia y estar siempre ahí para mí. **Cristina Navarrete**, **Cristina Pelaez** y **Natalia Alvarez**, gracias por recibirme como si nunca me hubiera ido.

A mis amigos de los Países Bajos, ustedes contribuyeron enormemente a esta tesis manteniéndome alejada de ella. **Antonio Manesco** y **Marilia Silva**, ustedes se convirtieron en nuestra familia (latina) lejos de casa. Gracias por asistir a mis carreras y exhibiciones de dataslip, por dejarme dormir encima de ustedes y enseñarme la diferencia fundamental entre “pao de queijo” y “pão de queijo”. Espero que podamos seguir teniendo brunchs, aventuras salvajes y luchando juntos contra los estereotipos latinoamericanos. **Felipe Bucci Ancapi** y **Kostas Vilkelis**, hicieron que los difíciles primeros días de la pandemia fueran mucho más agradables. **Carolina Carrillo**, recuerdo con cariño nuestras caminatas durante la pandemia. ¡Era una hoja!

A mis papis, gracias por el apoyo constante, por acompañarme y aconsejarme durante los días más difíciles del doctorado y por enseñarme (y recordarme) que puedo hacerlo todo. **Mami**, me enseñaste con tu ejemplo a ser una mujer fuerte, valiente y determinada. También me invitaste siempre a ser amable e interesarme por los demás. **Papi**, me mostraste mil mundos a través de la lectura y me enseñaste lo emocionante que es aprender. ¡Los quiero y admiro infinitamente! **Pauli**, cuando estoy contigo me siento en casa. Tú, más que nadie, me inspira a forjar mi camino y a romperlo todo cuando es necesario. Estoy muy orgullosa de ser tu hermana, y la tía cool de **Aloisia**. A mi familia en Colombia, gracias por el amor y el apoyo desde la distancia y por celebrarme siempre.

Gabi, te mudaste a un país frío y ventoso en medio de una pandemia para estar conmigo. También te casaste conmigo para que pudiéramos afrontar las incertidumbres de la vida después del doctorado con más facilidad. Estoy constantemente impresionada por ti. Eres una súper buena fuente de chismes y consejos, el primero en desafiarme y la persona más amable y que conozco. Has contribuido infinitamente y de muchas formas a esta tesis, las valoro todas – pero es *mi* tesis y no *nuestra* tesis. Gracias por amarme y cuidarme, incluso cuando no fui mi mejor yo, y por hacer el desayuno. Sé que lo que sea que venga después estará bien porque estás conmigo. ¡Con sentimiento para ti!

Preface

Everything in this thesis I owe to stubbornness and endurance.

Like many other things in my life as a migrant, my PhD experience began on a plane. I was flying from Bogotá to Frankfurt to visit my sister and finish arranging my move to Delft. Countless people were also flying at that time carrying – probably without knowing it – COVID-19 around the world, some of them on the same plane as me.

The COVID-19 pandemic was the first of many opacities in my PhD – unclear situations with undefined and mutable rules that no one communicates or explains. Would knowing these opacities a priori have changed anything? I'm not sure. When I decided to move to Delft and live in a small studio apartment, I didn't know that there, alone, I would spend several months trying to decipher many other opacities.

The PhD is inherently opaque. It is an individual process – lonely? –, shaped by the people who participate in it and the (power) relationships between them, the financing scheme, and the particularities of the institution in which it is carried out. It is difficult to know what to expect from a PhD beyond that it is a long and complicated process. But how long? How complicated? For all of us doing a PhD, it is the first time we are doing a PhD, we don't know what a *normal* PhD is or what it could/should be like. In fact, the process is very different for everyone and there is not a single way of doing a PhD.

Every PhD is opaque as you go through it. Opacity is not necessarily bad; it is part of a learning process. However, opacity allows certain non-normal – yet normalized – situations that would not be allowed otherwise. Furthermore, it delays the realization that situations that seem normal are not or should not be. My PhD – the research process that resulted in this thesis – was not normal, it was disempowering, yet it was formative.

To honor my experience during the opaque and non-normal – normalized – process that resulted in this thesis, and to accompany anyone who is in the middle of a non-normal process, I would like to highlight some non-normalities. But first, I want to emphasize that this is my thesis, and it is my voice that narrates an experience that was mine, but that I also shared with other people. Their voices are not reflected here, and therefore the non-normalities described here are partial.

First, it is **not normal** for (pre-existing) (power) relationships between the parties involved in your PhD to hinder your learning process. Second, it is **not normal** frequently not receiving timely and/or critical feedback on your work. Third, it is **not normal** for (thinking about) progress meetings to lead you to feelings of fear, boredom, or anger, especially if your fellow PhD colleagues feel the same way. Fourth, it is **not normal** that considerations that clearly affect your ability to do a PhD – such as your immigration status – are not considered when deciding what to do about your PhD. Fifth, it is **not normal** to have to create mechanisms to give and receive support and feedback because either they do not exist or those that do exist are not normal.

Now, on another plane, this time from Bogotá to Amsterdam, I think that, even if my

PhD was opaque, not normal, and disempowering, it was an enriching experience. But, above all, an experience of community, care and support. I took care of myself, other people took care of me, I took care of other people, and together we took care of each other. I also took care of this thesis, and put infinite perseverance, energy and a love into it. I hope that this thesis contributes to rethinking how we relate to and interact with our data and invites us (researchers) to take care of the people behind the data. Also, may this thesis invite us to care for and support ourselves and each other while we go through opacities, non-normalities, and beyond.

Alejandra Gómez Ortega
Traveling, May 2024

Prefacio

Todo en esta tesis se lo debo a mi terquedad y resistencia.

Como muchas otras cosas en mi vida como migrante, mi experiencia de doctorado comenzó en un avión. Volaba de Bogotá a Frankfurt para visitar a mi hermana y concluir mi mudanza a Delft. En ese momento volaban también incontables personas cargando – probablemente sin saberlo – el COVID-19 alrededor del mundo, algunas de ellas seguramente en mí mismo avión.

La pandemia de COVID-19 fue la primera de muchas opacidades en mi doctorado – situaciones poco claras con reglas indefinidas y mutables que nadie te comunica y mucho menos te explica y que de haber conocido a priori habrían cambiado algo, ¿o no? Cuando decidí mudarme a Delft y elegir un pequeño estudio no sabía que ahí, sola, iba a pasar varios meses tratando de descifrar otras opacidades.

El doctorado es inherentemente opaco. Es un proceso individual – ¿solitario? –, moldeado por las personas que participan de él y las relaciones (de poder) entre ellas, el esquema de financiación, y las particularidades de la institución en la que se realiza. Es difícil saber que esperar de un doctorado más allá de que es un proceso largo y complejo. Pero ¿qué tan largo? ¿qué tan complejo? Todos los que hacemos un doctorado por primera vez estamos haciendo un doctorado por primera vez, no sabemos que es o como se ve un doctorado *normal*. De hecho, el proceso es muy diferente para cada uno y no existe una única forma de realizar un doctorado.

Todo doctorado es opaco mientras se atraviesa. Esta opacidad no es necesariamente mala, es parte de un proceso de aprendizaje. Sin embargo, la opacidad permite ciertas situaciones no normales – pero si normalizadas – que ante un marco de referencia cualquiera no se permitirían. Además, dilata el darse cuenta de que situaciones aparentemente normales no lo son, o no deberían serlo. Mi doctorado – el proceso de investigación que resulto en esta tesis – no fue normal, fue horrible, formativo pero horrible.

Para honrar mi experiencia durante el proceso opaco y no normal – normalizado – que resulto en esta tesis, y para acompañar a cualquiera que esté en medio de un proceso no normal, me gustaría resaltar algunas no normalidades. Pero antes, quiero enfatizar que esta es mi tesis, y es mi voz la que narra una experiencia que fue mía, pero que también compartí con otras personas. Sus voces no están acá reflejadas, y por ende las no normalidades acá descritas son parciales.

Primero, **no es normal** que las relaciones (preexistentes) (de poder) entre las partes involucradas en tu doctorado dificulten tu proceso de aprendizaje. Segundo, **no es normal** que con frecuencia no recibas oportunamente retroalimentación critica. Tercero, **no es normal** que (pensar en) reunirte a discutir tu trabajo produzca sentimientos como miedo, tedio, o rabia – especialmente si tus compañeros del doctorado sienten lo mismo. Cuarto, **no es normal** que consideraciones que claramente afectan tu capacidad de hacer un doctorado – como tu situación migratoria – no se tengan en cuenta al tomar decisiones sobre los quehaceres del doctorado. Quinto, **no es normal** tener que crear mecanismos para dar y recibir apoyo y retroalimentación porque o no existen o los que existen no son normales.

Ahora, en otro avión, esta vez de Bogotá a Ámsterdam, pienso que, aunque mi doctorado fuera opaco, no fuera normal y fuera ¿horrible? – ¡que palabra tan fuerte! – fue una experiencia enriquecedora. Pero, sobre todo, una experiencia de comunidad, cuidado y apoyo. Cuidé de mí, otras personas cuidaron de mí, cuidé de otras personas, y nos cuidamos juntas. También cuidé de esta tesis, y puse en ella infinita perseverancia, energía y mucho amor. Así pues, espero que esta tesis contribuya a repensar la forma en la que nos relacionamos con nuestros datos y a cuidar a las personas detrás de los datos. También, que esta tesis invite a cuidar(nos) y apoyar(nos) mientras atravesamos opacidades, no normalidades, y demás.

Alejandra Gómez Ortega
Viajando, Mayo 2024


3 2021 03 01 00:15,21 2437 01 826.00, ...
4 2021-03 01,00 00,20 21,35.50,22.00, ,......
5 2021 03 01 00:15,21 2437 01 826.00, ...
6 2021-03 01,00 00,20 21,35.50,22.00, ,......
7 2021 03 01 00:15,21 2437 01 826.00, ...
8 2021-03 01,00 00,20 21,35.50,22.00, ,......
9 2021 03 01 00:15,21 2437 01 826.00, ...
10 2021-03 01,00 00,20 21,35.50,22.00, ,......
11 2021 03 01 00:15,21 2437 01 826.00, ...
12 2021-03 01,00 00,20 21,35.50,22.00, ,......
13 2021 03 01 00:15,21 2437 01 826.00, ...
14 2021-03 01,00 00,20 21,35.50,22.00, ,......
15 2021 03 01 00:15,21 2437 01 826.00, ...
16 2021-03 01,00 00,20 21,35.50,22.00, ,......
17 2021 03 01 00:15,21 2437 01 826.00, ...
18 2021-03 01,00 00,20 21,35.50,22.00, ,......
19 2021 03 01 00:15,21 2437 01 826.00, ...
20 2021-03 01,00 00,20 21,35.50,22.00, ,......
21 2021 03 01 00:15,21 2437 01 826.00, ...
22 2021-03 01,00 00,20 21,35.50,22.00, ,......
23 2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
Date,Time,avg(temp),avg(humidity),max(p1)
P1),max(tp1),min(tp2),max(tp2),min(tp3),
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,20 21,35.50,22.00, ,......
BANK
1234 5678 9012 3456
123

1

Introduction

1.1. Context and Motivation

As I write this thesis I'm interacting with various products and services that are collecting, storing, and using data about me, a smartwatch on my left wrist, a smartphone on the table, the digital music service where I listen to a playlist, and the internet browser through which I write. Like me, millions of people routinely interact with several products and services that are constantly collecting, storing, and using personal digital-trace data about them. Together, this data constitutes a digital trail [102] that is entangled with and can partially account for our behavior and even our most intimate experiences.

Two of the most important characteristics of the data within this trail are their multiplicity and temporal dimension. Multiplicity means data can be in more than one place at the same time. It can be copied and re-used for different purposes [155, 156]. Additionally, multiple copies of multiple types of data can be combined and aggregated. The temporal dimension anchors the data to a specific time and date. Thus, data can partially account for our behavior and experiences over time and it becomes possible to relate them to specific events and identify patterns and outliers. These characteristics render personal data highly valuable for scientific research. Multiplicity enables researchers to use data that was originally generated or collected for purposes other than research [22, 155] and to re-use existing (open) data, while the temporal dimension enables researchers to explore and investigate phenomena retrospectively and over time.

For decades, researchers and designers have developed different tools and methods to generate, collect, access, and re-use data. One emergent approach is **data donation**. Similar to the donation of blood or organs, the donation of data is an altruistic and voluntary transaction of personal data [22, 179] from a person who 'has'[1] data to another person (e.g., researcher) or entity (e.g., healthcare institution) who needs it.

Data donation is increasingly being discussed and applied across various domains. It offers opportunities to access new insights into people's behavior and experiences and opens the way to investigate critical research questions. For instance, researchers in the healthcare domain at the Robert Koch Institute in Germany developed the Corona-Datenspende App through which people could donate their health data (e.g., heart rate, sleep duration, step count) from wearable devices and fitness trackers (e.g., Apple Watch, Fitbit, Garmin) to monitor the COVID-19 pandemic [57, 166]. In the social sciences, Breuer et al. [34] developed a browser plugin that donors could install to donate information about public posts from their personal Facebook feeds (e.g., post title and author, number of reactions, and comments) to a research project about media use. Similarly, the human rights organization AlgorithmWatch [3] used data donation to investigate the functioning of algorithmic systems on TikTok. In the field of Human-Computer Interaction (HCI), Razi et al. [162] developed a web-based data donation system to collect private Instagram conversations from teens to identify online risks, including nudity and porn, sexual messages or solicitations, harassment, and violence, amongst others.

However, although data donation develops from values such as altruism, voluntariness, and solidarity [97, 155, 179] its application in practice is often transactional and appeals to other values, such as generalizability, efficiency, and speed. To encourage TikTok data donations, Zannettou et al. [215] offered potential donors financial com-

---

[1] 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 [97, 155].

pensation based on the types of data they choose to donate. It started at 5$ for the video viewing history (mandatory) and increased by 1$ for each additional type of data (e.g., like history, share history, app settings, comments, followers). They claimed that in this way they give *"power"* to data donors – the majority of which (95%) chose to donate *"almost all the data"* [215]. But did the donors really have the power to voluntarily decide? They were offered a choice, but not choosing (i.e., donating all data) could have been influenced by the monetary incentives, thus shifting the data donation away from voluntary to transactional. Additionally, were donors aware of the information they were sharing through fields such as messages? I mentioned earlier how data are entangled with our behavior and intimate experiences, including our social relationships. My TikTok messages might be mainly GIFs and emojis but still contain important information about me, my partner, my friends and colleagues, and my family. Is it worth 1$? Would 7$ represent a similar value to donors as a publication would to researchers?

In this thesis, I aim to develop an alternative approach to data donation informed by Data Feminism [51]. It proposes to interrogate and challenge (extractivist) practices in data science and data ethics informed by the ideas of intersectional feminism. For instance, by foregrounding power, who was it and who doesn't. Thus, it is a valuable lens to interrogate and challenge data donation practices such as that of Zannettou et al. [215], described above, where donors have power, but do they? D'Ignazio and Klein [51] provide strategies in the form of seven principles of Data Feminism:

1. **Examine Power:** It relates to the scrutiny of power dynamics by asking *who* questions (e.g., Who benefits and is neglected? Whose priorities are prioritized or overlooked?)

2. **Challenge Power:** It relates to challenging and changing an unjust status quo by taking action (e.g., collecting counter-data, imagining and proposing an alternative).

3. **Elevate Emotion and Embodiment:** It relates to valuing multiple forms of knowledge by challenging the perceived neutrality and objectivity of the data and elevating *"the knowledge that comes from people as living, feeling bodies in the world"* [51].

4. **Rethink Binaries and Hierarchies:** It relates to examining the assumptions behind counting, measuring, and classification in the process of creating knowledge (e.g., What is counted and what is not?).

5. **Embrace Pluralism:** It relates to including a broad range of perspectives and participants to contribute their knowledge to a data project at various stages by cultivating solidarity and a shared understanding. It builds upon Feminist practices in HCI [15, 16].

6. **Consider Context:** It relates to recognizing the context in which the data is generated; by accounting for its *situatedness* [90].

7. **Make Labor Visible:** It relates to recognizing, and valuing the work(-ers) involved in data projects by making them visible.

These principles call for challenging the differentials embedded in the relationships between some of the actors in data donation (i.e., data donors, data recipients, data) and raise important questions about the ethos and ethics of current data donation practices, such as: Who benefits from data donation and who is neglected? Who participates and how, people or their data? What forms of knowledge and data are prioritized?

In this alternative approach, which I call Sensitive Data Donation, I aim to include not only ethics and responsible research practices but also people's experiences of data donation and their latent needs. The latter has been seldom considered in empirical data donation research because of the domains, type of research questions, and contexts in which data donation is applied. To date, practical approaches to data donation (e.g., [34, 162, 204]) predominantly focus on developing practices, platforms, and methods from the perspective and needs of researchers, or simply use data donation to collect data (e.g., [204, 215]). These include people, as data donors, but their data donation experience is not discussed or accounted for. Similarly, data donation research that accounts for people's experiences mainly focuses on their motivations (not) to donate (e.g., [152, 153, 179]) and privacy concerns through large-scale surveys or interviews. They include people, as research participants, but these participants don't experience data donation (i.e., they do not participate in a process of donating data). Although I develop and propose similar practices, platforms, and methods, I strive to include people and their data donation experiences. Thus, I scope my research questions and methodological choices in a way that enables people to participate in and experience a specific form of data donation and reflect on this experience. Through this process, I adopt mostly qualitative methodologies and feminist approaches that advocate for reciprocity, collaboration, and some level of self-disclosure with research participants [53].

1.2. Thesis Position in Research Landscape

This thesis is informed and influenced by three research areas: Critical Data Studies (Section 1.2.1), Designerly Data Interpretation (Section 1.2.2), and (Data) Ethics and Privacy (Section 1.2.3). At the intersection of these three areas (Fig. 1.1), this thesis focuses on **personal digital-trace data** potentially available through data donation.

Personal data is defined in the European General Data Protection Regulation (GDPR) as any information related to an identified or identifiable person [65]. The GDPR further defines a special category of personal data: 'sensitive data.' It includes racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, genetic data, biometric data processed solely to identify a human being, health-related data, and data concerning a person's sex life or sexual orientation [65, Art. 9].

Personal digital-trace data corresponds to the data that is generated and collected as a result of people's interactions with connected products and services. These data could potentially contain or reveal sensitive information – and be categorized as sensitive data. Personal digital-trace data have multiple forms, types, and formats [209], including (1) digital communications (e.g., DMs on Tinder), (2) entertainment consumption (e.g., Netflix), finances (e.g., credit card purchases), (3) physical activity (e.g., daily steps), and (4) physiological signals (e.g., body temperature), among many others.

The (potential) availability of personal-digital trace data for scientific research and

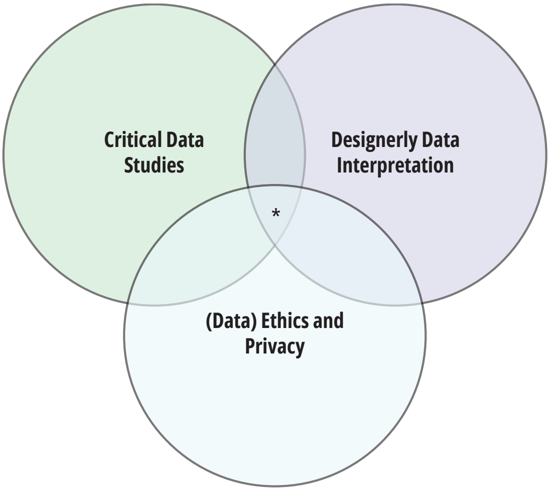


Figure 1.1: Research areas that ground this thesis.

other purposes was enabled by the GDPR in 2018. It defined the rights that data subjects (i.e., individuals) have around personal data collection and use from data controllers (i.e., product and service providers) [Art. 12-22][65]. Specifically, the *right to access* and *right to data portability*, which allow individuals to receive their personal data from data controllers, who are required to provide a copy in a machine-readable format.

1.2.1. Critical Data Studies

Several scholars in the fields of Human-Computer Interaction (HCI), Science and Technology Studies (STS), philosophy, and design propose to challenge common notions and imaginaries around data, such as that they are *clean* or *objective*. Feminist scholar Donna Haraway [90] introduced the concept of **situated knowledges** to illustrate how all forms of knowledge, including data, are produced by specific people in specific circumstances. Expanding on this concept, Yanni Loukissas [115] argued that data are **local**. Thus, deeply attached to and shaped by the places and contexts where they are created and used. Moreover, he challenges the assumption that data are *smooth* and *singular* and underlines that they are plural and heterogeneous. Building upon feminist scholarship, including Haraway's situated knowledge, D'Ignazio and Klein [51] propose Data Feminism as a way to foreground the power differentials embedded in (extractivist) data practices and to interrogate, challenge, and change certain practices in data science that reinforce existing inequalities. This thesis builds on these considerations: that data are not neutral or objective but situated and local, and that power differentials are embedded in practices around data collection.

1.2.2. Designerly Data Interpretation

Personal digital-trace data is increasingly used in designerly processes and activities. Design researchers and practitioners are incorporating digital-trace data into their practices and using it to design *from*, *with*, and *by* [188]. They have developed several methods and approaches that integrate behavioral data with contextualization, interpretation, and sense-making activities as part of exploratory design processes, including 'Participatory Data Analysis' [28], 'Data-Enabled Design' [26], 'Contextual Inquiry' [79], and 'Articulation Work' [195]. For instance, Bogers and colleagues [26] developed a prototype of a connected baby bottle to collect behavioral data (i.e., movement, temperature) about bottle feeding experiences and visualized the data using a canvas to facilitate participants' interpretation and promote reflection during interview sessions. Similarly, Tolmie and colleagues [195] developed a prototype that supported the legibility of sensor data (e.g., temperature, humidity, motion) collected at home and invited people to interpret and account for the data. These activities involve various orders of reasoning (e.g., place, time, people, practices, and events) and lead to relating data to specific events and reflecting through data. This thesis builds on these interpretative, subjective, and collaborative approaches to (designing with) data.

1.2.3. (Data) Ethics and Privacy

Practices around personal-digital trace data collection and use comprise several ethical challenges [20, 131, 192]. These include: (1) **accountability**, ensuring people have clear expectations – and that these are met [126, 199] – regarding the objectives of collecting, processing, and sharing data as well as the potential benefits and harms that could be derived from these [86, 137]; (2) **autonomy**, recognizing that people have different preferences about information disclosure (e.g., What information do they wish to disclose? For which purposes?), and should be able – have the capacity and mechanisms – to translate these preferences into concrete actions and decisions [77, 86, 126, 132]; (3) **inclusion**, providing people with meaningful opportunities to shape the data collection practices of which they are part [126, 157, 191]; (4) **transparency**, ensuring people have timely access to legible and relevant information on how personal data is collected, stored, and processed [126, 192, 199], and (5) **privacy**, ensuring people can protect their information from undesired disclosure.

More broadly, privacy relates to the (non) disclosure of personal information. Crabtree and colleagues [49] provided an overview of the various ways in which privacy is understood in HCI, including *privacy as control*, *privacy as boundary management*, and *privacy as contextual integrity*, among others. Privacy as control relates to the ability to control the flow of personal data through activities such as limiting information disclosure [208] and filtering what gets disclosed [169]. Privacy as boundary management is informed by the work of Irwin Altman [7], and relates to the selective disclosure of personal information as people move between privacy and publicity according to the context and intention [145]. Based on the boundary metaphore, Sandra Petronio proposes the Communication Privacy Management (CPM) theory, in which the disclosure of private information is based on privacy rules that are negotiated around personal and collective boundaries [149]. Finally, privacy as Contextual Integrity (CI) is a theory proposed by Helen Nissenbaum where privacy is understood in terms of the appropriateness of information flows according to social or cultural norms and grounded in specific contexts [137]. This thesis, which at its core relates to the (non) disclosure of personal digital-trace data, integrates and further explores these ethical challenges and privacy theories.

1.3. Research Aim

This thesis aims to develop an alternative approach to data donation informed by the principles of Data Feminism that empowers donors to know and care for their data and promotes different forms of knowledge and participation. Empowerment relates to the understanding and power donors have over their data and how power and value are distributed between donors and recipients. Participation relates to donors' involvement in the different research activities and how their involvement shapes their knowledge of their data, the data themselves, and the knowledge that is generated.

The alternative approach is grounded in existing forms of data donation. It departs from these approaches with a critique through the principles of Data Feminism. It challenges the perceived objectivity of digital-trace data and foregrounds multiple forms of knowledge, including people's embodied and situated knowledge. It recognises and addresses the knowledge gaps people might have over their digital-trace data and how these influence their ability to decide about their disclosure. It recognizes the context in which data is generated and how said context might influence its characteristics and the feelings around its disclosure. It strives to conduct research *with* people and their sensitive data as opposed to *on* people, or *on* their sensitive data. Moving away from engaging with people as contributors, who share or transfer data, to collaborators, who participate in interpreting the data, and co-creators, who participate in scoping the research. Thus, the main research question this thesis addresses is: **(RQ) How can data donation empower donors to engage with their data and promote their participation in design research?** It can be divided into five further research questions:

RQ1: How is data donation defined, operationalized and applied in scientific (design) research? With this question, I aim to identify and synthesize the relevant concepts and current practices in the data donation literature. Although data donation is an emerging practice, it has been discussed, developed, and applied across several disciplines including philosophy, psychology, health, social sciences and communication, and HCI. Delving into existing forms of data donation is essential to develop an alternative approach. What are the core concepts and practices in data donation? What are the open challenges in data donation research? How data donation should be approached in scientific research?

RQ2: How can data donation integrate people's embodied and situated knowledge(s)? With this question, I aim to incorporate the forms of knowledge that arise from people as "living, feeling bodies in the world" [51] into data donation. The "data" in data donation corresponds to digital-trace data. It contains detailed and structured information over time that is useful but limited. Especially in contexts where data is not seen as neutral or objective; but as prompt and support for subjective and situated inquiry. What types of knowledge(s) are excluded from data donation? How could these be embraced in data donation research?

RQ3: How can data donation foster an incremental understanding of data that invites donors to (re)assess their participation?
With this question, I aim to incorporate activities that support and promote ongoing consent practices into data donation. Informed consent requires that people understand how and why their data will be used and the advantages, disadvantages, and potential risks, associated with it [139]. Yet, since it is often hard for people to know their data [100] it is also hard for them to clearly understand how their data will be used, and even more so, the disadvantages and potential risks [140]. Can people consent to donate their sensitive data when they don't know them or their implications? How can people know their sensitive data and identify what (not) to disclose?

RQ4: How do donors perceive 'sensitivity' and 'intimacy' when faced with a comprehensive view of their data?
With this question, I aim to identify and characterize the sensitive and intimate aspects of personal data. Data donation often requires people to request and receive a copy of their data from data controllers (i.e., product and service providers). Previous research has demonstrated that this process is disempowering [4, 31]. Most data controllers do not fully comply, and when they do, the returned data are *"often difficult to understand, impractical to use, and raised new questions and concerns"* [31]. This poses a critical consideration for data donation research, which strives to provide people with agency and control over their sensitive data [24, 141, 204]. Can people act upon their sensitive data when they don't know them? What is sensitive about sensitive data?

RQ5: How can data donation promote different degrees of participation in research?
With this question, I aim to incorporate different degrees of participation into data donation, namely: contributors, collaborators, and co-creators [175]. The degree of participation of donors in data donation research is often limited, as they mainly contribute to a specific research project by transferring their sensitive data (e.g., [34, 204, 215]). Often these projects are relevant to them and their communities. Thus, their perspectives could be valuable to the research(ers). How else could people be involved in data donation research? How could people play a more active role in data donation research projects?

1.4. Research Approach

This thesis investigates how Data Feminism can inform an alternative data donation approach that empowers donors to know and care for their data and promotes different forms of knowledge and participation. The research follows a participatory Research through Design (RtD) approach [73]. I iteratively design and develop specific instances of a data donation process – embodied by a digital data donation platform[2].

---

[2] Data donation platform can be accessed through: `datadonation.ide.tudelft.nl`

These instances are then deployed in the field [103] and applied in a specific case study situated in an intimate context. Thus, enabling research participants (i.e., data donors) to experience a specific form of data donation in a context and reflect on their experience.

1.4.1. Intimate Contexts

In this thesis, I apply (an iteration of) the Sensitive Data Donation approach in three case studies, each situated in an intimate context mediated by data.

> **Case Study 1:** It aims to support people who track their menstrual cycle in interacting with the data they produce (i.e., menstrual tracking logs). Thus, it focuses on intimacy in relationship to the menstruating body and its entanglements with digital data. It is described in Chapter 3.
>
> **Case Study 2:** It investigates the perceived characteristics of data generated from people's interactions with their voice assistants (i.e., speech records). Thus, it focuses on intimacy around being at home and sharing it with connected appliances, specifically the Google voice assistant. It is described in Chapter 4 and Chapter 5.
>
> **Case Study 3:** It investigates athletes' perceptions of the impact of their menstrual cycle in sports through their data (i.e., physical activity logs). Thus, it focuses on intimacy around the relationship between people and their bodies and how these are (not) captured by data. It is described in Chapter 7.

Intimacy is a fluid concept [186]. It is related to other concepts such as privacy. What is intimate is often to be hidden and kept from others [91, 171]. Thus, it is non-public and private. It is also a gendered term [61] that is associated with female codes and characteristics, and often used as a euphemism for sex. Intimacy is related to closeness in relationships between different groups (e.g., couples, families, friends) and the relationships between people and technology. It also concerns the relationships between people and their bodies, which are many times taboo, and how these manifest through looking and touching [6, 38, 185]. As a result, intimacy usually happens in the home, an intimate space [108, 109, 146]. Intimacy also permeates our data as digital products and services that collect it enter our homes and intervene in our relationships with partners(s), family, friends, and our bodies. Data is close to us. In turn, we can get close(r) to our data and build intimacy with it. It entails understanding it as situated and entangled with ourselves, knowing what it is and what it means (i.e., implications), and trusting it [56], which are prerequisites for gaining agency and control over it.

Intimate contexts are inherently private and situated spaces. They are foundational spaces in which to explore and negotiate people's relationships with their data, and whose ethics and practices could be applied more broadly. Data donation is then a gateway to these spaces and could potentially be a bridge to getting close(r) to the data and building intimacy with them. Either way, data donation leads to vulnerability. Therefore, intimate contexts require a data donation approach that acknowledges and mitigates this vulnerability, where data is handled with care [129] and approached not as neutral or objective but as situated and contextualized [55, 90, 108].

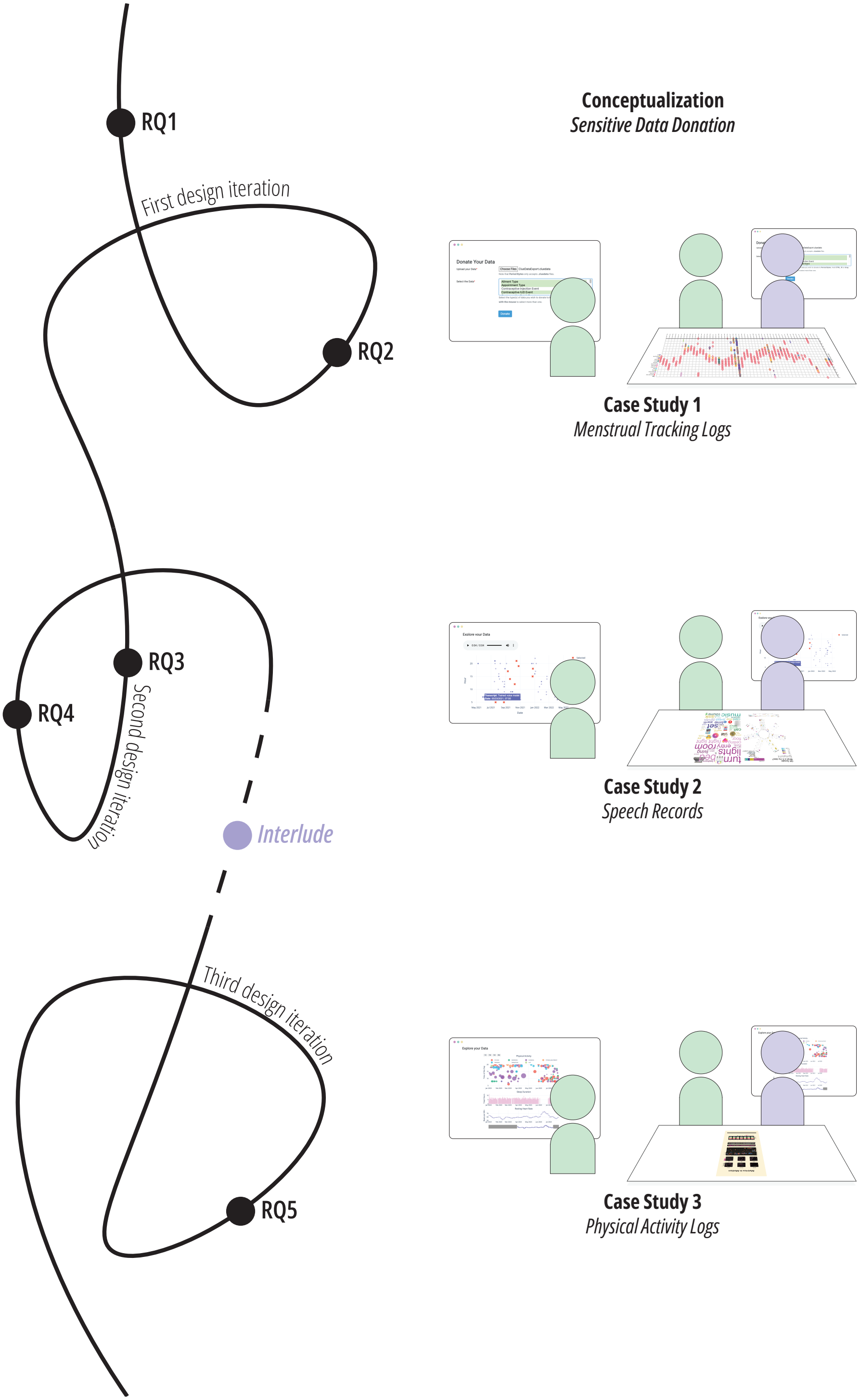


Figure 1.2: Research approach, describing the theoretical framing of the thesis and the resulting five principles that inform the design and development process followed by three design iterations.

1.4.2. Research through Design

This thesis follows an iterative Research through Design (RtD) approach illustrated in Figure 1.2. It starts with a conceptualization of Sensitive Data Donation, including the five principles and five-phase approach. It continues with an iterative process of designing and developing specific instances of a Sensitive Data Donation approach – embodied by a digital data donation platform – and deploying them in what Koskinen et al. [103] refer to as the ‘field’, a 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 represented with a loop in Figure. 1.2.

In RtD design actions and design activities play a formative role in the generation of knowledge [73]. 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 [216]. 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 [17, 73]. Artifacts are the concrete embodiment of a specific concept or idea and are a way to share and communicate it with others [73, 216]. They are shaped by design decisions made to represent said concept or idea [73]. Further, they are shaped by the technical opportunities and constraints around “making” them [73]. Thus, artifacts allow researchers to derive knowledge from “making” them as well as sharing them with others.

In this thesis, the artifacts correspond to various instances of a Sensitive Data Donation approach embodied by a digital data donation platform. 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. Along the iterative RtD 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 [149]) 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 [73]. 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.

I employ qualitative methods to analyze people’s experiences with the specific instance of the artifact and give voice to them on the subsequent designs. Given that my focus is on people’s experiences, this thesis is grounded in a relativist ontology and a constructionism epistemology. This means that I believe knowledge cannot be separated from individual contexts and experiences and it can be created or constructed through the process of research. Overall, the knowledge constructed in this thesis corresponds to intermediate-level knowledge [96]. It sits between specific instances (e.g., a particular artifact) and generalized theories. Specifically, the intermediate-level knowledge (Figure. 1.3) corresponds to methods and tools, guidelines, and principles.

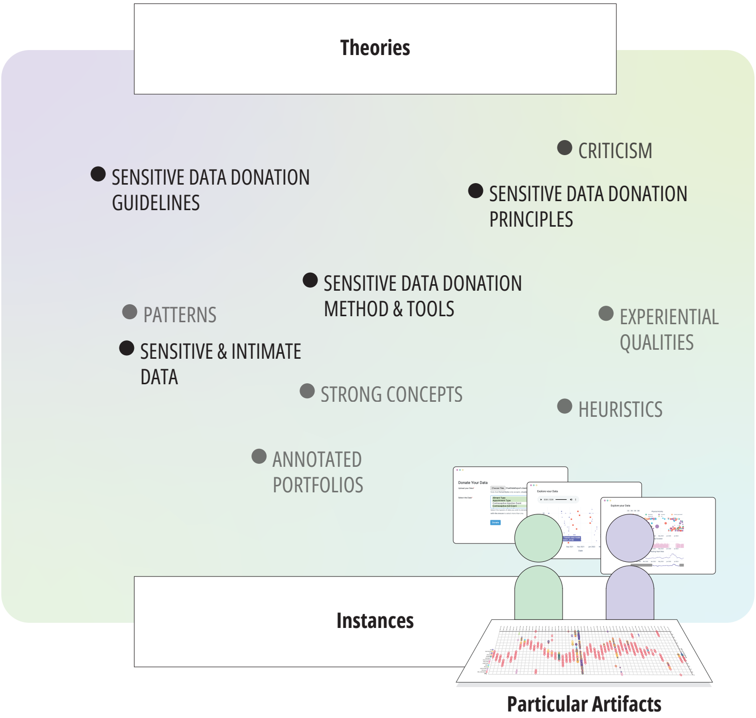


Figure 1.3: Adaptation of model by Höök and Löwgren [96] that highlights the type of intermediate-level knowledge developed in this thesis in **bold** and other types of intermediate-level knowledge in grey.

1.5. Research Contributions

The main scientific contribution of this thesis is a conceptual framework of **Sensitive Data Donation** describing an alternative approach to data donation in intimate research contexts informed by the principles of Data Feminism. It comprises five principles integrated into a five-phase approach and operationalized through a digital data donation platform. It is built on the following other scientific contributions:

- A *conceptual framework of data donation*, that synthesizes the multidisciplinary data donation literature, illustrating the concepts and actors involved in data donation.

- An *empirical understanding* of the content and perceived characteristics of digital-trace data: sensitivity and intimacy.

- A series of *empirical findings* that are relevant in the context of each case study. Including insights into how people who menstruate interact with the intimate data collected through menstrual tracking apps and cyclical metrics for athletes to reflect on how the menstrual cycle impacts their physical activity.

- A provocative artifact, *dataslip*, that aims to raise awareness around personal data collection. It materializes the personal data trail in the form of a receipt that is printed slowly and elicits creepiness.

1.6. Thesis Structure

Chapter 2 describes the key concepts and the current practices in the data donation literature, synthesises these into a conceptual framework, and introduces a critique through the principles of Data Feminism [51]. The critique informs a first approximation of the Sensitive Data Donation principles and the Sensitive Data Donation approach. **Chapter 3** initiates the artifact design and development process, by translating the Sensitive Data Donation principles into the first iteration of a data donation platform. This chapter focuses on incorporating into data donation activities that support donors in situating and contextualizing their data. **Chapter 4** continues the artifact design and development process, by enhancing the data donation platform with an interactive tool where potential donors can explore and curate their data. This chapter focuses on supporting donors in incrementally engaging with their data and (re)assessing their participation. **Chapter 5** delves into the data and its perceived characteristics; specifically, it focuses on 'sensitivity' and 'intimacy'. **Chapter 6** is a brief ***interlude***. It steps outside the core data donation research and introduces *dataslip;* a provocative artifact that aims to promote data awareness beyond data donation by materializing the data and eliciting creepiness. **Chapter 7** concludes the artifact design and development process, by enhancing the data donation platform with an interactive tool where potential donors can explore and curate different types of data locally; before they leave their devices. This chapter focuses on supporting different degrees of participation, namely contributors, co-creators, and collaborators, through data donation. Finally, **Chapter 8** summarises the findings of this thesis, consolidates these into the Sensitive Data Donation framework and the Sensitive Data Donation method, and highlights the scientific and societal contributions.

How is data donation defined, operationalized, and applied in scientific (design) research?

2

Interrogating Recent Approaches to Data Donation

Abstract

In this chapter, I investigate how data donation is defined, operationalized, and applied in scientific research (RQ1). To do so, I interrogate recent forms of data donation through a scoping review of the data donation literature in the last eight years. I focus on key concepts and current practices to develop a conceptual framework for data donation. I then interrogate this framework through the principles of Data Feminism and propose a critique. It raises questions about different aspects of data donation, including: Who benefits and who is neglected? What forms of knowledge are valued? Who gets to know what about data? Who participates and how? From this critique, I propose an alternative approach called Sensitive Data Donation. It comprises a set of five substantive principles and a five-phase approach – these will be applied in the following chapters.

This chapter draws on the following publication: **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2024) 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).

2.1. Introduction

I mentioned earlier how this thesis aims to develop an alternative approach to data donation informed by the principles of Data Feminism. The word *alternative* is key. It denotes other possible forms of data donation from which to depart. In this chapter, I interrogate and position this thesis with respect to these other forms of data donation. Data donation has been discussed and approached from multiple disciplines including philosophy of technology, psychology, health, social sciences and communication, computer science, and human-computer interaction. These disciplines have different and somewhat overlapping foci; from discussing the ethics of data donation (e.g., [97, 100, 106]) to investigating factors that motivate potential donors (e.g., [153, 165, 179]) or developing methods and tools to collect digital-trace data through data donation [24, 34, 140]).

I delve into these various foci by systematically reviewing the data donation literature from the last eight years with a focus on interrogating how data donation is defined, operationalized, and applied. I identify and integrate the central concepts and practices into a conceptual framework of data donation. It is the theoretical grounding from which I depart to propose an alternative. The word alternative also denotes a change or a challenge. Thus, I propose a critique of these other forms of data donation through the principles of Data Feminism [51]; which raise important questions about their ethos and ethics. The critique informs a first approximation to an alternative: Sensitive Data Donation. I conclude this chapter by introducing the substantive principles of Sensitive Data Donation and integrating them into a five-phase approach. These inform the artifact design and development process in the subsequent chapters.

2.2. Conceptual Framework Analysis

I identify and synthesize into a framework the main concepts and current practices in the data donation literature across multiple disciplines, including philosophy, psychology, health, social sciences and communication, and HCI. I used the conceptual framework analysis proposed by Jabareen [99], describing how conceptual frameworks can be constructed through an iterative qualitative analysis of existing literature, starting from a scoping review of multidisciplinary texts. The resulting framework is visualized in Figure 2.2.

2.2.1. Systematic Literature Review

I conducted a scoping review of the data donation literature to investigate how data donation is defined, operationalized, and applied in scientific research and what the degree of donor participation is. The review followed the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [197], structured in four phases (Fig. 2.1). I searched the Scopus database as it is useful for conducting literature reviews across multiple disciplines [158]. I restricted the search from 2015 to 2023 given that data donation was first proposed around 2015 [98, 174] and further developed with the introduction of the GDPR in 2018. Using Scopus, I searched for journal articles and conference papers mentioning “data donation” in their title, abstract, or keywords. I focused on scientific research, published in venues following peer-review processes, therefore, did not expand the search to include grey literature. I refrained from using synonyms for “donation” such as “sharing”, as the term “donation” implies specific motivations and expectations that are different than those of other terms [97], while personal data donation appeals to values such as altruism and solidarity and it is often situated within a research context, personal data sharing sits more broadly. It refers to how personal data is shared with other parties – including people [194] and organizations [167] – online and offline. Nonetheless, I initially sampled literature from terms such as “sharing” but found it too broad and excluded it.

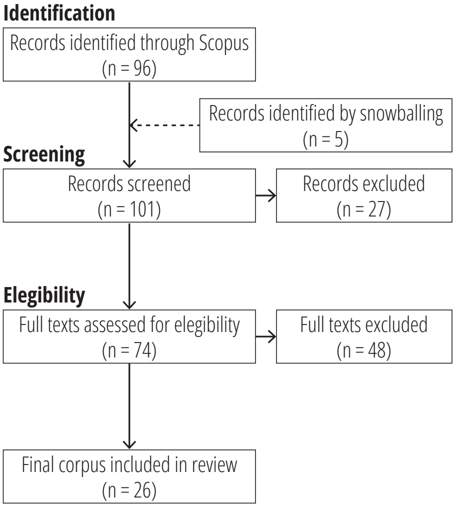


Figure 2.1: PRISMA-ScR flow of information through the different phases of the systematic literature review

I identified 96 records through Scopus and five more by screening the references (i.e., backward snowballing) of the initial set of papers. I screened the initial set of 101 papers, reading the title and abstract. Papers were excluded when they described other forms of donation (e.g., organs or tissue). Next, I assessed the eligibility of the remaining set based on the full text. Papers were included when they explicitly stated a theoretical framing or empirical approach contributing to data donation in the abstract, keywords, introduction, contribution statement, or conclusion. They were excluded when (1) they described or discussed data donation after death (i.e., posthumous medical data donation), or (2) they specifically situated data donation in the context of a hospital – where data corresponds to patient data and consent is obtained in clinical care situations and mediated by healthcare professionals [165]. I read the remaining papers and analyzed them based on the exclusion criteria.

This led to a final corpus of 24[1], focusing on (1) theoretical frameworks, methods, and (best) practices [22, 24, 88, 97, 100, 105, 136, 140, 141, 155, 190, 210] (2) motivations and willingness to donate data [93, 116, 151, 179, 181], (3) empirical approaches apply-

---

[1] In the paper, the literature review included two of my papers on data donation, I removed them from the report on this thesis.

ing data donation [18, 34, 142, 162, 204], and (4) systems and platforms enabling data donation [13, 25].

Following the conceptual framework analysis procedure [99], I read each paper, categorized them by discipline, and identified relevant concepts. Next, I iteratively grouped similar concepts into a new concept, with a unique name and description. The new concepts were iteratively discussed and synthesized with a co-author of the study, Gerd Kortuem. Moreover, I examined all papers in our corpus and clustered them based on their main contribution and how they defined, operationalized, or applied data donation. Further, papers applying data donation were classified according to the degree of participation (i.e., contributors, collaborators, co-creators [175]).

2.2.2. Defining, Operationalizing, and Applying Data Donation

Defining Data Donation

Data donation emerges from the availability of personal data collected from connected products and services [25, 181]. It is conceived as a voluntary act with an altruistic nature and the purpose of contributing to scientific research [97, 155, 190], especially in the healthcare domain. Further, it is considered a powerful method for data collection [25, 140] and a meaningful alternative to approaches enabling researchers to access data; such as Application Programming Interfaces (APIs) [34, 141] and dedicated tracking technologies [141].

Data donation enables researchers to collect private personal data at the individual level. This is a critical distinction between data donation and other approaches to data collection from which researchers often access public data (e.g. World Health Organization API) [34, 141, 204]. Data is considered private as it is generated and captured in (digital) private spaces (e.g., direct messages on Facebook) in contrast to the public data, captured in (digital) public spaces (e.g., aggregated health indicators). It is considered individual-level as researchers request and access it directly from individuals – who opt-in and consent to their participation [34, 141, 179].

Therefore, ethical considerations are critical in data donation and addressed in several studies. The main consideration is informed consent [24, 97, 100, 107, 140] and how donors can exert their autonomy [88, 97, 100, 107, 155] and preferences (e.g., deciding whether/ what/ and to whom to donate [190]). Further, it concerns ethical aspects related to the relationship between donors and their data and between donors and recipients. The former is shaped by data, which is opaque and potentially sensitive or invasive [97, 100, 136], introducing the challenge of understanding data [100] and other considerations, such as data minimization [25, 140, 141], mitigating harm [22, 155], and uncertainties around future use [97, 136]. The latter requires researchers to be transparent, provide sufficient information [97, 100, 155, 190, 210], and honor donors' contributions [107, 155]. Ethical considerations also concern potential donors. Studies focused on the motivations and willingness to donate data found that potential donors worry about their privacy, the perceived sensitivity of their data, and its possible misuse [93, 151]. These worries relate to trust in the data recipients and if and how trust is maintained [93, 116, 179, 181].

Operationalizing Data Donation

Several data donation infrastructures have been developed. These mainly constitute digital platforms to which donors can upload a copy of their data. Araujo et al. [13] proposed OSD2F where participants can see the content of their data in tabular form and select the data they (do not) want to donate. Likewise, Boeschoten et al. [24] developed PORT, where data is processed locally and presented to participants in tabular form before they consent to donating them. Studies focused on applying data donation have developed similar platforms. For instance, Razi et al. [162] designed a digital platform where participants can upload their Instagram data.

Applying Data Donation

Studies applying data donation employ two approaches: (1) digital platforms or repositories where donors upload a copy of their data, previously requested by a data controller [18, 34, 142, 162, 204], and (2) applications where donors consent to scrape data using their account(s) [34, 166]. Breuer et al. [34] applied and compared the two approaches using Facebook data and discussed the legal, privacy, and ethical implications. They concluded that the first offers higher transparency for donors but requires more effort and that both result in the collection of sensitive data. Razi et al. [162] and van Driel et al. [204] applied the first approach to collect Instagram data from teens, they instructed participants to upload their data to a digital platform and a repository respectively. Furthermore, Razi et al. [162] asked donors to annotate their data (i.e., private conversations on Instagram) by flagging them as ‘safe’ or ‘unsafe’. Finally, Ohme and Araujo [140] and Baumgartner et al. [18] applied the first approach to collect iOS screen time data, they required donors to upload a screenshot and screen recording respectively and not the data themselves.

Donors’ degree of participation [175] is contributory in most studies: they contribute to a project by transferring their data, and might further contribute by augmenting or annotating it. For instance, Breuer et al. [34] integrated a survey on privacy concerns, and digital habits into their Facebook data donation approach, while Razi et al. [162] asked donors to annotate their private conversations on Instagram. None of the studies in our corpus included donors participating as collaborators or co-creators.

2.2.3. Conceptual Framework of Data Donation

The conceptual framework includes the actors involved in data donation (in **bold**), the key concepts (in *italics*), and the relationships among them, mapped visually in Figure 2.2. It comprises a practical layer, describing concepts related to the implementation of data donation, and a conceptual layer, describing theoretical concepts.

Based on the literature review, I describe data donation as the voluntary and *consensual* non-reciprocal *transaction* of personal **data** [22, 155, 179]; made directly from an individual (i.e., **data donor**) to a person or institution, (i.e., **data recipient**). Data donations are *relational* transactions; as they strengthen or change the relationship between donors and recipients [155]. They operate in a specific context around a *research project*. Researchers initiate data donations by defining the data needs of their project [34, 141] and inviting individuals, from the general population or a specific community, to con-

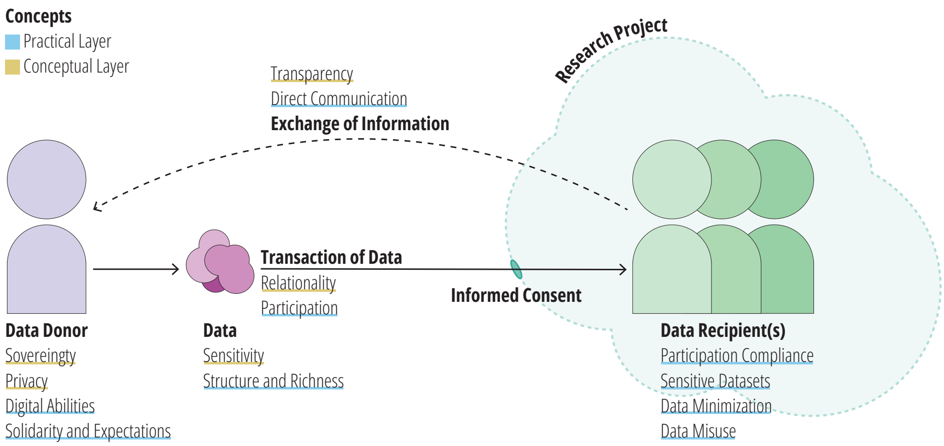


Figure 2.2: Conceptual framework of data donation

tribute by donating their data. Here, there is *direct* [93] and *transparent* [13, 34, 93, 116, 142, 210] *communication* between researchers and potential donors. Meaning, that researchers openly convey the purposes for which the data is requested, how the data will be collected and handled, and insights gained from the data that may be relevant to donors; which in turn builds trust [117]. Moreover, researchers must ensure that the personal data they request and receive is relevant to answer a specific research question (i.e., *data minimization* [34, 142]), set limits on its (future) use [97, 136, 155], and avoid *misuse* [24, 136, 140].

Donors respond to the invitation by transferring their data; in doing so, they exercise *data sovereignty* (i.e., the power to control their data) [93, 97, 140–142] and actively *participate* in scientific research [22, 97]. Moreover, when contributing to a research project, donors exert *solidarity* [97], recognizing sameness or similarity in the context of a project that is relevant for them [151], and *expect* something in return, not reciprocity, but knowledge, empowerment, recognition, or some type of benefit that derives from the research [22, 88, 155]. Additionally, they ponder their *privacy* concerns, their relationship with their data, and its perceived *sensitivity*. The individual-level private data that is available through data donation potentially contains sensitive information about donors and their relationship(s) to others [18, 24, 34, 97, 100, 140, 141, 151] and includes *rich* retrospective information that is timestamped and *well-structured* [24, 141, 204]. Thus, *privacy* is an important concept in data donation [13, 100, 210]; so much so that privacy concerns are a predictor of the willingness to donate personal data [93, 151]. The *digital abilities* of data donors are another predictor of the willingness to donate personal data [93], as they are essential for them to successfully navigate the various systems and platforms required to transfer the data to a data recipient [13, 34, 140, 142, 156]. Once data is transferred and consolidated into a dataset, it retains its *sensitivity* and non-publicity; as such it should be treated differently than public and open datasets [24, 140, 162].

2.3. Critique: Data Donation through a Data Feminism Lens

Current forms of data donation are insufficient in **intimate contexts** that reject the notion of data as neutral and objective and instead aim to account for its *situatedness*. These engage with data as disconnected from the specific contexts (i.e., times, places, bodies, devices) where it is generated and shaped and tend to reduce a complex experience to its digital traces [121]. Feminist scholar Donna Haraway [90] introduces the notion of *situated knowledges* to illustrate how all forms of knowledge, including data, are produced by specific people in specific circumstances. Understanding data as situated enables those whose data is about to interpret, contextualize, and contest it; outlining aspects that are invisible with only the data [75, 108, 195] and that are currently disregarded in data donation. Building upon feminist scholarship, including Haraway's situated knowledge, D'Ignazio and Klein [51] propose Data Feminism as a way to interrogate, challenge, and change certain practices in data science that reinforce existing inequalities. They introduce seven principles: (1) Examine Power, (2) Challenge Power, (3) Elevate Emotion and Embodiment, (4) Rethink Binaries and Hierarchies, (5) Embrace Pluralism, (6) Consider Context, and (7) Make Labor Visible, these are described in Chapter 1. The principles of Data Feminism raise important questions about the ethos and ethics of current data donation practices. Such as: Who benefits from data donation and who is neglected? Who participates and how, people or their data? What forms of knowledge and data are prioritized? In the following, I explore these questions.

2.3.1. Power and Value Asymmetries in Data Donation

In most data donation research donors participate as contributors (Section 2.2), illustrating the power asymmetries between donors and researchers, who conduct research *on* their data instead of *with* them. Donors, as contributors, are a source of data in a project that is shaped by the researchers; whereas they could be collaborators and co-creators in a shared project that is relevant to their community and shaped from within. These asymmetries are reflected in Figure 2.2 with the unidirectional exchange of information between researchers and data donors; as they directly communicate the project goals and the participation procedure [13, 34, 93, 116, 142, 210]. Researchers define the project and participation criteria, and donors comply or not.

Similarly, there is a value gain asymmetry. Researchers clearly gain value from data donation, they gain access to data and through it are able to advance their research. It is less clear how data donors can gain a similar value [75]. Conceptually, data donations are non-reciprocal transactions [155]. Meaning, they are not a direct exchange (i.e., data is not exchanged for another thing), and donors are not to expect something directly in return. Yet, when donors contribute to a research project through data donation, they often perceive it as relevant to them or someone close to them [97, 151], and their contribution implies labor and effort [74, 141, 155]. They generally expect to benefit in some way, for instance from the use of the data in the project [116]. In practice, the most common way that donors benefit from their participation is through financial compensation (e.g., [18, 34, 140, 204]); which contradicts the charitable nature of data donation. This can lead to ethical issues such as coercion [147] and limit donors' ability to voluntarily offer consent [201]. More broadly, compensation is not considered a benefit derived from participation but an incentive for participation compliance [82, 201], or the willingness to follow the procedure related to participation, which is low in data donation and similar approaches requiring a transaction of digital trace data [18, 24, 142, 204].

For instance, to encourage data donation, [215] implement a system where TikTok users can upload a copy of their data. They offer donors financial compensation based on the types of data they choose to donate. It starts at 5$ for the video viewing history (mandatory) and increases by 1$ for each additional data (e.g., like history, share history, app settings, comments, followers). They claim that in this way they give *"power"*, or sovereignty, to data donors – the majority of which (95%) chose to donate *"almost all the data"* [215]. I argue that this specific implementation of data donation is problematic. Donors get a choice but that choice is shaped by the researchers and not choosing (i.e., donating all data) is strongly incentivized; thus influencing the voluntariness of the transaction and placating potential concerns such as privacy. It further illustrates the power, and value gain asymmetries in data donation. Moreover, it treats data as a commodity, failing to consider that *"data are people"* [217]; meaning data are intertwined with individuals, their behavior, and social relationships [51, 76]. For instance, TikTok viewing history could potentially reflect a person's views and interests while her followers and comments could account for her interactions with friends and colleagues; are these worth 7$? Would $7 represent a similar value to donors as a publication would to researchers?

2.3.2. Knowledge Asymmetries in Data Donation

In Figure 2.2, I highlighted how informed consent is treated as a single moment and static decision in data donation. It generally takes place when donors agree to participate in the research (e.g., [162, 204]) or when they transfer their data (e.g., [34, 140, 215]). Previous research extensively describes the shortcomings around the static nature of informed consent [37, 101, 189], especially considering the sensitive and obscure nature of personal digital trace data [76, 217]. Data donation research has recognized these limitations and poses meaningfully informed consent as one of its great challenges [93, 100, 140, 141, 151, 155, 190, 204, 210]. They have approached it mainly by offering data donors a choice over which data to donate before they transfer it [2, 25, 215]. For instance, PORT [25] extracts the relevant data (e.g., visited places) from the files uploaded by potential donors (e.g., location history from Google Takeout) and displays an overview in the form of a table for them to approve (e.g., Date: January 2020, Number of Places: 24).

Nonetheless, at the moment of informed consent, data donors often *"don't know what they don't know"* [100] about their data and are unable to fully understand the implications of sharing it and to define clear boundaries around it [9]. The individual-level data that is transferred through data donation is not only sensitive data [18, 34, 141, 151], but it has opaque and has unknown sensitive elements [76, 100, 141]. Thus, sensitivity is a characteristic of personal data that is not known upfront and needs to be discovered and disentangled from the data through a process of exploration and interpretation [76]. This process is not compatible with data donation approaches that conduct research *on* donors, or their data, instead of *with* them. Integrating such a process into data donation entails embedding informed consent into an ongoing, dynamic, and affirmative process [37, 101, 189].

2.3.3. Emotion, Embodiment, and Context in Data Donation

The forms of data that constitute a data donation project are primarily digital traces, generated or collected in the wild through people's interactions with digital products and services (Section 2.2). Recent data donation projects have focused on social media data from Facebook [34], Instagram [162, 204], TikTok [215], and YouTube [2], and data from wearable devices [166], connected appliances [74], and digital apps [18, 75, 142]. These private individual-level data have several qualities that render them valuable for scientific research, they are structured, spatiotemporal, and rich. Nonetheless, they are decontextualized and unable to adequately reflect people's behavior, feelings, or experiences [195]; although they serve and are often used as a proxy for these.

I mentioned earlier how *"data are people"* [217]; referring to how the information encoded in data is deeply related to people and it is difficult to disassociate data from a specific individual. For instance, a person's interactions with a voice assistant can illustrate her morning routine, interests, and concerns [76]. However, people are *not* data. Their experiences, feelings, and contexts are not entirely captured by their data. It offers a glimpse. It is limited, decontextualized, open for interpretation, and potentially wrong or incomplete [75, 108, 195]. By narrowly focusing on digital trace data, current forms of data donation exclude other forms of knowledge, especially people's embodied and situated knowledge [51, 90], and other forms of participation, such as collaboration and co-creation. Data donation research has begun to address this limitation by integrating ways for people to augment, annotate, and contextualize the data; including surveys [34] and manual annotations [162].

2.4. Sensitive Data Donation

The principles of Data Feminism raise important questions about the ethos and ethics of recent data donation practices. For instance, how current forms of data donation often neglect donors' expectations and efforts, or how these prioritize the researchers' perspectives, biases, and values. While current data donation practices often conduct research *on* people's sensitive data, Data Feminism highlights the importance of conducting research *with* people and their sensitive data – including a broad range of perspectives and participants and recognizing the specific contexts (i.e., times, places, bodies, devices) where data is generated. In the following, I investigate how Data Feminism can inform an alternative form of data donation that rejects the notion of data as neutral and objective and assumes the importance of balanced participation, context, and pluralism. Specifically, I propose the Sensitive Data Donation method, which comprises five substantive principles, arising from the critique (Section 2.3), and a five-phase approach incorporating procedural principles related to ethical data (donation) practices.

2.4.1. Substantive Principles of Sensitive Data Donation

(P1) **Balanced Value:** This principle calls for recognizing and honoring donors' contributions and efforts by intentionally integrating activities into the data donation process that allows them to derive value. These activities could include acknowledgment of donors by name wherever appropriate [155], dedicated learning activities for donors [75], and new knowledge and empowerment derived through data exploration resulting in data not being *"about donors but theirs"* [74]. It invites the questions: **What contributions will donors make to the research project? What would be a fair benefit to donors?** It derives from the "who" questions posed by Data Feminism around power (i.e., Who benefits and is neglected? Whose priorities are prioritized and overlooked?) Currently, recipients are in charge of setting up the research and infrastructure, while donors are involved in activities that enable them to transfer their data. There is labor and effort on both sides, and both should derive a similar value from their contribution [76, 107, 155].

(P2) **Sensitive Data:** This principle calls for recognizing the sensitive nature of the data as a critical prerequisite for balanced and informed participation. Here, sensitivity is understood as potentially sensitive information that is unknown in the data [75, 76, 100] and as sensitive private information that must be protected and not disclosed [24, 74, 141, 204]. It invites the question: **How can donors know their sensitive data and draw clear boundaries around its disclosure?** It derives from shortcomings of current approaches that fail to consider the practicalities of donating sensitive data. Data donation frequently requires people to request and receive a copy of their sensitive data from data controllers (Section 2.2.2). Previous research has demonstrated that this process is *"disempowering"* [4, 31] as most data controllers do not fully comply with access and data portability requests, and when they do, the returned data is *"often difficult to understand, impractical to use, and raised new questions and concerns"* [31]. This reiterates how at the time of donation – and informed consent – donors *"don't know what they don't know"* [100] about their data and its sensitivity and must be encouraged and supported to do so.

(P3) **Multiple Knowledge:** This principle calls for involving donors in interpreting and contextualizing their data; prioritizing their embodied and situated knowledge. It builds upon previous research that strives to include donors in these activities [108, 195, 203]. It invites the question: **How can donors participate in interpreting, contextualizing, and situating their data?** It derives from the third principle of Data Feminism, proposing to value multiple forms of knowledge. Especially in intimate research contexts where data is not seen as neutral or objective; but as prompt and support for subjective and situated inquiry. For instance, Tolmie et al. [195] demonstrated how the assumption that people's everyday interactions could be "read off" from sensor data inside their homes is flawed; they argue interpretation and contextualization are key when relying on data to reconstruct an otherwise incomplete human experience.

(P4) **Ongoing Consent:** This principle calls for embedding informed consent as an ongoing incremental process that accounts for donors' preferences regarding participation and disclosure of sensitive information over time. It invites the question: **How can donors be empowered to increasingly know their sensitive data and identify whether and what information to disclose?** It derives from the obscure nature of digital-trace data [76, 97, 100] and the need to promote data sovereignty and meaningful choice in data donation [24, 97, 140]. Explicit informed consent is an ethical and legal requirement for research involving human participants and a prerequisite for processing sensitive data under the GDPR [65, Art. 9]. It requires that donors understand how and why their data will be used and the advantages, disadvantages, and potential risks, associated with it [139], which is challenging due to the nature of the data and the potential sensitive information it contains or could reveal [74, 100]. Thus, it must be approached as an ongoing process and not as a single instance [74, 101].

(P5) **Shared Goals:** This principle calls for supporting different degrees of participation in data donation and inviting interested potential donors to relate to and shape the research project and goals from the start. It invites the question: **How could donors play a more active role in data donation research projects?** It derives from the limited participation of donors in data donation research (Section 2.2.2) and how it bounds the questions and contexts that are investigated and prioritized. These are often relevant to individual donors and their communities. Thus, their perspectives could be valuable to the research(ers) and should be included.

In Table 2.1, I summarize these five principles and contrast them with current approaches to data donation.

	Recent Approaches	*Sensitive Data Donation*
Data donor(s)	Participate by contributing their digital trace data and contribute additional self-reported data through surveys and annotations.	Participate by contributing their data and can further participate as collaborators, situating and interpreting the data, and co-creators, scoping the research questions and goals.
Sensitive data	Decontextualized digital trace data often annotated or augmented with additional self-reported data.	Situated and contextualized digital trace data enriched with embodied knowledge derived form collaboration.
Transaction	Transactional, often incentivized through financial compensation.	Relational, promoting a similar value gain that harnesses the relevance of the research and relatedness to a community.
Informed Consent	Static and treated as a single moment.	Dynamic and treated as an ongoing process that promotes an incremental understanding of data and accounts for its sensitive nature.
Research Process	Defined by the data recipients, prescribed and linear. One-directional information flow.	Defined by the data recipients in collaboration with a community of potential donors, open-ended and iterative. Bi-directional information flow.

Table 2.1: Comparison between recent data donation approaches and the sDD method.

The principles of Sensitive Data Donation build upon principles and practices described in the data donation literature. These include Jones' [100] incongruities around data awareness, underlining how donors *"don't know what they don't know"*, approaches inviting donors to interpret and contextualize data [74, 75], implementations of data donation platforms and systems [13, 25, 162], and legal and ethical considerations [34, 97, 140]. Contrasting with recent approaches, it stems from and advocates for a different way of thinking about data. When applying these principles, the goal is not to achieve ecological validity [162, 204] or generalizability [24, 190] but to foreground the content and characteristics of data and how it relates to individuals' knowledge and experiences in a given context. Thus, building up on the conceptual framework of data donation (Fig. 2.2), I synthesize the five principles into a conceptual framework of Sensitive Data Donation (Fig. 2.3).

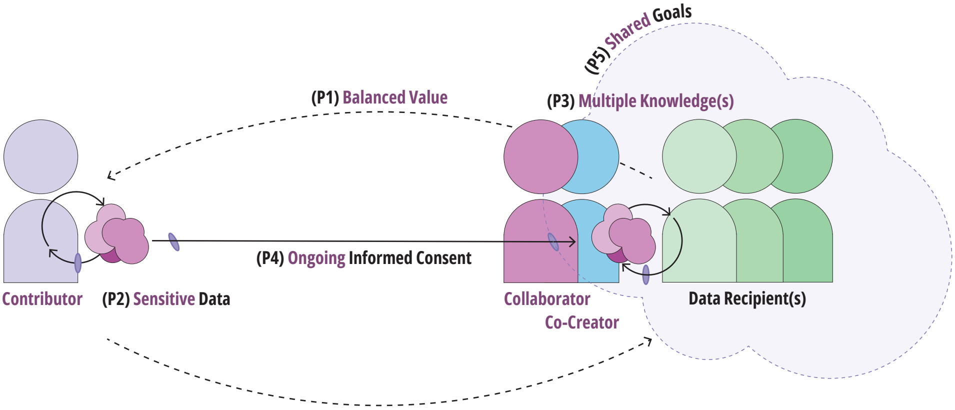


Figure 2.3: Conceptual framework of Sensitive Data Donation

2.4.2. Sensitive Data Donation Approach

I integrate the five substantive principles into a five-phase approach, represented visually in Figure 2.4. These phases stem from how previous research has approached digital-trace data donation (e.g., [24, 34, 162, 204]) and how the substantive principles can be implemented. Additionally, they integrate procedural principles derived from previous literature around ethical data (donation) practices.

1. **Identify, prepare, and communicate:** Researchers and potential donors co-create and scope the research questions and goals and co-define the value-gain strategy, data needs, and how these can be flexible enough to suit individual preferences. This requires researchers to explicitly consider how potential donors can derive value from their participation and change how they conceive the data needs of a given project: from not receiving donations from individuals who decide to partly share their data [24] to defining practices that support data sovereignty and preparing to use partial and messy datasets. Additionally, researchers invite potential donors to participate in the research. Here, I incorporate the procedural principle of **inclusion** by allowing potential donors to shape the research process and incorporate their perspectives [127, 157].

2. **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. This requires researchers to become

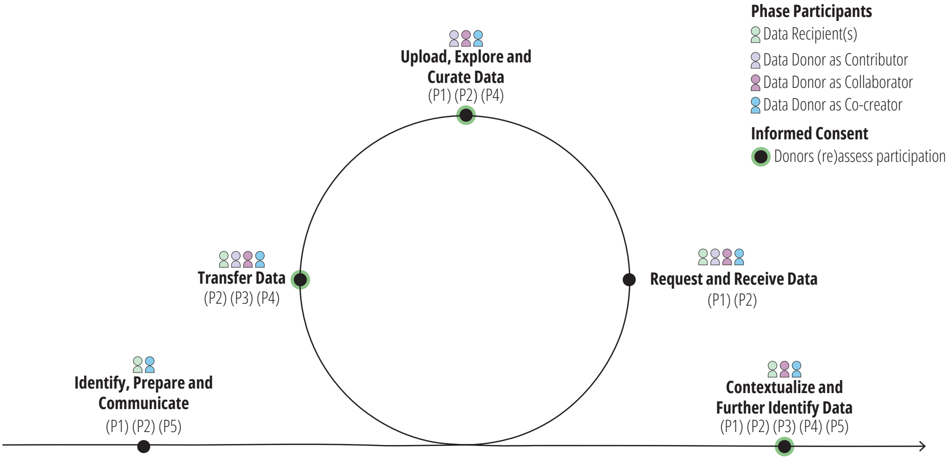


Figure 2.4: Overview of the five phases of the sDD method and involvement across the different degrees of participation.

familiar with the practicalities of the process and be available to guide and support as well as to clarify questions about the research process and goals. Here, I incorporate the procedural principle of **transparency** by clearly communicating why and how data will be used, and the benefits and risks associated with participation [127, 157, 199, 204].

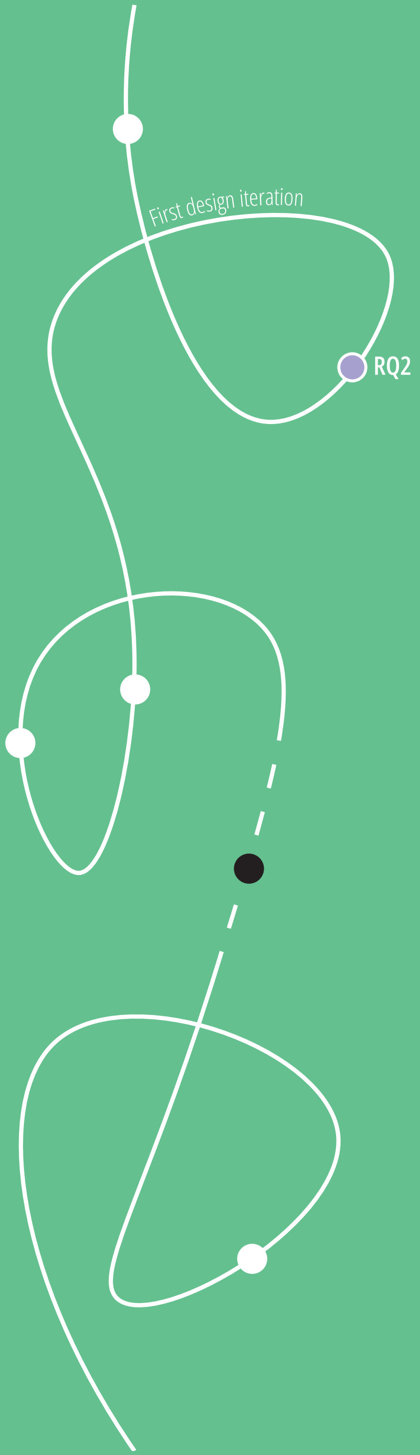
3. **Upload, explore, and curate data:** Researchers support potential donors to autonomously explore and draw boundaries around their data and the information they wish to disclose *before* they disclose it. This requires researchers to facilitate potential donors' interaction with their data even before they become research participants by making the (opaque) data visible, inspectable, and easy to understand and manipulate. In doing so, they encourage and support donors to engage with the content and characteristics of their data and reflect on what is sensitive about them. Here, I incorporate the procedural principle of **privacy** as boundary management [149] and contextual integrity [137] by explicitly inviting individuals to define where sensitive information resides and limiting its disclosure.

4. **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. Additionally, they decide on their preferred degree of participation (i.e., contributors, collaborators, and co-creators [175]). This requires researchers to delay the first moment of informed consent and the actual data donation until donors have explored the data. Here, I incorporate the procedural principle of **autonomy** by explicitly enabling individuals to determine whether and what data they wish to donate and how they wish to participate [75, 97, 127].

5. **Contextualize and further identify data:** Donors participating as collaborators and co-creators are invited to interpret and contextualize their data with the researchers, (re)evaluate their participation (i.e., confirm or revoke consent), and further scope the research questions and directions. This requires researchers to prepare and represent the data as a tool to elicit and invite multiple forms of knowledge and to communicate the progress and directions of the research. Here, I incorporate the procedural principle of **accountability**: researchers ought to inform donors about the process, progress, and limitations of the research, and deliver on the expected ways to provide value [75, 126, 131, 157].

2.5. Chapter Takeaways

The five principles of Sensitive Data donation constitute the theoretical grounding of this thesis. They steer the artifact design and development process. In the following chapters, I incorporate these principles into a data donation process embodied by a digital data donation platform and apply them to intimate contexts – where participants can experience and reflect upon a specific form of data donation. The first two principles, namely Sensitive Data **(P1)** and Balanced Value **(P2)** are explored and included in all three iterations. **Chapter** 3 describes the first artifact design iteration, which focuses on the principle of Multiple Knowledge(s) **(P3)**, by incorporating into data donation activities that support donors in situating and contextualizing their data. **Chapter** 4 describes the second artifact design iteration, which focuses on the principle of Ongoing Consent **(P4)**, by embedding data donation into an incremental process that supports donors in exploring their data and (re)assessing their participation, autonomously and in collaboration with me. **Chapter** 7 describes the third artifact design iteration, which focuses on the principle of Shared Goals **(P5)**, by promoting three different degrees of participation in data donation research: contributors, collaborators, and co-creators.



How can data donation integrate people's embodied and situated knowledge(s)?

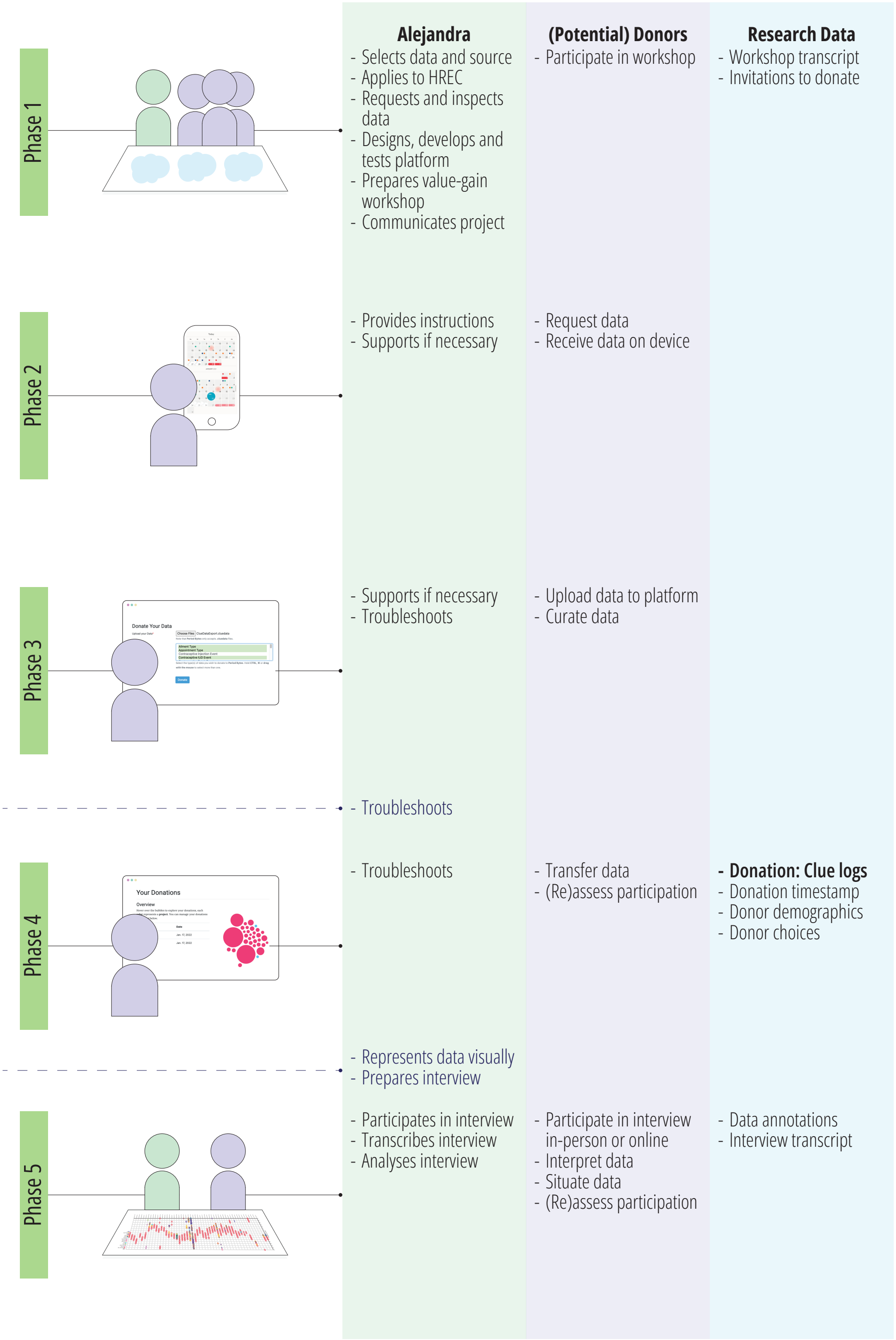
3

Integrating Multiple Knowledge(s) into Data Donation

Abstract

In this chapter, I investigate how data donation can integrate multiple forms of knowledge other than digital-trace data, especially people's embodied and situated knowledge (RQ2). To do so, I design and develop the first iteration of a sensitive data donation approach embodied by a digital data donation platform. It focuses on value creation, data curation, and data interpretation. I apply it in a research project aiming to support people who track their menstrual cycle in interacting with the data they produce. In this context, I focus specifically on the menstrual tracking app Clue which allows its users to self-report up to 31 different types of data. 35 donors participated by donating their data, of which 27 volunteered to interpret them and reflect on their data donation experience, and 13 were invited to do so. I discuss their data donation experience and conclude this chapter by discussing how involving donors in data interpretation influences the roles and relationships between donors and recipients and the properties of the donated data, leading to dynamic datasets.

This chapter draws on the following publication: **Gomez 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 (NordiCHI '22).



	Alejandra	(Potential) Donors	Research Data
Phase 1	- Selects data and source - Applies to HREC - Requests and inspects data - Designs, develops and tests platform - Prepares value-gain workshop - Communicates project	- Participate in workshop	- Workshop transcript - Invitations to donate
Phase 2	- Provides instructions - Supports if necessary	- Request data - Receive data on device	
Phase 3	- Supports if necessary - Troubleshoots	- Upload data to platform - Curate data	
	- Troubleshoots		
Phase 4	- Troubleshoots	- Transfer data - (Re)assess participation	**- Donation: Clue logs** - Donation timestamp - Donor demographics - Donor choices
	- Represents data visually - Prepares interview		
Phase 5	- Participates in interview - Transcribes interview - Analyses interview	- Participate in interview in-person or online - Interpret data - Situate data - (Re)assess participation	- Data annotations - Interview transcript

Figure 3.1: Overview of the case study showing the involved people and activities mapped around the five phases of Sensitive Data Donation.

3.1. Introduction

This chapter investigates how data donation can integrate people's embodied and situated knowledge(s) through the first Research through Design (RtD) iteration. I design and develop an instance of the data donation approach described in Chapter 2 – embodied by a digital data donation platform – to gain insights into data donation. I apply this instance of data donation in the context of menstrual tracking technologies – where people who menstruate rarely interact with the data they (self-)report. Aligned with the principle of Multiple Knowledge(s) **(P3)**, this iteration focuses on incorporating into data donation activities that involve donors in interpreting and contextualizing their data – prioritizing their embodied and situated knowledge.

The principle of Multiple Knowledge(s) **(P3)** draws from several perspectives, including a call to recognize data as 'situated' [90], not objective or neutral [51] but local, incomplete, and interpretative [115]; and the Human-Centered Design maxima: *"people are experts of their own experience"* [170]. These perspectives have seldom been considered in data donation research because of the domains, type of research questions, and contexts in which data donation is applied. These have different needs and priorities, such as sampling a representative group of a specific population [34, 142], or ensuring the ecological validity of the data [162]. One example of data donation research that encompasses more than the digital trace data is that of Razi et al. [162], who applied data donation to investigate online risk detection on Instagram and invited teenagers to annotate their direct messages based on pre-defined categories for risk level (i.e., low, medium, high) and risk type (i.e, nudity/porn, sexual messages, harassment, among others). This approach attempts to capture teenagers' perceptions and experiences and account for nuances, yet it relies on pre-defined categories somewhat rigid and limited.

Outside data donation research, designers and researchers have been exploring people's engagement and interpretation of their (intimate) data, including sensor data from connected homes (e.g., [109, 195]) and digital-trace data from connected appliances inside the home (e.g., [108, 154]) and wearable devices close to the body (e.g., [71, 134]). Interpretation is key as data – on their own – is ambiguous and incomplete. It lacks contextual information (e.g., physical, social, emotional) [148]. For instance, data captured during a period of disruption or non-routine circumstances, such as pregnancy, can be a source of concern if decontextualized [134]. Previous work have employed (interactive) data visualizations to support people in understanding and interpreting their data (e.g., [108, 134, 154, 195]). Thus, building upon previous research, the data donation process proposed in this chapter integrates interpretative activities that invite donors to contextualize their data through interactive visualizations.

Overall, with respect to empirical data donation approaches (e.g., [34, 162, 204]), the data donation process described in this iteration comprises:

- **Value creation**, potential donors are invited to co-define a value-gain strategy and consequently, participate in a dedicated non-transactional value-gain activity. It aligns with the principle of Balanced Value **(P1)**.

- **Data curation**, potential donors are explicitly prompted to decide whether and what types of data to donate – among the 31 different types collected by Clue – before donating them. It aligns with the principle of Sensitive Data **(P2)**.

- **Data interpretation**, donors are welcome to participate in a creative session to interpret and contextualize their data. Aligned with the principle of Multiple Knowledge(s) **(P3)**.

These activities are described in Section 3.3. With this being the first RtD iteration and consequently, the first time I apply data donation in practice, this chapter reports on several empirical insights. These include: how (potential) donors respond to the invitation to donate, the different choices made by donors when selecting which types of data to donate, their motivations and expectations experiencing data donation, and the outcomes of interpreting and contextualizing their data. Additionally, it introduces insights from the interaction between people who menstruate and their tracking data collected through digital apps. This chapter concludes with a reflection on the actors involved in data donation: data, donors, and recipients and the data donation process.

3.2. Intimate Context: Menstrual Tracking

I situate and apply data donation in the context of menstrual tracking technologies, which are digital technologies that allow people who menstruate to track and monitor their menstrual cycle. These technologies are often embedded into health platforms (e.g., Apple and Google Health, Fitbit) or widespread apps (e.g., Flo, Clue, Eve). Menstrual tracking technologies regularly collect personal and intimate data from their users (e.g., breast tenderness, sexual activity, sleep pattern) [69, 129]. They have been adopted by a wide range of users, with apps like Flo and Clue having 43 and 8 million active users, respectively [44, 67]. Design and HCI research in this area has explored why and how *women* track their menstrual cycles [63], the privacy implications of menstrual tracking and fertility technologies [69, 94, 129], and the design space for menstrual technologies that mediate self-touch and augment self-knowledge [38, 187]. Several shortcomings of menstrual tracking technologies are well documented in the literature. For instance, their gendered design, since these tend to use stereotypically feminine attributes (e.g., pink, flowers) [63]; or the fact that most technologies are built with normative assumptions that cast aside plural bodies, gender, and sexual minorities, as well as people with irregular cycles or reproductive health disorders [38, 42, 63, 69, 94]. Open questions remain about people's regular interaction with these technologies and how the data collected by these could better support its users.

Within this context, I aim to apply data donation to explore how to better support people who track their menstrual cycle through digital apps in interacting with the data they produce. This context underlines the need to incorporate Multiple Knowledge(s) **(P3)** into data donation, as it is critical to gain access to the data (i.e., menstrual tracking logs) but even more so, to situate and contextualize them within individual tracking practices and experiences. This chapter also introduces insights from the interaction between menstruating people and their menstrual tracking data collected through digital apps.

3.3. Data Donation Process

In this section, I describe how I design, develop, and apply a simplified version of the Sensitive Data Donation approach (Fig. 3.2). This iteration focuses on on value creation

(P1), data curation **(P2)**, and data interpretation **(P3)**. I integrate these principles into the design of the data donation platform and the resulting five-phase data donation process. The activities, participants, and data collected during each phase are summarized in Figure 3.1.

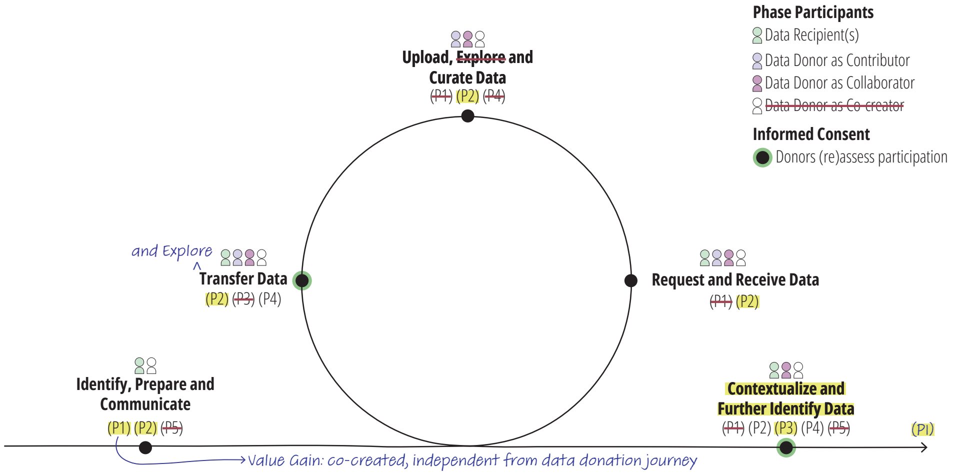


Figure 3.2: Simplification of the Sensitive Data Donation approach for this iteration, highlighting in yellow the main focus of this chapter and crossing out in red the omissions from the approach proposed in Chapter 2.

Additionally, I apply the procedural principles of transparency and autonomy through the phases of Sensitive Data Donation, specifically Phases 1 and 2 – where I communicate about the research and the process with potential donors – and Phases 3 and 4 – where I invite donors to choose whether and what data to transfer.

- Transparency, relates to the choice of information that is made accessible [192, 199]. Donors should have timely access to intelligible and relevant information on how their data will be handled. In this way, they are equipped to make informed decisions from the beginning and over time.
- Autonomy, relates to the ability to act [70, 97]. Donors exercise autonomy by inviting others to access and use their data [97, 210]. Donors should have the capacity and the mechanisms to act upon their data by defining their own terms and limitations.

These procedural principles are informed by previous literature on data donation [22, 107, 210], the ethical challenges posed by the practice of using data [68, 192], and the principles of Human-Data Interaction (i.e., legibility, agency, and negotiability) [48, 132].

Phase 1: Identify, Prepare and Communicate

To identify the relevant data sources and types of data, I explored how to request and obtain a copy of the data from menstrual tracking apps among the top five menstrual tracking technologies on the Google and Apple stores in The Netherlands in August, 2021 [10, 78]. From the plethora of menstrual tracking apps, I decided to focus on the Clue app, as Clue users can get a copy of their data by downloading a takeout file directly from the app. This contrasts with other apps. For instance, users of Flo and Eve are required to contact the support team to obtain a copy of their data, which complicates the data donation process. In the case of Period Tracker and My Calendar, it is not clear how to obtain a copy of the data from their privacy policy.

Users of the free version of Clue can log up to 31 types of data arranged in three groups: (1) single choice categorical data (e.g., *sleep duration:* 0-3 hours, 3-6 hours, 6-9 hours, 9 hours or more, *menstrual bleeding:* light, medium, heavy, spotting), (2) multiple choice categorical data (e.g., *menstrual pain:* cramps, headache, ovulation pain, tender breasts, *sexual activity:* unprotected, protected, high sex drive, withdrawal), (3) manual input (e.g., *weight:* numeric value, *text:* alphanumeric value). Data logged in Clue is stored with a date timestamp with no indication of time (e.g., 2022-09-18T00:00:00).

To define a value-gain strategy **(P1)**, I conducted a generative workshop with four participants, active users of menstrual tracking technologies. The workshop consisted of three activities and lasted 90 minutes. First, I invited participants to reflect individually on their feelings and experiences with menstruation and menstrual tracking technologies. Second, I introduced the concept of data donation and invited participants to envision the timeline of their preferred data donation experience. Finally, I invited the participants to pair up, discuss each other timelines and their preferred ways of gaining value from this experience. The workshop resulted in potential value gain strategies relevant to the specific context of menstruation. Beyond the motivation of helping others, broadly contributing to society (e.g., advancing research, raising awareness), and benefiting from the research outputs, already highlighted in the literature [179], participants expressed the wish for a *community feeling* and *personalized insights and feedback*. I integrated these into the data donation process by organizing a value gain event with an expert in reproductive health and sexuality. In doing so, I aimed to create a space for donors to ask questions, learn from the expert and each other, and share experiences. The event took place online in December 2021 and was advertised along with the invitation to donate.

In the design of the data donation platform, I incorporated the procedural principles described above.

- Transparency, by providing donors with concise and understandable information on how their data will be used and why [30, 132]. I opted for using plain language as well as graphs and diagrams when possible. Additionally, I provided my contact details should donors wish to reach out for additional information.

- Autonomy, by encouraging meaningful choices that allow donors to decide whether to donate, what to donate, and how to participate. In addition, I incorporated the mechanisms for donors to visualize and manage (e.g., delete) their donations at all times.

To prepare the platform to parse and visualize the data, I requested data from Clue through various dummy accounts and explored the structure of the resulting JSON file. The platform has three open-source components that manage (1) the user profiles and authentication, (2) the data storage and sharing, and (3) the donation process. The first two were implemented using TypeScript, and the third was implemented using the Python web framework Django. Data was passed between system components using web APIs.

Phase 2: Request and Receive Data

I reached out to potential donors and invited them to participate in the research by donating their data. To do so, I opted for convenience and snowball sampling by using my personal social media, reaching out to existing communities around menstruation, and contacting people and institutions that work around menstruation. In the call to donate, I specified who was conducting the research by stating my name and affiliation, what the goal was, who could participate (i.e., adult active users of the Clue app) and how (i.e., detailed instructions to obtain their data from Clue), and the opportunity to take part in the value gain event. Additionally, I shared a link to the data donation platform, where potential donors could read more about the project and find detailed instructions on how to donate their data (i.e., downloading/uploading their data from Clue). The call was open to donors from all over the world, as Clue is an EU-operating company and it is GDPR compliant for non-EU citizens, meaning they can also get a copy of their data. I advertised the project for five weeks, from the 15 of October until the 19 of November of 2021.

Phase 3: Upload and Curate Data

I designed the data donation platform to invite potential donors to (1) choose whether to participate in reconstructing the context of the data, (2) choose whether to receive updates from the project, (3) provide their demographic data (i.e., self-reported age and gender), and (4) curate their data by choosing which types of data to donate among the 31 types of data collected by Clue. There was no ‘select all’ and no options were pre-checked. Thus, the availability of data was subject to the choices made by potential donors throughout the uploading process **(P2)**.

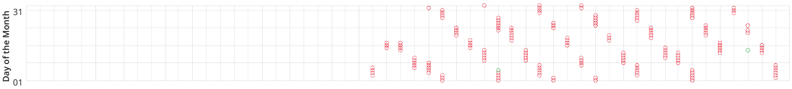
Phase 4: Transfer and Explore Data

After curating their data, potential donors could decide between transferring them or removing them from the data donation platform. If they decide to transfer their data, they can explore an interactive visualization representing the different types of data they shared. From this point onward, they can revoke access to the data recipient and delete their data from the platform through a button.

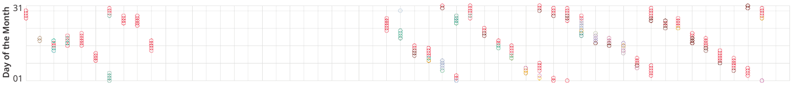
Phase 5: Contextualize Data

To involve donors in situating and contextualizing their data **(P3)**, I conducted semi-structured interviews with a subset of the donors who agreed to participate in this step. I opted for semi-structured interviews as this is the most common approach in the previous HCD and HCI literature (e.g., [26, 28, 195]). The interviews lasted between 30 and 45 minutes and broadly comprised three activities. First, I invited donors to briefly discuss the highlights of their experience with menstrual tracking technologies. Second, I invited donors to explore, interpret, and reflect upon their data. Third, I invited donors to envision future menstrual tracking technologies that fit their needs.

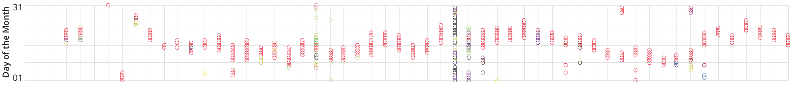
(a) Donor logs menstrual bleeding and contraceptive events. Questions include: Why do you log your data? How do you log your data?



(b) Donor logs menstrual bleeding and other events in days close to menstruation, stops using clue for 16 months. Questions include: How do you decide what to log?



(c) Donor logs menstrual bleeding and other events in days close to menstruation, and daily events for a few mohtns. Questions include: What data do you log more?



(d) Donor logs daily events. Questions include: When in your daily routine do you use Clue? How frequently do you look at your data?

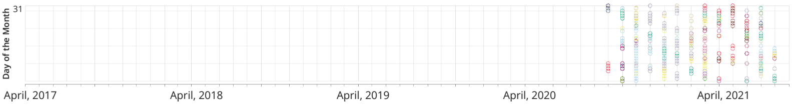


Figure 3.3: Timeline visualization of four donations, each with a different starting date and different types of data.

I created interactive data visualizations to support the interviews. I visualized every interaction with the Clue app on a timeline, where the x-axis shows the month and year, and the y-axis shows the day of the month (Fig. 3.3). If a single interaction is selected, a tooltip shows the type and value of the data point (e.g., *type:* menstrual pain, *value:* tender breasts). I opted for this visualization, as previous research highlights how viewing data as points on a graph arranged by time helps people reason about it [154]. When focusing on specific interactions and patterns, I visualized individual types of data on the timeline by using markers (e.g., *sexual activity:* unprotected, protected, high sex drive, withdrawal) or varying the point size (e.g., *menstrual bleeding:* light, medium, heavy, spotting). In addition, if the donation contained multiple types of data I visualized them on a bar chart to show an overview over time. When visualizing the data I prepare for the corresponding interview, by focusing on specific events or patterns. In addition to introductory questions (e.g., Why did you start using Clue? When do you use Clue?, How did you choose which types of data to track and donate?), I prepare questions informed by the data and specific to each donor. I illustrate this process with four examples in Figure 3.3. I conducted interviews in person and online via Zoom. For the in-person setting, I brought the data visualizations printed on A3 sheets of paper and interactive data visualizations displayed on a laptop. For the online setting, I used the whiteboard tool Miro. The interviews were conducted in English and Spanish, depending on the preferences of the donors.

Data Donation Procedure

The resulting data donation procedure (Fig. 3.4) comprises six voluntary activities. It begins with donors receiving an invitation to participate in the research. It continues with donors requesting and receiving a copy of their data on the Clue App, and uploading it into the data donation platform. Here, the data donation platform explicitly invites donors to select which types of data they would like to donate, or transfer, and to decide if and how they want to participate in the subsequent activities. Donors who expressed a desire to participate in an individual session to interpret, situate, and contextualize their data receive an invitation to schedule it in person or online. Further, all donors receive an invitation to attend the online value-gain event.

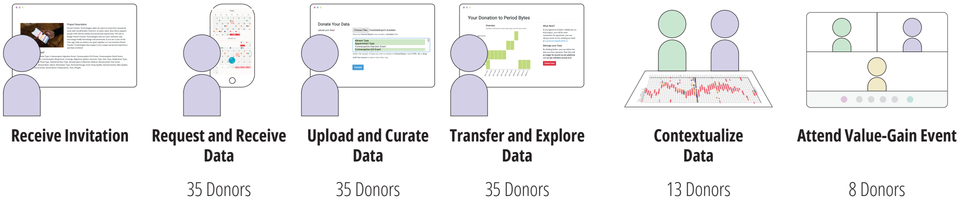


Figure 3.4: Data donation procedure illustrating the various activities, mediated by the data donation platform, in which potential donors participate and how many donors participated in each activity.

3.4. Applying and Reflecting on Data Donationn

3.4.1. Participants

In total 35 donors (D1-35), aged 19 to 44 (mean: 30.0, median: 29, 2 did not provide age data), contributed to the project by donating their data (Fig. 3.5). Out of these, 33 identified as female, one identified as non-binary, and one did not provide gender data. 27 donors indicated that they wanted to participate in the interview. I invited the first 19 to schedule the interview at their convenience, I did not send an invitation to the rest (n=8) due to limited availability on my side – I underestimated how much time it took to prepare each data visualization(s). Ultimately, 13 donors, aged 20 to 36 (mean: 28.8, median: 29) identified as female, scheduled an interview. They were spread across eight countries: Argentina, Brazil, Canada, Colombia, Germany, The Netherlands, Portugal, and The United States. In addition, eight donors accepted the invitation to participate in the value gain event. Most of the donors (n=27) did not participate in the value gain event, four responded to our invitation mentioning that they had other commitments.

3.4.2. Data and Analysis

To reflect on my experience and that of data donors, I generated and collected three types of data. First, the different channels and times through which I distributed the in-

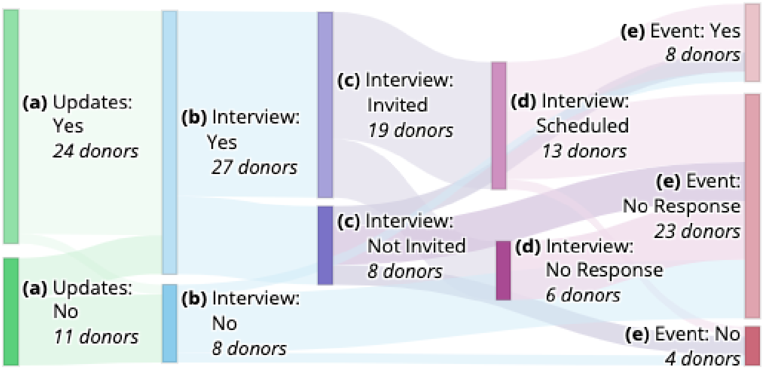


Figure 3.5: Preferences of the donors (35) throughout the data donation journey. From left to right: (a) receiving updates from the project, (b) participating in interview, (c) receiving an invitation to schedule the interview, (d) scheduling the interview, and (e) attending the value gain event.

vitation to donate and the donation timestamps (i.e., time and date the donation was made), to reflect on the overall dissemination strategy. Second, the donors' demographics (i.e., self-reported age and gender), and choices made throughout the journey (i.e., whether to participate or receive updates, which (types of) data to donate), to better understand if and how they were able to set their terms and boundaries. Finally, the audio recordings from the semi-structured interviews. I transcribed these using MS Office and manually reviewed each transcript. I analyzed the interview transcripts using Reflexive-Thematic Analysis [32, 33] within a constructionist paradigm. To do so, I went through the transcripts to familiarise myself with the data and inductively coded the entire dataset using ATLAS.ti. Then, I reviewed the codes and grouped them into tentative themes. I iteratively reviewed and refined the themes. The resulting themes and subthemes were discussed and refined with a co-author of the study, Jacky Bourgeois. In the next section, I present results related to these three types of data.

3.5. Results

3.5.1. Invitation to Donate

I distributed the call to donate across different channels, including my personal social media (i.e., LinkedIn and Twitter), online communities (i.e., subreddits Menstruation and MenstruationStation, Get a Clue a Facebook group for users of Clue), and by contacting people and institutions that work on menstruation and related topics. Some of my efforts were unsuccessful, for example, moderators on one subreddit did not approve my request to post the message, while on another, my post was deemed *spam* and removed minutes later. Similarly, some of the people and companies I contacted did not respond or were hesitant to share the message with their community. For instance, Clue abstained from sharing it as *"it may be confusing to their users."*

Those who agreed to share the message include social media influencer @VanillaCoolDance, the Period! magazine, and the company of menstrual products Cute Cotton, all three based in the Netherlands. Figure 3.6 provides an overview of the donations I received over time as well as the estimated views of the invitation throughout the different channels it was advertised. For LinkedIn and Twitter, the estimated views correspond to the number of views reported by each platform, re-posts from my personal contacts are not accounted for. While for Facebook, Instagram, and the Cute Cotton Newsletter they were calculated as a percentage of the number of followers or subscribers. In total, the invitation to donate had an estimated of 8.000 views. My efforts were more successful (i.e., more people donated their data) when the invitation was shared with existing communities by their members (e.g., @VanillaCoolDance, Cute Cotton), followed by when the call was shared with my personal network. This is consistent with the research of Skatova and colleagues, which highlights that familiarity and trustworthiness with the data recipients can influence the decision to donate [179]. Having established members of existing communities as intermediaries facilitates building trust and familiarity, while using my personal social media means the request was seen and shared by people who are already familiar with me, my research and institution. In contrast, my efforts were the least successful (i.e., fewer people donated their data) when I tried to personally share the invitation with existing online communities (e.g., subreddits, Get a Clue). For instance, in the Get a Clue community no one engaged with the post.

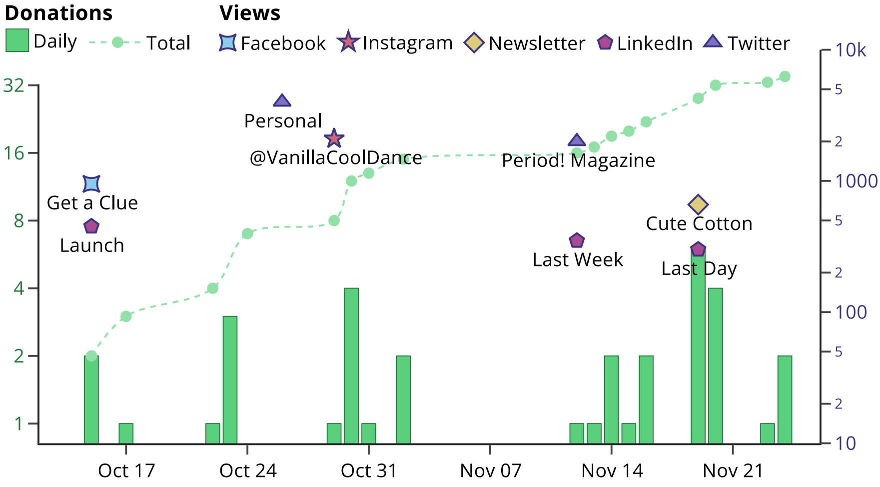


Figure 3.6: Daily donations over time, derived from the timestamp of the donations, and estimated views of the invitation to donate on different channels.

3.5.2. Receiving Data Donations

Donors could choose which types of data to donate, by manually selecting each one on the data donation platform, there was no ‘select all’ option. 17 donors selected all types of data upon donation while 18 made a specific selection (Fig. 3.7). Out of these, 6 donated only their menstrual bleeding, 3 their menstrual bleeding, menstrual fluid and menstrual pain, while the rest made a different range of choices including up to 15 types of data. The received datasets have unique characteristics that result from data donation. First, different temporal ranges and go back in time for months and even years. From the date where donors first logged their data into the Clue app, different for everyone, to the date of donation. It leads to the availability of data over an extended period of time without having to conduct a longitudinal study. The earliest starting date dates to mid-2015, while the latest one dates to mid-2021. Second, the data collection is embedded into the donors' routine and is not attached to a research project or a research instrument, thus less prone to observation bias. Third, the datasets contain different types of data because of the selections made by the donors and the unique ways they interact with Clue. In some cases, this meant I didn't have enough information without having to make assumptions or interpretations. For example, one donor donated only the menstrual bleeding and chose not to participate in the interview, leaving me with open questions about the data I did and did not receive. The fact that slightly more than half of the participants made a choice over which data to donate sheds light on the importance of supporting different privacy preferences and adapting to work with the data that is made available. This might be not ideal for designers and researchers, I found it inconvenient in some cases, but it would be an interesting shift towards giving people more control of their data in design and research activities.

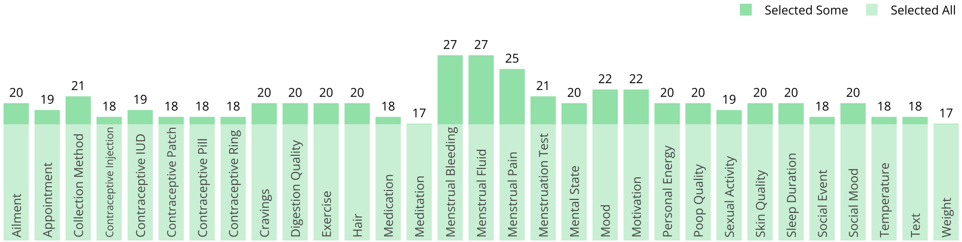


Figure 3.7: Selections made by the donors at the time of donation for each type of data.

3.5.3. Reconstructing the Context of the Data

Donors were positively surprised to see an overview of their data over time which seemed *"more clear"* (D16) and *"say a lot more"* (D15) than the calendar view provided by Clue. During the interviews, I experienced first-hand the value of situating and contextualizing the data, as this allowed me to delve into the lived experiences behind them and fill in the gaps. For example, in Figure 3.3.a, understanding why (and why then) the tracking began.

> "I stopped using the the pill in [date], so that's where I really wanted to like track it again" (D33).

Interpreting gaps, such as the 16 month period with no data on Figure 3.3.b. Which could easily be misinterpreted, for example, by attributing it to pregnancy.

> "I've been on and off also trying different methods to actually figure out when my period would come because it's a bit different. But, it's also one

> of these things I'm doing it for myself so I don't have to be 100%, so it's really useful on and off."

Explaining changes, such as the near everyday tracking in November and December 2019 in Figure 3.3.c.

> "In [date] I was in [country] and doing this course with different women and they told me about tracking your cycle so it's more in tune. So then I started tracking [more aspects]" (D29).

Or, in Figure 3.3.d, managing the intricacies of tracking and dealing with limited options and categories.

> "Let's just try to see if my mood has some different changes near [menstruation], but then, I don't know, like on the same day I could be happy, sad, and sensitive" (D2).

In addition, the interviews allowed me to learn about the donation experience from the donors' side. Below, I report on my learnings specific to the donation experience and the contextualized insights on menstrual tracking.

3.5.4. Motivation to Donate Intimate Data

Donors expressed different drivers that led them to contribute to the project. First, contributing to society by helping advance research, as expressed by D29:

> "It feels that you're doing something good for overall society and hope it will make a difference."

There was a consensus among donors on the importance of contributing to research, *"feeling that I'm contributing to research seems cool to me, it seems important to me"* (D5). Some of the donors had previous experiences with research, either by conducting research themselves (D5, D15, D16, D22) or by participating in data-intensive research studies (D8, D29). As described by D8:

> "I haven't done this [data donation], but there is like a big health research in [country]. So I'm part of that. Which I think is really interesting because they're just kind of tracking you throughout the years".

Moreover, donors expressed an eagerness to contribute especially to research about women, menstruation, and FemTech, *"even though there is research on medicine or whatever, I feel that we have a large scope to explore how different all women are and we have always seen each other lagged behind in research"* (D15). In fact, a second driver was interest in the research topic and the research outcomes, as expressed by D33:

> "I learned over time that this is something that has barely been researched like period. In like the big spectrum. So when I saw this I was immediately, like oh, that sounds interesting, I want to participate in this! And it's quite interesting to see what your research will conclude."

Finally, a third driver for those in our personal network was contributing to *my* research, as expressed by D5: *"it is also cool that I know you, you are my friend and I am helping your research"*.

3.5.5. Experiencing Data Donation

Donors described the data donation process as a positive experience that led to positive feelings, *"it felt nice to be able to donate data because it's free but it can help so much"* (D29). They found the process of downloading and donating their data fairly easy and appreciated the instructions we provided.

> "It was actually quite easy because you all described it quite well on how to do it. So for me to actually upload data it was really easy." (D33).

Although some highlighted that without the instructions *"it was not so easy to export the data, it was not so obvious"* (D5). While one donor described the process as a bit burdensome on our end *"I did have some struggles with the website from your university 'cause I had to click like a few times, but then it was surprisingly easy as well"* (D29). The choice of whether or not to donate was straightforward and resulted from a *why not* mindset. Yet, donors had mixed attitudes with respect to sharing their data and privacy. Some *"are not bothered about sharing data"* (D5) and *"didn't even think of it. Just downloaded* [the data] *and put everything up"* (D22). While others expressed higher privacy concerns and thoroughly informed themselves before making the donation.

> "I read everything that you wrote about it. So, what you were researching. And also, there was the option to select which data you wanted to donate, and there's a possibility to withdraw at any moment. So yeah, I felt quite comfortable." (D33)

Similarly, donors had mixed feelings about disclosing and discussing their menstrual cycle. Some consider it to be *"private"* (D7) and *"find it a bit uncomfortable to discuss certain topics"* (D16). While others are actively making an effort to discuss these topics *"openly in daily life"* (D22). Overall, donors trusted the data donation approach: *"I feel like you're doing a really good job the way you're going about* [the data donation process] *"* (D8). The institution: "[the researchers] *are working in a university I attended, and someone that I knew shared* [the call] *so I had no reason to distrust"* (D13). And the data protection regulations in Europe: *"I feel like in Europe in general, universities are quite good at keeping* [data] *well, they do have privacy rules, right? And regulations"* (D22).

3.5.6. Into the Context: Tracking the Menstrual Cycle

Similarly to Epstein and colleagues [63], I identified 6 reasons why people track their menstrual cycle with Clue: (1) get predictions about menstruation, (2) gain a better understanding of the menstruating body throughout the cycle, (3) monitor a specific health condition or pain, (4) get pregnant or avoid getting pregnant, (5) inform conversations with healthcare providers that often begin with *"when was your last period?"*, and (6) prepare, avoid accidents, and plan for various life events. In addition, I identified three ways in which people interact with Clue to track the menstrual cycle, these are complementary and changing throughout life.

First, logging menstruation, which is approached in a *"systematic way"* (D16). Most of the donors who interact with Clue in this way expressed logging their data as soon as they menstruate and retroactively logging *"a few days down the line to fill in any blanks"*

(D29). In this case, tracking is primarily a means of receiving accurate predictions and it is seen as a *"monthly task"* (D7). Common among the donors is the notion that logging the data is essential to receiving accurate predictions, as expressed by D7:

> "What makes it more accurate is that one begins to enter all the monthly information so that it is much more precise".

Second, logging menstruation and related events. In this case, tracking enables a better understanding of the menstruating body throughout the cycle and facilitates identifying patterns and managing pain. Here, what is logged, apart from menstrual bleeding, varies greatly between and within donors.

> "So I usually put the main things that I never forget, like my period and my skin as well because I was having a lot of acne because I stopped the pill. Then menstrual pain and the energy levels and sexual activity of course, because I don't want to get pregnant." (D4)

Within donors, the types of data that are logged tend to change over time due to various factors, including pregnancy, stress, cycle variations, and a new understanding of the cycle events and symptoms, among others. For instance, for D8 logging changed after identifying a pattern in her cycle:

> "I think that's the main reason why I stopped tracking that stuff because I was able to figure out it is my period causing this as it is and not some other random thing."

While changing what to log or reviewing log settings is possible with Clue, it's not made easy for people. In fact, donors expressed not being aware of some features and using other ways to log data that they could have logged with Clue if they knew it was possible. For instance, D4 uses another app to track her weight *"until about two weeks ago I realized that Clue also had* [an option] *for* [tracking] *the weight."*

Third, logging daily events. In this case, tracking is seen as a way to gain a more *"holistic view"* (D16) of the menstruating body throughout the cycle. Similar to the previous interaction pattern, what is tracked varies greatly between and within donors. In this case, tracking is often prompted by a notification from Clue, *"the app reminds me like, can you track your symptoms?"* (D2). Most of the donors engaging in this pattern of interaction expressed that they rarely looked back at their data, for instance D2:

> "I just put it there and then don't really look back. I actually don't know how to do it on Clue. Like I don't know there's a way to see like the past things, right?"

Furthermore, donors often get discouraged and tend to abandon this pattern of interaction over time *"I started tracking* [daily], *but I never actually did it more. Yeah, I guess I was bored of it"* (D29).

3.6. Discussion

In this chapter, I explored how data donation can involve donors in interpreting, situating, and contextualizing their data. Over five weeks, I received diverse datasets from 35 donors and situated and contextualized 13 of them, demonstrating the feasibility of this approach. I anticipated it to be a hurdle as it entails donors being available and willing to discuss topics that might be considered personal and awkward. Hence, I was positively surprised when the majority of donors (27) expressed their willingness to participate in this step, exceeding my planned interview capacity. The attitude of the donors, reflected in their choices, is promising for future data donation projects that aim to embrace Multiple Knowledge **(P3)**. Moreover, the flexibility of the material I used for the interviews allowed me to conduct them in-person and remotely. The flexible setup and the possibility to conduct remote research were suitable for the context of the COVID-19 pandemic. Yet, opting for interviews meant that the process was time-consuming, and during the last few weeks of the project it was difficult for me to schedule a meeting with everyone who was willing to do so. Future data donation research, could benefit from approaches that allow contextualization activities to be carried out remotely and asynchronously.

In this section, I abstract my experience from the research activities and translate it into a reflection on the actors involved in data donation and the procedural principles. Furthermore, I provide design recommendations for approaching Sensitive Data Donation.

3.6.1. Reflections on the Actors

Data: Shaped by the Process and Fostering Intimacy

Data are at the core of data donation and have unique properties that enable and result from the process. First and foremost, data are *personal*, they relate to a person, are partially created by a person, her behavior, her characteristics, and her interactions [155], and they are donated and contextualized by a person. Second, data are *dynamic*, spatiotemporal, and changing through people's actions and choices. Donated data could be from the past, present, and future and could be enriched by reconstructing its context. Third, data are *multiple*, defined by Prainsack [155, 156] as being able to be in more places that one at the same time, and being able to be copied and used by several people at the same time, independent of what the others are doing. Multiplicity enables data donation and allows donors to simultaneously contribute to one or more projects.

Furthermore, I argue that through Sensitive Data Donation data are *intimate*. The term 'intimate data' has been used in HCD and HCI to refer to data collected from intimate activities (e.g., sleeping, showering, urinating), or activities whose data might be perceived as *unusual* or *awkward* [6, 109]. Recent work highlights different perspectives and approaches. Kwon and colleagues explore the collection and sense-making of intimate data in the context of showering, where they combine sensor data (e.g., gyroscope, temperature, water flow) with contextual interviews where participants reflected on the data collection [109]. They argue that sensor data are not intimate per se, but they get intimate connotations by the articulation of shower practices. In the area of bodily care and women's health, often subject to taboos and associated with notions of filth and shame [5], intimacy is deeply rooted into the nature of the experiences and interactions, that are then captured and transformed into data [6, 14, 92, 129]. For instance, Helms [92] reflects on how her own urinating data transforms a bodily experience into something that is sensed or measured, and Almeida and colleagues [6], set out to support and empower body literacy through digital technologies that promote looking at the vagina and harness the *awkwardness*. In the case of menstrual logs, intimacy derives both from the data, depicting a bodily experience, and from reconstructing their context. In this way, Sensitive Data Donation produces intimate data, where intimacy is a product of the data themselves, depicting a person's behavior and characteristics, and their further contextualization. Moreover, as data gain intimacy while being shaped by a person's behavior and further contextualization, they foster intimacy between donors and recipients.

Donors: Intentionally giving

Donors are people who knowingly and willingly contribute to a project by donating and actively reconstructing the context of their personal data. In this way, donors intentionally contribute to research or design projects in specific contexts. Intentionality is important since it brings to light specific characteristics of the donors, such as their high regard for research in topics they consider relevant [179], and their willingness to participate in research activities [29]. Nonetheless, a variety of factors influence donors' willingness to donate including the type of data being donated, the data donation effort, the way information is presented, the context of the project, familiarity and trustworthiness of data recipients, as well as privacy concerns, as highlighted by previous research [179, 180]. For this reason, even when donors are highly motivated it is fundamental to enable them to define their own privacy preferences [22, 107, 210]. I suggest supporting privacy preferences on an individual and granular level as it might lower the donation threshold for donors with higher privacy concerns who might be hesitant. Besides, donors' privacy preferences could become design research insights themselves.

On the other hand, intentionality means that donors donate their data with underlying expectations over the research process and benefits, and with the hope that an outcome will be achieved. This is reflected in the choices donors made throughout the data donation process, where the majority (24) agreed to receive updates from the project (Fig. 3.5). Focusing on outcomes is not realistic, especially in contexts where projects are open-ended and exploratory. Still, donors' expectations over the process and benefits should be honored within reason. In Sensitive Data Donation, where reconstructing the context is part of the process and involves donors' engagement over time, honoring donors' expectations is central in building and maintaining a trusted long-term collaboration resulting in richer and more intimate datasets.

Data Recipients: Driving collaboration through accountability

Data recipients are designers and researchers working on a topic or project that involves personal data who seek the active collaboration of people. Sensitive Data Donation requires them to take on various roles and tasks. For example, considering the needs and wants of donors and defining ways for them to benefit [22, 97, 107, 155], interacting with regulatory bodies (e.g., Human Research Ethics Committees, Institutional Review Boards) to define a process that minimizes risks to donors, and receiving, shaping, and contextualizing the data, where they engage donors in activities of interpretation and sense-making. The last two are already part of HCD and HCI research activities [30, 77, 80, 205].

Data recipients initiate data donation by inviting potential donors, but it is up to donors to carry through [97]. In this way, Sensitive Data Donation, and consequently the quality of the data, is reliant on the donors' willingness to donate and reconstruct the context of their personal data. I previously described how donors carry underlying expectations over the data donation process and the outcomes, for the recipients these expectations translate into accountability. I suggest data recipients be accountable to donors and regulatory bodies as data donation unfolds. In this case, accountability includes delivering on the expected ways to gain value, informing donors about the process, progress, and limitations, notifying donors in case of data and security breaches, updating donors about the ways they are expected to contribute, and including donors in the final outputs. Thus, data recipients build and maintain a bond with the donors that fosters a feeling of ownership over the process and promotes collaboration.

3.6.2. Reflections on the Procedural Principles

The data donation platform was a central element in defining the process and shaping my approach, and a first effort to apply the procedural principles of *transparency*, and *autonomy*. Based on my experience, I discuss the shortcomings in putting these principles into practice, propose future considerations, and suggest a third sub-principle: *awareness*.

1. *Transparency*, I aimed to provide donors with access to understandable information on how their personal data is handled by providing clear details up-front. Although not the case of our project, the objectives and ways of handling data can change throughout design research projects that are open and exploratory, meaning transparency should be maintained along the way. Donors should be informed in a timely manner across the process, through reminders and updates that can be adapted to donors' privacy preferences.

2. *Autonomy*, I facilitated autonomy by allowing donors to decide whether to donate, what to donate, and how to participate. However, during the one-on-one session, I learned that some donors were uncomfortable sharing specific events *"I was a little concerned with* [a data point], *I don't feel as comfortable sharing that and letting others know"* (D16). Donors could choose which types of data to donate, but within each type, they could not choose whether to donate or exclude certain data. Donor autonomy should be supported in an even more granular way. For instance, by facilitating the filtering of specific data points. In addition, I provided the mechanisms for donors to donate their data, revoke access to the researchers and delete their data from the platform. As of April 2022, no donor has revoked access to the researchers or deleted their data from the platform. Similarly, in traditional design research participants seldom withdraw. As design researchers, and data recipients, we shall become more proactive towards helping people exercise their autonomy and reassess the terms of their participation. Beyond providing the mechanisms, supporting the process, and establishing checkpoints.

3. *Awareness*, across the data donation process, I failed to provide donors with a way to explore their data before making a donation. Due to the multiple ways in which

data can be collected and generated (e.g., manual self-reporting, automated logging or sensing), people may have an active or passive role in the process and may be more or less aware of what the data contains. Besides, most data controllers give users a JSON takeout file with their data, and these files are not easily explored by people without a technical background. There is an opportunity to better support donors in understanding what is on these files, resulting in a better-informed decision. Donors should have the capacity and the mechanisms to explore the content of their data before making a donation. For instance, by visualizing and exploring the data before donors make a decision to donate.

I am aware that these principles entail a trade-off (e.g., more information could be overwhelming, reminders could be annoying, and exploring/filtering datasets could be burdensome) which is why they ought to be adapted to the (privacy) preferences of each donor and the characteristics of the design process.

3.6.3. Recommendations: Approaching Sensitive Data Donation

I conducted several activities that strengthened my understanding of how to approach Sensitive Data Donation. Based on my experience, I develop five recommendations for data recipients, designers and researchers.

1. Donors enable data donation and in doing so facilitate design and research. There are no intermediaries or third parties, donors directly and willingly contribute to a design or research project, as a result, donors should directly gain value and their contributions should be valued. Data recipients should find ways to offer direct and time-bound value to donors. For instance, providing insights obtained from the data or benefits derived from research outputs.

2. Data recipients should support donors throughout the data donation process. For example, by providing clear instructions on how to donate and making the process as simple as possible. Yet, this process is limited by the data portability alternatives offered by data controllers. Moreover, data recipients should encourage donors to (re)define their privacy preferences even if this might result in less data available.

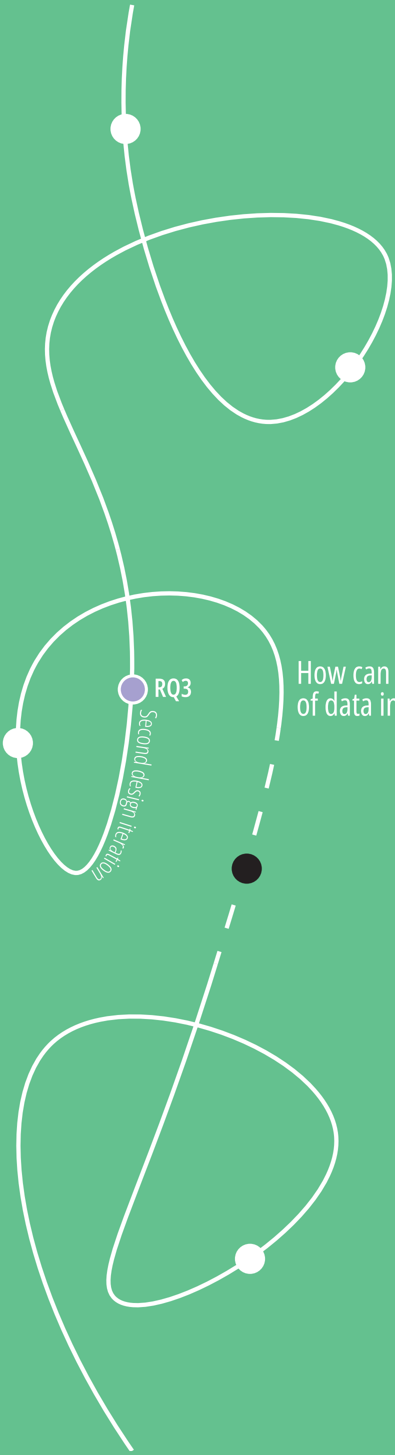
3. Data recipients should carefully consider where and how to invite potential data donors. Having members of existing communities as intermediaries facilitate spreading a call to donate, but might hamper diversity and lead to bias.

4. Data obtained through data donation was collected and generated in-the-wild, where people behave freely, naturally, and even wildly (e.g., in unique and unexpected ways). Data recipients should be prepared to work with datasets that are dynamic, diverse, and reflect this wildness (e.g., unique, incomplete).

5. Shaping and reconstructing the context of the data obtained through data donation requires data recipients to be flexible and adaptable to multiple contexts (e.g., remote research, in-person research) and (privacy) preferences. In addition to finding suitable ways to visualize the data and communicate with donors through it.

3.7. Chapter Takeaways

In this first iteration, I focused on integrating into the data donation process activities that invited donors to contribute beyond their digital-trace data. Specifically, augmenting the data with their embodied and situated knowledge through interpretation and contextualization **(P3)**. Inspired by current practices in HCD and HCI I approached the principle of Multiple Knowledge(s) **(P3)** through an individual interview supported by the data. This process was engaging and exciting for donors, yet it was time-consuming on my side. Additionally, I applied the procedural principles of *transparency* and *autonomy*. However, donors' experiences revealed that they are not necessarily aware of the content of their sensitive data at the time of donation. Thus, it is necessary to promote *awareness* and better support donors in exploring, understanding, and building intimacy with their data autonomously and before they decide whether to donate it. This illustrates the limitations of how data curation **(P2)** was implemented in this iteration. Moreover, it introduces the need to support donors in knowing and exploring the data before they are transferred to recipients. Awareness could be a way to foster value creation in itself, which is the focus of Chapter 4. Moreover, I approached value creation **(P1)** through a dedicated activity in which not all donors participated. How to approach value creation in a way that better suits the research process and activities remains an open question.



How can data donation foster an incremental understanding of data inviting donors to (re)assess their participation?

4

Incorporating Dynamic Consent into Data Donation

Abstract

In this chapter, I investigate how data donation can foster an incremental understanding of data that invites donors to (re)assess their participation (RQ3). To do so, I design and develop the second iteration of a sensitive data donation approach embodied by a digital data donation platform. It focuses on incrementally supporting donors in knowing their data before they decide to donate it and throughout their voluntary participation in subsequent activities. I hypothesize this incremental process yields a return of value to them and enables them to be *adequately informed* about their data, how and why they will be used at the time of initial informed consent, and (re)assess their participation in the research throughout the data donation process. I apply the data donation process in a research project aiming to investigate people's perceptions of their Google Assistant speech records. 22 donors participated by donating their data, of which 17 volunteered to interpret them and reflect on their data donation experience. I discuss their experience and conclude this chapter by illustrating the gap between knowing that data is collected and knowing what data collection feels like. Further, I discuss how bridging this gap through a data donation process is perceived as valuable and requires an ongoing (re)assessment around disclosure and participation.

This chapter draws on the following publication: **Gómez Ortega, A.**, Bourgeois, J., Hutiri, W.T., and Kortuem, G. (2023) Beyond Data Transactions: A Framework for Meaningfully Informed Data Donation. AI & Society.

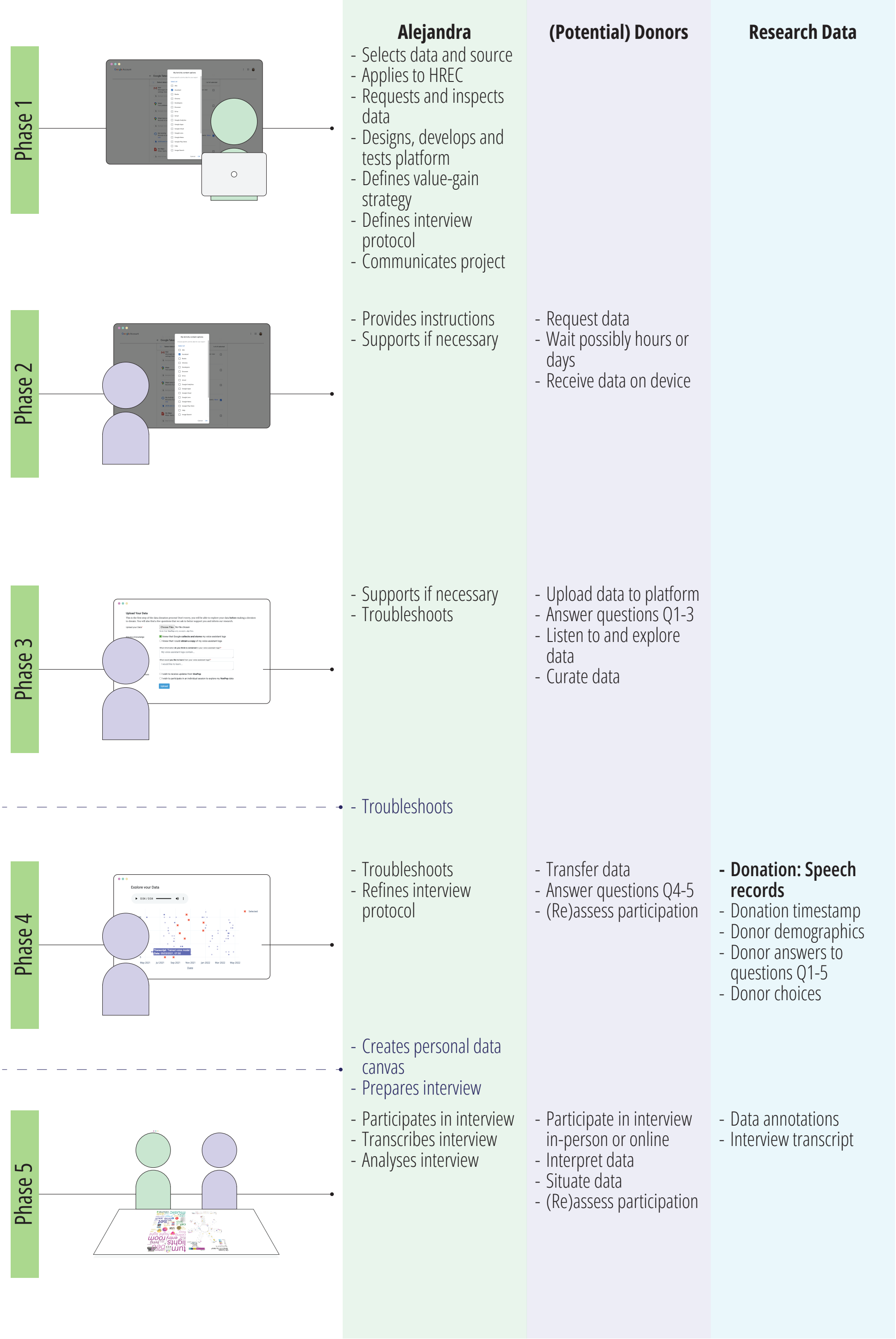


Figure 4.1: Overview of the case study showing the involved people and activities mapped around the five phases of Sensitive Data Donation.

4.1. Introduction

This chapter investigates how a data donation process can foster an incremental understanding of data that invites donors to (re)assess their participation through the second Research through Design (RtD) iteration. I design and develop an instance of the data donation approach described in Chapter 2 – embodied by a digital data donation platform. Additionally, I incorporate the learnings from Chapter 3, specifically how donors are often unaware of the content and characteristics of their sensitive data at the time of donation. I apply this instance of data donation in the context of domestic interactions between the inhabitants of a house and their voice assistants. Aligned with the principle of Ongoing Consent **(P4)**, this iteration focuses on integrating ongoing consent practices into data donation cemented on an increased understanding of data. More specifically, I hypothesize that incrementally supporting donors in knowing their data yields a return of value to them and enables them to be *adequately informed* about their data, how and why they will be used at the time of initial informed consent, and (re)assess their participation in the research throughout the data donation process.

The principle of Ongoing Consent **(P4)** builds on informed consent, an ethical and legal requirement for research involving human participants. It is considered one of the main challenges of digital-trace data donation [22, 100, 140, 190], and, more broadly, of research and other activities involving transactions of Big digital-trace data [9]. Informed consent requires that people understand the content and characteristics of their data, how and why their data will be used, and the advantages, disadvantages, and potential risks associated with the usage of their data [139]. In this context, Breuer et al., [34] compared two of the most common approaches to data donation – of Facebook logs – (1) applications to which donors give permission to scrape data using their personal account(s) and (2) digital platforms or research repositories to which donors manually upload a copy of their data, previously requested from a data controller. They argue that in both approaches it is possible to obtain informed consent from donors and the second one offers higher transparency to donors; who can *“see exactly what types of data they will share with the researchers.”*

However, they – and I the iteration described in Chapter 3 – fail(ed) to consider (1) donors’ general (and limited) understanding of their data at the time of informed consent and (2) the practicalities of requesting and obtaining a copy of the data, which do not lead donors to *“see exactly what types of data they will share with the researchers.”* First, at the time of informed consent, often prior to and independent of the data donation (e.g., [34, 123, 162, 166]), the content of the data remains obscure and abstract. Most donors *“don’t know what they don’t know”* [100] about their data. Thus, it is hard for them to anticipate the (privacy) implications of donating them. Second, arguing that the approach where donors manually upload a copy of their data to a digital platform or research repository offers higher transparency assumes donors can obtain information from it. Bowyer et al., [31] conducted a study inviting 11 people to obtain a copy of their data from different organizations; they found that in most cases people were left *“in the dark”* and the returned data was *“often difficult to understand, impractical to use, and raised new questions and concerns.”* Similarly, Alizadeh et al., [4] concluded that people require support in understanding and making sense of the files and the data that are returned. This means donors might be able to see the files of data they will share with the researchers but might not *adequately* understand the (personal and sensitive) information they contain (and that they are giving away) and its implications.

Overall, with respect to the data donation approach described in Chapter 3, the data donation process described in this iteration comprises:

- **Value creation through data**, potential donors can derive value from their participation in the research through an incremental process of knowing and getting close to their data. It aligns with the principle of Balanced Value **(P1)**.
- **Data exploration before curation**, potential donors are invited to explore their data locally through an interactive tool on the data donation platform. They can familiarize themselves with their data, and identify its content and characteristics *before* deciding whether and what to donate. More importantly, *before* data is transferred. It aligns with the principle of Sensitive Data **(P2)**.
- **Ongoing participation (re)assessment**, potential donors are explicitly invited to ongoingly (re)assess their participation preferences as they engage differently with their data in the various stages of the data donation process. It aligns with the principle of Ongoing Consent **(P4)**.

These activities are described in Section 4.3. This chapter reports on donors' experiences throughout the data donation process. It describes their knowledge and understanding of their data (i.e., personal data literacy) and perceived value gain. Additionally, it underlines some of the limitations and future challenges of data donation processes, such as their complexity and the relational aspects of data that should be accounted for. This chapter concludes with a reflection on the principles of Balanced Value **(P1)**, Sensitive Data **(P2)**, and Ongoing Consent **(P4)**.

4.2. Intimate Context: Voice Assistants in the Home

I situate and apply data donation in the context of domestic interactions between the inhabitants of a house and their voice assistants. Voice assistants are routinely used by millions of people around the world as part of their daily and social lives [154]. It is reported that in 2022 Google Assistant and Apple's Siri were each used by over 500 million people worldwide, while Amazon's Alexa was used by over 100 million people worldwide[1]. Voice assistants are always listening and activate when users use the wake word, *"OK Google"*, *"Hey Siri"*, or *"Alexa"*. Afterward, they process, respond to the user's query, and store a *speech record*; containing a timestamp, transcript, and audio recording.[2] Users of voice assistants integrate these devices into various tasks and activities throughout the day, including managing smart appliances, getting ready for bed, and cooking [21, 172] Thus, speech records allow for a detailed picture of voice assistant users and their routine activities [154].

Previous research suggests that most voice assistant users have an incomplete under-

---

[1] Voice assistant users worldwide, from Smart Speakers Global Market Report (accessed in September 2022)

[2] In response to the European General Data Protection Regulation (GDPR) as of 2020, voice assistants only store the audio recordings if the user has opted in.

standing of how speech records are collected, stored, and processed [111, 154], as well as the security and privacy implications [40, 123]. Bentley et al., [21] collected speech records through Mechanical Turk from 88 Google Assistant users, they concluded that users interact with these devices approximately between 2 and 18 times per day, with an average of 4.1 times per day. Hence, it is difficult for voice assistant users to be aware of what information is stored on their speech records, especially over time. Similarly, Pins et al., [154] argue that it is difficult for voice assistant users to understand the extent of data collection and processing by the system or vendor; as these are introduced through vague and unclear terms of use statements and privacy policies. They developed a prototype to support exploration where 11 users of Alexa and Google Assistant uploaded a copy of their speech records. Moreover, Malkin et al., [123] developed a browser extension to retrieve speech records from 116 users of Alexa and Google Assistant; they used individual speech records as survey prompts and found that almost half of the users (51.7 %) did not know their speech records were permanently stored and the majority (56.0 %) did not know they could review their past speech records. Hence, their understanding of the information and infrastructure behind speech records prevents them from being *adequately* informed when pondering whether to donate (or share) them with researchers.

Within this context, I aim to apply data donation to investigate people's perceptions of their speech records when faced with a comprehensive view as opposed to individual data points (Chapter 5).

4.3. Data Donation Process

In this section, I apply a simplification of the Sensitive Data Donation approach (Fig. 4.2) with a focus on the principles: **(P1)** Balanced Value, **(P2)** Sensitive Data, and **(P4)** Ongoing Consent. I approach these principles as interrelated under the overarching goal of fostering personal data literacy – the necessary knowledge of their data that a person must have to make informed decisions about them – through data donation. I hypothesize that incrementally supporting donors in knowing their data (i.e., increased personal data literacy) yields a return of value to them and enables them to be *adequately* informed about their data at the time of initial informed consent and (re)evaluate their participation throughout the data donation process. Although not the focus of this chapter, this process includes interpretation and contextualization activities that promote the principle of Multiple Knowledge(s) **(P3)**. These aim to support donors in better understanding the different types of information captured through their data, how they relate to their experiences, and reflect on their implications.

In the following, I describe how I applied these principles through the five phases of the data donation approach. The activities, participants, and data collected during each phase are summarized in Figure 4.1.

Phase 1: Identify, Prepare and Communicate

To identify the relevant data sources and types of data, I explored how to request and obtain a copy of the data from Google Assistant, Amazon's Alexa, and Apple's Siri. I decided

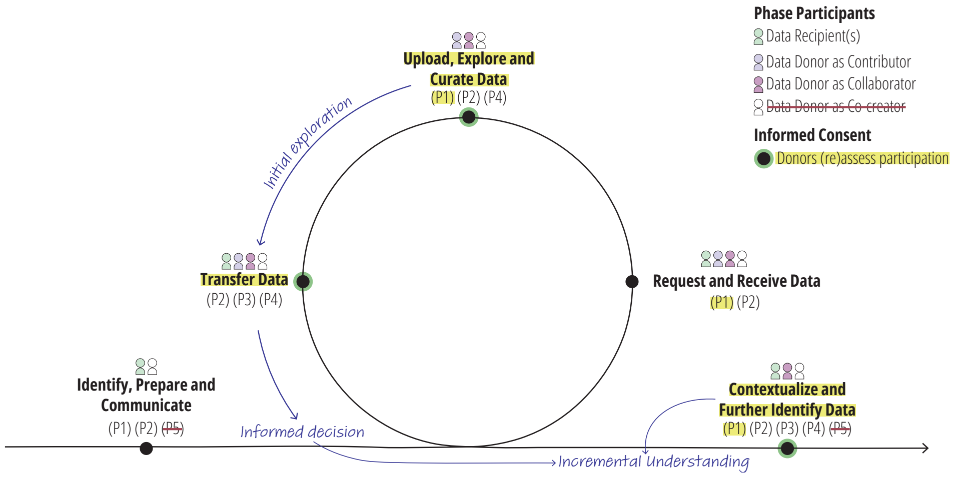


Figure 4.2: Simplification of the Sensitive Data Donation approach for this iteration, highlighting in yellow the main focus of this chapter and crossing out in red the omissions from the approach proposed in Chapter 2.

to focus on Google Assistant, as Google has an accessible Takeout page from where their users can export their data to a downloadable file. At the time, it was less clear how users of Alexa and Siri could request and obtain a copy of their data.

In this iteration, the value-gain strategy is tightly coupled with tools and activities in the data donation platform that invite potential donors to know their data. These are informed by a set of recommendations provided by Pins et al., [154], who developed a prototype that supported the exploration of speech records. Their recommendations to foster awareness and support personal data literacy in this context include: (1) support right from the start, (2) support to structure the data into categories, (3) reconstruct the context, (4) draw attention to unintended interactions, (5) tell users how the vendors (might) see them, and (6) disclose the communications between devices and services (e.g., third-party services and applications).

To prepare the platform to parse and visualize the data, I requested data from Google Takeout from my account and various dummy accounts. I explored the structure of the resulting folder and files. As in Chapter 3, the platform has three open source components that manage (1) the user profiles and authentication, (2) the data storage and sharing, and (3) the donation process. The first two were implemented using TypeScript, and the third was implemented using the Phython web Framework Django. Data was passed between system components using web APIs.

To invite potential donors to participate, between April and June 2022, I reached out to Google Assistant users (e.g., Assistant App, Google Home, Google Nest) worldwide, and we invited them to donate their speech records. I used snowball sampling by periodically posting our ‘call to donate’ across different channels for three months. These included our personal social media (e.g., Twitter and LinkedIn), online communities (e.g., subreddits r/GoogleAssitant and r/GoogleHome, Google Home users on Nextdoor), and local cafes and universities. Additionally, I reached out to people and institutions (e.g., a privacy foundation and an internet podcast) who shared the ‘call to donate’ with their communities through social media, newsletters, mailing lists, and events in which I took part. In total, the ‘call to donate’ had an estimated of 35.000 views.

Phase 2: Request and Receive Data

Between April and June of 2022, I reached out to Google Assistant users worldwide (e.g., Assistant App, Google Home, Google Nest) and invited them to participate in the research by donating and reconstructing the context of their data. For this, I used a combination of convenience and snowball sampling. I advertised the research by periodically posting on my personal social media (e.g., Twitter, LinkedIn), existing online communities (e.g., subreddit r/googleassistant, local mailing lists and newsletters), posting flyers in local cafes and universities, and advertising the research at community events. In the call to donate, I specified who was conducting the research by stating my name and affiliation, what the goal was, who could participate (i.e., adult active users of the Google Assistant) and how (i.e., detailed instructions describing the process to obtain their data from Google). Following the recommendation by Pins et al., [154] of *“supporting right from the start”* I provided detailed visual instructions describing the process. Donors were required to: (1) visit takeout.google.com[3] and log in with their Google credentials; (2) select the type (i.e., speech records), format (i.e., JSON, and MPEG) and size of the data to export; (3) wait *‘a long time (possibly hours or days)’*[4] for the export to complete; and (4) receive an email with a ZIP file containing their speech records. Here, it is important to note that donors receive via email a ZIP file containing a JSON file listing all speech records and several MP3 files (one per speech record). As described by previous research (e.g., [4, 31]) these files and formats are hard to understand for most people.

In this step, I aimed to foster (personal) data literacy by highlighting that (1) speech records are collected and stored by Google and (2) users can obtain a copy of their speech records (and other data).

Phase 3: Upload, Explore, and Curate Data

After receiving an email containing a copy of their speech records, donors could upload their ZIP file to the data donation platform. When uploading their data to the platform, donors could find information about the research project (i.e., project goals and activities) and team (i.e., names, affiliations, and contact information of the researchers) and detailed visuals describing the data management and storage. With this information, I aimed to (1) *adequately* inform donors about the research goals and activities and (2) support them in (re)evaluating their participation. When uploading the data, donors provide initial informed consent to participate in the research. They have not (yet) con-

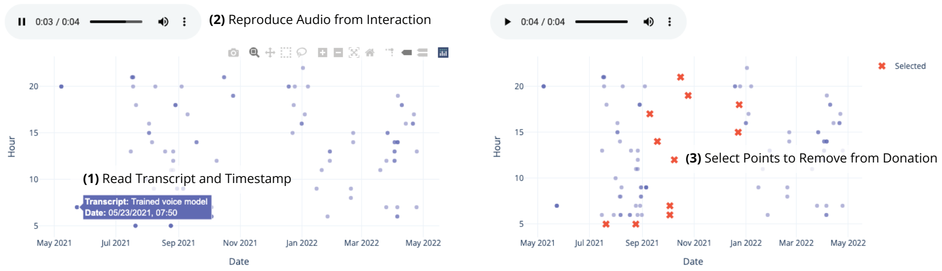
---

[3] Google Takeout Page: takeout.google.com

[4] 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. You’ll receive an email when your export is done."

sented to donate (transfer) their speech records. Thus, data is uploaded and stored in the platform, but researchers do not have access to it until donors have explored it and explicitly (re)evaluated their decision. After uploading their data, donors were invited to explore it, understand it, and (re)evaluate their participation. Here, donors could assess whether and what data to donate (transfer) to the researchers. With this activity, I aimed to foster personal data literacy by enabling donors to delve into the (1) content (2) dimensions (i.e., timestamp, transcript, audio recording), (3) amount, and (4) temporal distribution of the uploaded speech records.

To support and enable exploration, I build upon the data visualization prototype developed by Pins et al., [154], who visualize speech records as points on a graph arranged by time to help people reason about it. I augmented their prototype by allowing donors to *listen* to their audio recordings when hovering at a point, in addition to reading the transcripts and timestamps (Fig. 4.3). Donors could visualize (and listen to) an overview of their speech records over time through an interactive graph where each point represents an interaction with the Google Assistant. In the graph, the x-axis represents the time of the day and the y-axis represents the date. When donors hover over a point, they can listen to the audio recording and read the transcript and the exact date and time of the interaction. Together with the visualization, donors could see the following message: *"I invite you to explore **(and listen to!)** your data by hovering over the dots, each dot represents an interaction with your voice assistant."*



Transcript: Trained voice model
Date: 05/23/2021, 07:50

(a) Hovering over points to explore. **(b)** Selecting points to delete.

Figure 4.3: Example of the interactive graph where potential donors can explore an overview of their speech records over time.

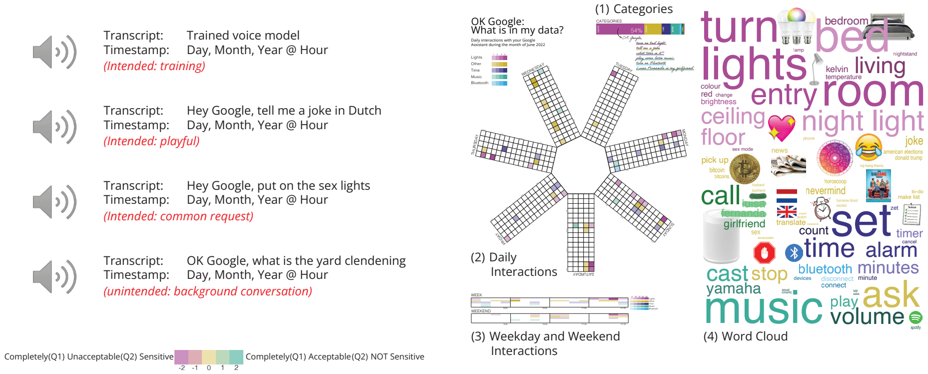
Phase 4: Transfer Data

After exploring their data, donors were invited to (re)evaluate their participation by choosing to: (1) withdraw from participating and delete all their speech records from the platform; (2) consent to donate (transfer) all their speech records to the researchers, who immediately gain access to it; (3) remove specific speech records (e.g., single data point, all data from a given time) from the platform and consent to donate (transfer) the remaining speech records. If donors consented to donate their speech records, I invited them to provide their demographics (i.e., self-described gender, age), location (i.e., city), and information about their Google Assistant (i.e., device type, language).

Phase 5: Contextualize Data

To involve donors in situating and contextualizing their data, I conducted semi-structured interviews with donors who volunteered to participate in this step. Here, I aimed to foster personal data literacy by facilitating donors' exploration and interpretation of their speech records. I developed a *personal data canvas* following the recommendations by Pins et al., [154]: (1) *"drawing attention to unintended interactions"*, (2) *"supporting to structure the data into categories"*, and (3) *"telling users how the vendors (might) see them."*

The *personal data canvas* introduces speech records as single and multiple interactions. First, I focused on introducing the dimensions of the data (i.e., timestamp, transcript, audio recording) through single interactions and *"drawing attention to unintended interactions"* (Fig. 4.4.a). Second, I focused on introducing multiple interactions through a data visualization (Fig. 4.4.b). In doing so, I supported donors to *"structure the data into categories"*, and broadly *"told users how Google (might) see them."*



Transcript: Trained voice model
Timestamp: Day, Month, Year @ Hour
(Intended: training)

Transcript: Hey Google, tell me a joke in Dutch
Timestamp: Day, Month, Year @ Hour
(Intended: playful)

Transcript: Hey Google, put on the sex lights
Timestamp: Day, Month, Year @ Hour
(Intended: common request)

Transcript: OK Google, what is the yard clendening
Timestamp: Day, Month, Year @ Hour
(unintended: background conversation)

(a) Single interactions.

(b) Multiple interactions.

Figure 4.4: Example of a personal data canvas. Shown with permission of the donor

In the visualization, I focused on conveying the information from the timestamps and transcripts of multiple interactions. Specifically, I identified common interactions for each dataset and grouped them into categories (e.g., weather, music, time). I visualized the distribution of these categories throughout the dataset with a bar graph (Fig. 4.4.b$_{(1)}$), and I represented each category with a different color throughout the visualization. Additionally, I presented the number of (daily) interactions for each category per hour of the day and day of the week with a heat map (Fig. 4.4.b$_{(2)}$) where I focused on the 16 hours of the day with more interactions, the start and end times vary by the donor. Similarly, I used a heat map to present the number of interactions of each category per hour of the day during the weekdays (Monday through Friday) and weekends (Saturday and Sunday) (Fig. $4.4.\text{b}_{(3)}$). Finally, I presented a word cloud (Fig. $4.4.\text{b}_{(4)}$) with the most frequent words grouped and color-coded by category, and additional images were visually representing some of the terms. I added the images to make the interactions more prominent and easier to explore.

I used the *personal data canvas* as a prompt during semi-structured interviews where I supported donors in exploring the data, reflecting on their behavior (as captured by the data), and identifying patterns and potential inferences. It was presented as two slides on a screen; if the interviews were in person, the visualization (Fig. 4.4.b) was also printed on A3 paper. The interviews revolved around three stages. First, I invited donors to describe their data donation experience (up to that point, *downloading, uploading,* and *exploring, understanding, and donating* the data) and whether they considered removing points from their donation. Second, I introduced the *personal data canvas* and explored the different attributes of the data (Chapter 5). Here, I supported donors to lead and articulate the interpretation and contextualization of their data. Third, I invited donors to describe any feelings or emotions that emerged throughout the data donation experience (comprising the interview) and discuss their perspectives on the value gained. I phrased value gain in terms of getting something out of the experience or wishing something for a future experience[5]. During the interview, I reminded donors of the possibility of (re)evaluating their participation and withdrawing their donation.

Data Donation Procedure

The resulting data donation process (Fig. 4.5) comprises five voluntary activities. It begins with donors receiving an invitation to participate in the research. It continues with donors requesting and receiving a copy of their data from the Google Takeout page, and uploading it into the data donation platform, where they are explicitly invited to explore and curate them. Here, donors are also invited to decide if and how they want to participate in the subsequent activities. Donors who expressed a desire to participate in an individual session to interpret, situate, and contextualize their data receive an invitation to schedule it in person or online.

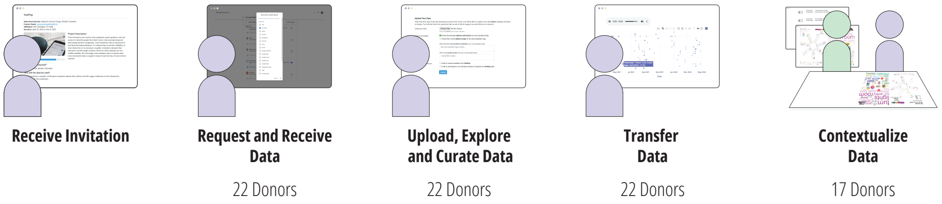


Figure 4.5: Data donation procedure illustrating the various activities, mediated by the data donation platform, in which potential donors participate and how many donors participate in each activity.

---

[5] Example of questions: What did you get out of this experience? What would you like from a data donation experience in the future?

4.4. Applying and Reflecting on Data Donation

4.4.1. Participants

In total, 22 donors (D1-D22), aged 21 to 58 years (mean = 30.8, median = 38), 1 identified as non-binary, 7 identified as female, and 15 identified as male, positively responded to the call by donating their data. Donors were primarily located in the Netherlands (54%), with some based in other countries, including Germany, Italy, Colombia, and Argentina. Obtaining a copy of the speech records, enabled by the GDPR, was also possible for donors outside the EU[6]. 17 donors (5 identified as female and 12 as male) agreed to participate in the follow-up data exploration interview.

4.4.2. Data and Analysis

To reflect on my experience and that of data donors, I generated and collected three types of data. First, when donors first uploaded their data into the platform, I collected information about their initial understanding of Google's data collection and storage practices. Through a short questionnaire. I invited donors to answer the following questions[7]:

Q1: Did you know that Google collects and stores your speech records?

Q2: Did you know that you could download a copy of your Google Assistant speech records?

Q3: What information do you think is in your speech records?

These served as a baseline to determine donors' initial awareness with respect to that of participants in previous studies (e.g., [40, 111, 123]). Second, after donors explored and transferred their data I collected information about their demographics and understanding of their speech records. I invited donors to answer the following questions:[8]

Q4: To what extent do you agree with the following statements?
  - Seeing my data is helpful for understanding how much data my assistant collects
  - Seeing my data is helpful for understanding what types of data my assistant collects
  - Seeing my data is helpful for understanding how long my assistant has been collecting data
  - Seeing my data is helpful for deciding whether to donate them

Q5: What did you learn from seeing your data in this way?

These served to determine donors' understanding of their data after the exploration and their perceptions regarding the usefulness of the interactive visualization (Fig. 4.3).

---

[6] The GDPR applies to the population of the European Union. Yet, in practice, the right to data portability is available worldwide, since international companies rarely limit it by geography [31].

[7] Answers to Q1, Q2 are Yes/No, answers to Q3 are open.

[8] Answers to Q4 are 5-Point Likert scale from strongly disagree to strongly agree, answers to Q5 are open.

Third, I collected information about their overall data donation experience through the semi-structured interviews. I conducted the interviews in English between June and July 2022. Interviews lasted between 35 and 55 minutes; 5 took place in person and 12 via Zoom. I conducted one interview with the two members of a household who share a device ($D9_{a,b}$), the remaining 16 interviews were one-on-one as most donors were single-users of their Google Assistant. The interviews were audio recorded and transcribed. I made an initial transcript using MS Office 365, then manually reviewed and edited it.

My analysis primarily focuses on the answers to the questionnaires (Q1-Q5) and interview transcripts. The answers to the closed questions (Q1, Q2, Q4) are used to illustrate donors' understanding of their data at different points of the data donation process. The answers to the open questions (Q3 and Q5) are combined with the interview data, and analyzed using reflexive thematic analysis [32, 33], within a constructionist framework. I independently read through the transcripts to familiarize ourselves with the data and coded the entire dataset using ATLAS.ti. Through this process, I aimed to capture all the aspects of the data relevant to the data donation experience and the perceived value gained. I independently reviewed the codes and subsequently discussed and grouped them into tentative themes with a co-author of this study, Jacky Bourgeois. I iteratively reviewed and refined the themes.

4.5. Results

4.5.1. Getting my data

Donors gained awareness of Google's data collection practices and the possibility of obtaining a copy of their data. When answering to **Q2** over half of the donors (12 out of 22) indicated not knowing it was possible to obtain a copy of their data.

> "I didn't know, when I saw in the beginning like the instructions about how to download this data. It was the first time for me, and actually, it was very interesting." (D16)

For them, reading the *call to donate* and the *instructions on how to donate* was a way to discover their rights and with them new ways of engaging with their personal data. Additionally, throughout the download process, data went from an abstract entity to a nearly tangible (and material) one that is available and can be explored digitally, seen, read through, and listened to. Data is there, stored somewhere, and accessible (to donors and others).

Data Literacy

Donors' personal data literacy increases by understanding how to intervene and participate in Google's data collection practices and becoming familiar with their individual (data) rights (e.g., right to data portability). This understanding extends beyond the context and scope of the research, as described by $D9_b$.

> "What I found the most interesting was, while I was downloading the data, to see how organized it was. We were only following instructions, so we deselected everything and then we uploaded just the voice commands. But I was

> genuinely excited to see that I could look up my YouTube history, my Google searches, my Google Maps. Everything is in a specific folder, so I think this research empowers you to look stuff up that you otherwise wouldn't." ($D9_b$)

It illustrates how all kinds of (personal) data are stored in structured databases and how these are searchable and accessible upon request. Hence, it presents the opportunity for donors to access and explore data from other Google services (e.g., browsing history, location) and other data holders (e.g., Spotify, Twitter) if only out of curiosity.

Value Gain

Donors engage differently with their (personal) data; an abstract concept that gained clarity and materiality. Speech records are opaquely generated as a product of the (many) interactions between people and their voice assistants and are stored *"somewhere in a cloud"* (D19). They became available and inspectable through the act of 'obtaining a copy'. Moreover, speech records became something donors *have*, *own*, and *control*.

> "Can you say that you own a dataset or that the data about your life is yours if you are not really capable of using it, or donating it, or doing anything about it? Because, I feel like that data [the speech records] is Google's data. I mean, if I don't have a server, if I don't have the technical ability, if I never use it in my daily life, is that data mine? It's about me, but I don't really feel it is mine. Thanks to this project, we kind of gain ownership over that. If I'm a passive agent, I feel like it is about me, but if I'm an active agent it is mine." ($D9_a$)

$D9_a$ articulates the difference between data being *about her* and *hers*. Through data donation donors gained ownership of data that is *about them*; and became *theirs*. Here ownership is not limited to *having* a copy of the data. It extends to **actively** being able to *control guard* and *use* it; for example, by deciding to donate it. Nonetheless, *having* is important. It means donors can (re)use the data as they wish; although donors acknowledge this process is not straightforward and it requires technical skills and resources.

4.5.2. Knowing my Data

Through inspecting and exploring the data donors become aware of *"how much information is stored and the kind of information that is stored"* (D14); and how data relates to themselves and their interactions with Google Assistant. Exploration led to the paradoxical realization that interactions are recorded and become data, paradoxical given that most donors (17/22) indicated being aware that Google collects and stores their speech records when answering to **Q1**. It underlines the gap between using a device *knowing* that it collects and stores (personal) data and *knowing* how data *looks*, *sounds*, and *feels like*.

> "I don't know how to describe it. But one thing for me is like to use it [Google Assistant] and the other thing is like listening to my voice now. By listening to it, you become aware of the fact that this is recorded and was stored somewhere by Google, and it makes you feel a bit unsettled." (D5)

Data Literacy

Donors' personal data literacy increases by realizing that the interactions with Google resulted in data points, indefinitely stored *"Google records all this information for, I don't know how long"* (D18). This realization was described as *surprising* (D2,D14,D17,D21), *crazy* (D14), and *creepy* (D1, D19). Especially considering the *observed* nature of speech records, collected implicitly and in the background of all interactions donors have with their Google Assistant. It is nearly impossible for a person to keep track of every interaction over time, *"I cannot remember what kind of stupid things I've said to Google"* (D18); hence *knowing* how data *looks, sounds,* and *feels like* can lead to uncomfortable feelings, *"Google feels like a stalker"* (D19), and emotions.

> "I remember when I downloaded it, I hadn't made a Google takeout before. I did not know that they stored all this data. And like all the data that is there, that for me was a moment of real emotional response like oh, OK." (D21)

The data visualization supported donors in *knowing* how data *looks, sounds* and *feels like.* In addition, it helped most donors grasp how much data was collected, what types of data were collected, and for how long. Further, it enabled donors to identify behavioral patterns in their data, *"there are patterns on the time of the day when I'm using the assistant, which reflect, somehow, my routine* (D22). In this way, donors gained awareness of data being personal; related to themselves and their behavior and reflecting specific aspects about themselves and their behavior.

Value Gain

The gained awareness and familiarity, often described as *knowledge* and (increased) *understanding,* were perceived as valuable takeaways from participating in the data donation journey. These led to a (better) informed opinion on the data and how it relates to themselves and others.

> "What I took out of this experience is knowledge. It's knowledge about what Google collect[s], the fact that you can download the data, listening to ourselves, it was exciting. And then, having an informed opinion about [data]. Before I had like a fear, and now I have an informed opinion, or at least semi informed." ($D9_a$)

Donors appreciated how data donation enabled them to look 'behind the curtain' and gain knowledge into how Google 'sees' them and how it works; what it listens to, *"listening to it, it's like oh OK, it was recorded"* (D5); what gets recorded, even when it should not, *"it was a personal conversation, I was not aware that* [it] *was being recorded"* (D14); what it does (and does not) understand, *"the transcripts are not always the things that I said"* (D18); and how much and how often it collects and stores data, *"I could realize how much information is on my phone, about me"* (D16).

4.5.3. Knowingly giving away my data and contributing to research

In addition to supporting donors' understanding and exploration of their data, the data visualization enabled them to be aware of what exactly they were 'giving away' to the research, *"it is interesting to hear the recordings because you get a sense that there is a level* *of control of what you're giving away"* ($D9_b$). Here, data became a way for donors to participate in and support research activities; informed and enabled by increased personal data literacy.

> "We were able to listen to the recordings and it's very intimate, but it kind of gave me peace of mind because at the end of the day, what is there is not what I value the most when I think about my privacy." ($D9_a$)

Data Literacy

Donors' (personal) data literacy enables them to be aware of the information donated (or transferred) to researchers. This awareness, invites them to reflect on (and (re)evaluate, if necessary) their privacy boundaries. I aimed to further support donors define their boundaries and *control* what they were 'giving away' by enabling them to remove specific points from their donation. Yet, most of the donors (21/22) decided **not** to remove any points, *"I didn't find anything that was recorded that I thought well, no, no, I don't want to share it"* (D10). This was primarily motivated by how most data points corresponded to simple and mundane interactions (e.g., 'OK Google, what time is it?', 'OK Google, set an alarm'), *"when I looked at the data, it was really, like, 'How is the weather', so, I did not see anything that I would have needed to remove"* (D2).

Value Gain

Through the data donation journey donors *knowingly and actively* contributed to the research. Donors expressed having high regard for scientific research and considered this research *a good cause*, *"when I donate, also money I wanted to go for a good cause. And I'm convinced, it makes me convinced, that my data went for a good cause"* (D2). Hence, the action of contributing to research was perceived valuable and led to positive feelings. The motivation to contribute to research also shaped donors' decisions on whether to donate their data *"I understand the research process and I understand that they* [researchers] *need this kind of information, so I'm completely open to do it"* (D17), and which data to donate, *"I feel that if I share more data, more interactions, it will be more useful for the research. So, my decision was to help as much as I can"* (D14). Additionally, having *knowingly and actively* contributed to this research meant donors had expectations regarding the research progress and its outcomes. These underline opportunities for researchers further provide value to donors by being open and accountable.

> "If I took the time to donate my data to a project, it's because I'm actually interested in it. So, I want to keep knowing what is happening or how is my data being used for." (D1)

4.5.4. Reflecting on my data, my relationship with Google, and myself

During the semi-structured interviews donors gained deeper insights into how (personal) data relates and reflects aspects about themselves and their behavior. These, however, are incomplete and limited by the specific ways in which people interact with their Google Assistant, *"it does give a sort of accurate picture, but it's not the picture that*

I would put together. I think it gives, let's say to certain topics, more, uh, prominence than how prominent they are" (D2). For example, the interaction 'OK Google, turn off the alarm' can indicate when a person wakes up, and the interaction 'OK Google, turn off the bedroom lights' can indicate when a person goes to sleep; broadly reflecting her sleep routine but not providing any insight into what happens in between. Yet, although limited and incomplete, data can support reflection. D22 illustrates the process of reflecting on her routine through the (lack of) data:

> "Realizing about Fridays, that I don't use Google on Fridays. It's like why? And then I thought like yeah, OK so I was not at home on Fridays. I mean, my working day ended like at noon. So, it was like super interesting, I didn't realize about those patterns [before]. Because I know that I wake up at 5:30 and I go to sleep at 10, but then I was never, like, OK the last three Fridays I was doing this." (D22)

Data Literacy

Donors' personal data literacy increases by engaging with the data and reflecting on the nuances of the context captured by it. Through this process, donors realized the many ways data is embedded in and partly reflects their daily lives and interests and the potential inferences that could derive from data. In doing so, they gained greater awareness resulting in a *tipping point* in their perspectives on (personal) data. Data was no longer considered *nothing* (i.e., simple and mundane) and became *something* (i.e., personal and sensitive).

> "When I signed up for this study, I was like, OK, my Google home data? I don't think there is anything to find in it, so why wouldn't I share [it]? And even after ticking and ticking through [in the data visualization on the data donation platform], like OK, what am I sharing? I was still convinced. [...] And now I'm surprised, it's not like there is nothing in the data. For a brief moment, I was even like, OK, I'm glad that nothing more surprising came out from there [laughs]." (D2)

Increased awareness, and the change of perspective derived from it, resulted in the intention to change how donors interact with their Google Assistant to minimize (sensitive) data collection, *"I learned about myself, but also, I think I would be a little bit more careful with what I'm going to ask Google from now"* (D17). Additionally, it enabled donors to put their privacy concerns into perspective. For some, it led to the realization that Google is not that bad, *"I mean it might be able to tell if I'm sleeping, or at what time do I wake up, but like the things that are really important for me in terms of privacy are not there and so I was kind of relieved"* ($D9_a$). For others, it led to the realization that Google collects and stores too much data and is too creepy, *"Google feels like a stalker"* (D19). It prompted donors to reconsider their relationships with digital technologies as well as their participation in this research and future research activities that entail personal data sharing. Still, when invited to (re)evaluate their decision to donate at the end of the data donation journey none of the donors (0/22) wanted to withdraw their donation.

> "It was a learning process for me. To understand better what information I

> share and what information I let Google know about me. And it's something that I believe we need to improve, because as I told you, there are things in the [personal data canvas] that you did that I don't want to be sharing with anyone, even with Google." (D14)

Value Gain

Donors appreciated the multiple viewpoints set in place for them to explore their data. These allowed them to engage with data through different lenses and direct their attention to specific details, including amount, temporal distribution, (un)intendedness, aggregation, and potential inferences. The knowledge gained during the process was incremental. It led donors to challenge their assumptions (e.g., my data is just 'How is the weather') and account for the personal nature of the data; that relates to and reflects aspects of their behavior, especially when combined and considered over time.

> "Even though the data, if you take 1 by 1, is not something important or relevant. Those behavioral patterns are quite sensitive, like what you did [with the personal data canvas], like you can infer what my days look like a little bit." (D1)

The incremental knowledge donors gained throughout the process was enabled and supported by the guidance and materials I provided. These were highly appreciated and perceived as valuable takeaways in themselves. Especially the *personal data canvas* (Fig. 4.4.b), *"it's a very nice visualization. Specifically, I appreciate the visualization"* (D17). It provided a structure to interpret the data, *"this distribution* [Fig. 4.4.b$_{(1)}$], *I just love it"* (D21); and prompted donors to reflect on their behavior.

> "It is super interesting seeing all this data like classified, as it was. [4.4.b$_{(2)}$] And as I said, it helped like identifying the patterns of my day, of my routine." (D22)

The *personal data canvas* became a tangible outcome of the process, which we gave to the donors at the end of the interview, *"for me, getting back this visualization is useful. It really makes sense and it can actually tell something about me"* (D2). Beyond the material, donors found value in the guided exploration of their data. This process offers the opportunity to bring to light (personal) insights and disentangle the abstract construct of (personal) data. Hence, it could be relevant even beyond the context of our research.

> "I would pay for this. I would pay to have this kind of consultancy. Like not having to go into Amazon, and Spotify, and Facebook, and Twitter to understand. I think I would be one of the people that would pay for someone to go, explore and tell me from the platform's perspective how is my life. Kind of like people who pay for astrology? For other people to tell you who you are, you are fearless, you are... [laughs] Definitely, if I had someone doing data explorations with me, that would be something I would be interested in paying [for]." ($D9_a$)

4.5.5. Shortcomings of the Data Donation Experience

Data Donation: A New Experience

Data donation is a new experience. It enables donors to engage with their personal data and in doing so, open their personal space to others (i.e., researchers). This process can be confronting and uncomfortable. In the context of *speech records*, which correspond to *observed data* that is generated implicitly from people's behavior, this process is also a window to the unknown and the unexpected that is entangled in and captured by the data.

> "I must say that I feel a bit **naked**. In the sense that this [personal data canvas] tells a lot about me, much more than I expected" (D17)

Additionally, data donation entails donors to 'give away' a copy of their personal data. Although this is something donors do *knowingly and actively*, it is a leap of faith. Meaning, donors 'give away' their data to researchers in a specific context and under certain conditions. But they cannot guarantee that researchers will use their data in said context and under said conditions. They can only trust.

> "I'm relieved there's not more data out there. And that just triggered also thoughts in me about, like, how this data donation is really cool, but I'm also giving you permission to like do whatever with it." (D2)

Data Donation: A (Not So) Easy Process

Data donation, as operationalized in this research, entails a journey that comprises several steps and interacting with at least three digital platforms (i.e., Google Takeout, email provider, data donation platform). This was by-design as we aimed to support and promote awareness throughout the process. Yet, it meant that some donors faced difficulties and were confused.

> "It's not exactly super confusing, but it's confusing enough that I feel like I'm not sure if I'm doing this right when I'm downloading the data, when I'm uploading the data, I was usually expecting something different to happen." ($D9_b$)

The complexity of the journey might discourage potential donors, especially those less experienced in interacting with digital technologies. Hence, there is a need to balance the awareness gained throughout the process with its complexity.

Data Donation: (Not) A Window Into-The-Wild

On 2020 Google announced that the Google Assistant was no longer collecting and storing the audio recordings from every interaction unless users had explicitly opted-in to allow voice data collection. I was aware of this and when disseminating the *call to donate* I invited potential donors to check their configuration and, if necessary, opt-in to allow voice data collection and interact with their Google Assistant for a couple of weeks or months before donating their data. Four donors (4/22) opted-in to allowed voice data collection and generated data while being aware that it was going to be used for our research, *"when we turned it on, we were like wow, Alejandra is going to listen to this"* ($D9_a$).

I instructed these four donors to interact naturally with their Google Assistant. Yet, being aware of the research led to interesting interactions (e.g., ‘ OK Google, what do you know about Alejandra Gomez Ortega?’) and behaviors.

> “It was also interesting because I had the settings on and I had some guests. So first, I thought well, I have to make a little note [saying] that you can be recorded. Then I forgot, and after the visit I thought maybe I have to inform them. So, I did informed them afterwards.” (D10)

Previous research claims that the data that is available through data donation ‘is embedded into the donors’ routine and is not attached to a research project or a research instrument, thus [is] less prone to observation bias’ [75]. This research, where this partially applies (18/22), illustrates how data donation is limited by the infrastructures in which data is embedded in.

Data Donation: A (Not So) Individual Journey

In this project I received 22 datasets and identified more than one speaker being recorded in all of them (22/22). Different speakers were more frequent in multi-user environments (e.g., $D9_{a,b}$), where more than one person shares a physical space where a device is present. Yet, there were still recorded in single-user environments where other people (e.g., occasional visitors) are around (e.g., D2). Hence, although donations were made by a person *knowingly and actively* giving away her data –except for $D9_{a,b}$ who gave away *their* data– other people were indirectly involved and information about them (e.g., their voice) was donated.

> “If I’m trying to donate data to some project, then I then I would like to donate data that I know it has picked up from me, and well...” (D2)

Donated data captures people’s relationships and interactions with others (e.g., partners, family members, friends, neighbours), who are present in the dataset. Moreover, it accounts for people’s relationships with others (e.g., ‘OK Google, call *my mom*’, ‘OK Google, *my girlfriend* is [name]’). Hence, data that is donated could indirectly involve other people, who are captured by the data, *“I thought, well, it is my uncle’s privacy, I don’t want to compromise, someone else’s privacy”* (D10). Underlining the importance of accounting for them. Although doing it is not necessarily trivial. Donors expressed having informed others (e.g., partners, family members, friends) of the data collection and the data donation, *“I would want my partners OK that this data is being shared* (D11). But in some cases, figuring out who to inform can become a puzzle.

> “Oh, first of all, it’s not even me. [laughs] I don’t think I know who [it is]. 10th of December. Because, it [the Google Assistant] is close to a window, so, but I don’t think the window was just open. Like stuff from the street.” (D2)

4.6. Discussion

In this chapter, I explored to what extent supporting donors in knowing their data creates value for them and enables them to be adequately informed about their data and (re)evaluate their participation in research. I led donors on a journey of engaging with and understanding their speech records, collected by Google Assistant in the background of their every interaction. Through this process, I received datasets from 22 donors and situated and contextualized 17 of them.

In this section, I abstract my experience from the research activities and translate it into a reflection on the principles of Sensitive Data Donation and the actors involved in data donation. Furthermore, I provide recommendations for promoting personal data literacy trough data donation.

4.6.1. Reflections on the Actors

- **Data:** They evolve through data donation. Understanding data donation as an encompassing process implies understanding data as dynamic rather than static (always accessible and reusable). The data donation process is centered around the voluntary transaction of personal data and is further shaped by its exploration and understanding. Through data donation, personal data goes from an abstract entity to a concrete one that is situated and deeply entangled with people's behaviors and intimately relates to and reflects them. Donors and recipients should harness the dynamic nature of data, as it enables meaningful collaborations shaped by both parties and from which both parties benefit.

- **Donors:** They enable data donation by transferring their data. The data donation process invites them to actively contribute to research. As part of the process, they have the opportunity to explore their personal data and gain incremental and situated knowledge about their content. Through these activities, donors gain a return of value through a better understanding of their data (i.e., increased personal data literacy). They gain ownership of their data, from being about them to theirs (i.e., they own an actual copy). Furthermore, they are empowered to make better-informed decisions throughout the research and beyond. Moreover, as described by previous research (e.g., [75, 140]), data donation should support their agency and autonomy by enabling them to set granular boundaries and define the terms of their participation.

- **Recipients:** Designers and researchers initiate data donation by inviting people to participate in their research. The data donation process requires them to intentionally engage with the personal and dynamic nature of the data. As part of their study design or design process, they must contemplate: (1) facilitating data exploration and interpretation; (2) supporting donors in (re)evaluating their participation and (re)defining their boundaries (e.g., curating their data); (3) creating opportunities for donors to gain value; and (4) nurturing communication and accountability. In turn, these activities could contribute to their understanding of people's relations with their data and trigger reflection in their research and practice.

4.6.2. Reflections on the Principles

Balanced Value: Knowing Data

In this chapter, I hypothesized that a better understanding of the data (i.e., increased personal data literacy) yields a return of value for donors. The results illustrate that most donors perceive the incremental knowledge regarding their (personal) data as valuable, even if uncomfortable or creepy. In addition, most donors appreciated the *empowerment* derived from acquiring new knowledge that can be applied to other contexts and gaining *ownership* of their data.

Fostering a better understanding of the (donated) data on a personal (i.e., how it relates to and reflects their behavior) and infrastructural (i.e., how data is collected, stored, and regulated) level is a promising avenue for donors to gain value from engaging in data donation. It is a value-gain strategy that harnesses the abilities and strengths of researchers (e.g., shaping the data and translating it into something graspable). It can support (better) informed transactions and collaborations through data. Additionally, it can trigger researchers and designers to better understand and engage with people's entanglements with their data. These, in turn, can invite us, researchers, to reflect on our practice (e.g., *how do we do research?*) [81] and prompt us to design digital systems that invite different relationships with (personal) data. Hence, it is an opportunity for mutual benefit that is feasible and relevant for both parties.

Sensitive Data: Incremental Understanding

Current framings of data donation (e.g., [140, 162, 180]) are focused mainly on the instance of the voluntary transaction of data (i.e., when a person donates her data). The results of this chapter illustrate how the sensitive nature of the data calls for an *encompassing process* around the transaction of sensitive data. This process should dynamically and iteratively support and invite donors to: (1) access meaningful information about how their data is used and handled; (2) explore and understand their data on a personal and infrastructural level, and (3) (re)evaluate and (re)assess their boundaries and participation.

Approaching data donation as an encompassing process de-emphasizes the transaction of sensitive data as the primary instance of informed consent. Yet, it still is the entry point into the process. I recommend concrete actions around this instance. *Before* the transaction of data, researchers and designers should inform donors about the goals and activities and enable them to explore and familiarize themselves with the data; grasp its content and characteristics. In this chapter, I encouraged exploration *before* the transaction of data through an interactive graph where donors could engage with their speech records over time. *After* the transaction of data, and *throughout* the process, researchers and designers should support donors' incremental understanding of the data. I supported the incremental understanding of the data by facilitating interpretation and articulation during semi-structured interviews prompted by the *personal data canvas*. Although these are not the only ways, they illustrate how these activities could manifest in practice. In addition, researchers and designers should remain available and accountable to data donors. Accountability is important as data donation is a *"leap of faith"*.

That is, once researchers and designers gain access to the donated data, nothing (beyond ethics) prevents them from using it in a different way than agreed. Therefore, accountability could provide reassurance and increase donors' trust in the process.

Ongoing Consent: Dynamic Datasets

An underlying outcome of the data donation journey introduced in this chapter was illustrating the gap between (1) *knowing* that voice assistants collect and store data; (2) *knowing* what data *looks*, *sounds*, and *feels* like; (3) and *knowing* how it relates to a person and reflects her behavior. This gap challenges the notion of being *adequately* informed in situations that involve a transaction of digital-trace data. It echoes the limitations of informed consent highlighted in previous literature [9, 36, 133, 139]. Breuer et al., [34] argue that it is possible to obtain informed consent from data donors. They rely on a template proposed by [183] listing the information that should be provided for consent to be informed. It includes: (1) why data is being collected; (2) what will be done with the data; (3) what data will be collected; (4) how data will be stored; and (5) what the risks of disclosure might be. Although such information is important and must be provided, I argue it is not enough for people to be *adequately* informed; that is, to have a clear understanding on *what* and *how* their data will be used [133]. The information provided is the mere formalization of a unilateral transaction of data, an element that remains opaque and abstract. Stating *what* data will be collected is widely different from supporting people in knowing and understanding data and its implications.

The results underline how the content of personal data is not only opaque to donors but also to researchers. I was technically equipped to understand and analyze the received speech records and prepared to encounter contextual insights. However, I could not have anticipated the information that unravelled through the process. Thus, I propose an iterative and incremental process of supporting participants in *knowing* (what) data that invites them to (re)evaluate and (re)assess their participation, preferences, and privacy boundaries. Hence, improving the informed consent process towards one that is ongoing and dynamic [101]. This process requires researchers and designers to reconsider their relationships with participants and adopt new procedures that harness the dynamic nature of the data, continuously changing through the actions and preferences of participants [75].

4.6.3. Recommendations: Fostering Data Literacy in Data Donation

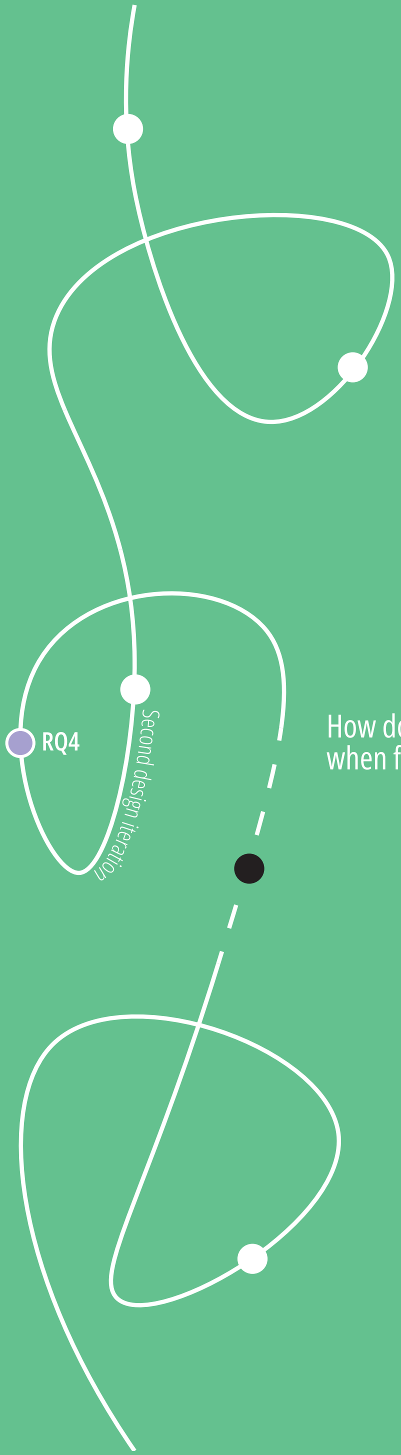
Data literacy is a widely used concept often associated with specific skills and abilities (e.g., combining data, visualizing data). I provide recommendations to foster data donors' personal data literacy. Concretely, I focus on the information that should be communicated to support informed data donation (and broadly data sharing) decisions.

1. **How is data collected?** Provide information on data collection (and storage) and their relationship to people's behavior and interaction with digital products and services. For example, a speech record is generated and stored every time a person interacts with her Google Assistant.

2. **How can I access my data?** Provide information regarding specific data collection practices and policies that support people in navigating the process of gaining *ownership* of their data. For example, speech records collected by Google Assistant are available upon request via Google Takeout.
3. **What exactly is (in) my data?** Enable playful ways for people to understand the content and characteristics of the data, what they *look*, *sound*, and *feel* like. For example, illustrate how often a data point is generated, what information it contains, and how many are on the entire dataset.
4. **What makes data about me?** Illustrate how data relates (and reflects) to people, their behavior, and experiences. For example, by shaping data in a way that underlines behavioral patterns and facilitates interpretation.
5. **What data I'm donating?** Support awareness of the data (and personal information) being shared and its potential implications. Enable and facilitate setting boundaries and identifying potentially sensitive elements. For example, allow for granular data-sharing decisions throughout the process.

4.7. Chapter Takeaways

In this second iteration, I focused on integrating into the data donation process opportunities for donors to (re)assess their participation facilitated by an increased understanding of their data. I approached the principles of Balanced Value **(P1)**, Sensitive Data **(P2)**, and Ongoing Consent **(P4)** as interrelated. Thus, I supported donors in knowing their sensitive data as a way to provide value to them and invite them to (re)evaluate their consent and participation. To do so, I incorporated into the data donation journey activities that supported donors in exploring their data before and after the transaction: (1) autonomously through an interactive tool in the data donation platform, and (2) collaboratively through the personal data canvas. Expanding on the collaborative exploration process, Chapter 5 delves into the perceived characteristics of the data: sensitivity, and intimacy. These activities can guide the implementation of dynamic and ongoing consent practices; serving as explicit instances for donors to (re)assess their participation in the latter phases of Sensitive Data Donation. The results of this chapter indicate that supporting an incremental understanding of the data enables donors to be *adequately* informed. Additionally, the knowledge and empowerment derived from the increased understanding of the data and how it relates to and reflects a person's behavior are perceived as valuable takeaway from the process. Challenges remain around donors' opt-out decisions throughout the process.



How do donors perceive `sensitivity' and `intimacy'
when faced with a comprehensive view of their data?

5

Conceptualizing Sensitivity and Intimacy through Data Donation

Abstract

In this chapter, I explore how donors perceive the ‘sensitivity’ and ‘intimacy’ of their data, specifically their Google Assistant speech records, when faced with a comprehensive overview (RQ4). I take a step away from the design iterations to focus on a specific phase of the data donation journey: interpreting, situating, and contextualizing data. I describe the contextualization and interpretation activity introduced in Chapter 5 and delve into donors’ perceptions of their data as they gain an incremental understanding of them. Speech records are an interesting context to explore donors’ perceptions as there is a misalignment between experts, regulators, and users on whether and what data is ‘sensitive’, partly due to how data has been presented to users; as single interactions. Thus, I present data to donors comprehensively as single interactions, patterns, and inferences. Through this process, donors experience a tipping point in perceived sensitivity and intimacy as they delve deeper into their data and the information derived from it. I conclude this chapter by proposing a conceptualization of sensitivity and intimacy that accounts for the tipping point and the fuzzy nature of data and must disentangle from them. I discuss the implications for Sensitive Data Donation.

This chapter draws on the following publication: **Gomez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2023) What is Sensitive about (Sensitive) Data? Characterizing Sensitivity and Intimacy with Google Assistant Users In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23).

	Alejandra	(Potential) Donors	Research Data
	- Familiarizes with and classifies data - Creates personal data canvas - Prepares CI-inspired interview		
Phase 5	- Participates in interview - Transcribes interview - Analyses interview	- Participate in interview in-person or online - Answer questions Q1-2 - Interpret data - Situate data	- Data annotations - Interview transcript

Figure 5.1: Overview of the case study showing the involved people and activities mapped around the five phases of Sensitive Data Donation.

5.1. Introduction

This chapter investigates people's perceptions of the sensitivity and intimacy of their Google Assistant speech records, introduced comprehensively as single interactions, patterns, and inferences. Aligned with the principle of Sensitive Data **(P2)**, I aim to determine whether people are concerned about their speech records, the information they capture, or their potential disclosure; and what about speech records makes them 'sensitive' and 'intimate' data. Broadly: What is sensitive about (sensitive) data? I build upon the data donation process introduced in Chapter 4. Thus, in this chapter, I don't describe a design iteration. Instead, I describe the semi-structured interviews where I contextualized data in collaboration with donors, the design decisions that shaped the personal data canvas, and donors' perceptions (Fig. 5.1).

I continue to explore the intimate context of the domestic space that is shared with a voice assistant. Users of voice assistants interact with them through their voice, which is considered a convenient and natural way to communicate, more intuitive than clicking or typing [173]. In doing so, users integrate these devices into their routines and physical spaces, including pockets, bedrooms, living rooms, and kitchens [21, 154, 172]. Voice assistants collect personal data that is (1) volunteered, explicitly created and shared by a person (e.g., name and date of birth when filling out a registration form); and (2) observed, implicitly collected and captured by recording the actions and behavior of a person (e.g., timestamped speech records and textual transcriptions, generated and stored from each interaction) [31, 213]. Thus, voice assistants collect personal data containing various information about a person, some potentially sensitive.

The GDPR defines sensitive data as a special category of personal data that includes racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; and data concerning a person's sex life or sexual orientation [65, Art. 9]. Outside of the GDPR, the term sensitive data is used more broadly by Human-Computer Interaction scholars, referring to information that is stigmatized (e.g., mental illness [8], HIV status [208]), should not be disclosed [114], could be easily compromised if disclosed [198], or whose disclosure could expose people and lead to inferences about their behavior and experiences [113, 168]. In addition, HCI scholars have introduced the term intimate data referring to personal information about intimate practices (e.g., cooking, sleeping, showering) taking place in intimate spaces (e.g., home) [87, 109] and bodily experiences (e.g., menstruating, urinating) [6, 92].

Speech records are collected in intimate spaces, namely people's phones and homes [109, 177]. They include data from all kinds of interactions, most of which are simple and mundane (e.g., "OK Google, set an alarm"[1]) [21, 154, 172]. Do they contain 'sensitive' or 'intimate' data? Let me say a person interacts with a Google Assistant while showering, "OK Google, play music on Spotify", could this be sensitive or intimate information? If so, why? Is it her voice? The voice is biometric information and is considered sensitive under the GDPR if used to identify a person. Is it the content? The context? She is in the shower, an intimate space. Is it the aggregation of multiple interactions that could lead

---

[1] The examples of interactions I present in this paper come from the 8735 speech records I received.

to inferences about her showering routine?

Privacy experts and regulators argue that speech records correspond to sensitive information (especially the audio recordings) and emphasize that sensitive information about people's behavior can be inferred from them [144, 154]. Yet, voice assistant users seem to express a contrasting opinion. Previous research has shown that voice assistant users consider that individual speech records do not correspond to sensitive information [111, 123]. These studies focused on single interactions people had with their voice assistants, asking them about the acceptability of randomly selected interactions [123], and prompting them to reflect on interactions logged in a diary [111]. Speech records illustrate ambiguities around the interpretations of sensitive data. These are partly due to the focus on data as single interactions, disregarding data aggregation and potential inferences, of which voice assistant users have limited understanding [40, 111, 154]. Thus, there is a need to investigate how people articulate the notions of 'sensitivity' and 'intimacy' of speech records when introduced comprehensively, considering patterns and inferences in addition to single interactions.

5.2. Intimate Context: Misalignments in the Home

Users of voice assistants integrate these devices into their daily and social lives at home and outside the home [21, 154, 172]. Interactions start with a wake word, "OK Google", "Hey Siri", or "Alexa", which the always-on [123] voice assistant recognizes, processes, and stores. Hence, every interaction generates a speech record, which contains a timestamp, indicating the date and time, a transcript and an audio recording[2]. Thus, speech records allow for a detailed picture of voice assistant users and their routine activities [21, 154, 172]. This picture can include, for example, how someone lives in or near a city ("OK Google, what is the weather like in [city]?"), wakes up early in the morning during the week ("OK Google, set an alarm for 6:00"), and plays a podcast on Spotify first thing in the morning ("OK Google, open Spotify").

Research on the field of user intent mining (e.g., [35, 159, 173]) has been used to categorize these interactions. Broder [35] proposes a taxonomy of web search with three key needs that are also applicable to interactions with voice assistants: (1) informational, where the purpose is to obtain information, (2) transactional, where the purpose is to perform an activity (mediated by the voice assistant), and (3) navigational, where the purpose is to invoke a third party application. Similarly, Qu and colleagues [159] propose a taxonomy that includes follow-up questions and greetings/gratitude. While Shani and colleagues [173] argue for considering playful interactions as well (e.g., "OK Google, tell me a joke"), and propose a taxonomy of playful interactions including relief (e.g., "OK Google, order poop"), incongruity (e.g.,"OK Google, give me a high five"), and superiority (e.g., "OK Google, you suck").

Due to the nature of interactions with voice assistants and where they take place, speech records could contain and lead to potentially sensitive and intimate information about users. Yet, previous research (e.g., [111, 123, 124, 130, 144]) illustrates nuances and mis-

---

[2] As of 2020 the Google Assistant only stores the audio recording if the user has opted-in.

alignments around perceived sensitivity and intimacy. For instance, a longitudinal study of ubiquitous surveillance in the home concluded that people consider audio recordings among the two most sensitive and disturbing data, the other being video recordings [144]. Similarly, researchers have documented privacy concerns around voice assistants in the home, including that they listen 24/7, can record private conversations, and collect personal information [41, 111, 124, 130]. In contrast, Malkin and colleagues [123] found that *'on the whole, data currently stored with voice assistants is not considered sensitive'* by users. Their research is limited to individual interactions (i.e., a single speech record) and does not consider what could be inferred from multiple interactions. Lau and colleagues [111] conducted a diary study with voice assistant users and obtained similar results, *'users did not consider their speech records sensitive and did not make use of privacy controls'*.

Moreover, speech records could be (re)used to learn potentially sensitive information about a person, which could be misused. From the audio recordings alone, it is possible to estimate the age of the speaker [182], recognize her emotions [41, 143], identify activities such as laughing, crying, and eating [41, 160], diagnose a broad range of psychiatric disorders, including depression and schizophrenia [41, 118], and determine the size and shape of the room where the device is located [41]. In addition, speech records generate in a shared physical space, where there is a physical closeness between a user and their voice assistant [87, 146]. Hence, they could contain intimate data from the intimate activities occurring within that space.

5.3. Methodology

5.3.1. Collection of Speech Records through Data Donation

I conducted a data donation campaign (Chapter 4) to collect speech records generated in-the-wild. Between April and June 2022, I reached out to Google Assistant users (e.g., Assistant App, Google Home, Google Nest) worldwide, and invited them to donate their speech records. I used snowball sampling by periodically posting the 'call to donate' across different channels for three months. These included my personal social media (e.g., Twitter and LinkedIn), online communities (e.g., subreddits r/GoogleAssitant and r/GoogleHome, Google Home users on Nextdoor), and local cafes and universities. Additionally, I reached out to people and institutions (e.g., a privacy foundation and an internet podcast) who shared the 'call to donate' with their communities through social media, newsletters, mailing lists, and events in which I took part. In total, the 'call to donate' had an estimated of 35.000 views.

I asked interested users to download a copy of their speech records from Google, upload it to the data donation platform, and decide whether to participate in the interview. Due to Google's 2020 policy change requiring users to opt into voice data collection, four donors had to opt in and collect data for a couple of months before donating it. These donors knew their interactions with their Google Assistant would be used for this research.

Participants

22 users of Google Assistant (referred to as donors D1-D22) volunteered to participate in the research by donating their speech records (N = 22, 1 identified as non-binary, 7 as female, and 14 as male). They ranged in age from 21 to 58 years (mean = 30.8, median = 38). Out of these, 17 ($n_{interview}$ = 17, 5 identified as female and 12 as male) agreed to participate in the interview. Donors were located in the European Union (EU) and South America. Obtaining a copy of the speech records, enabled by the GDPR, was also possible for donors outside the EU[3].

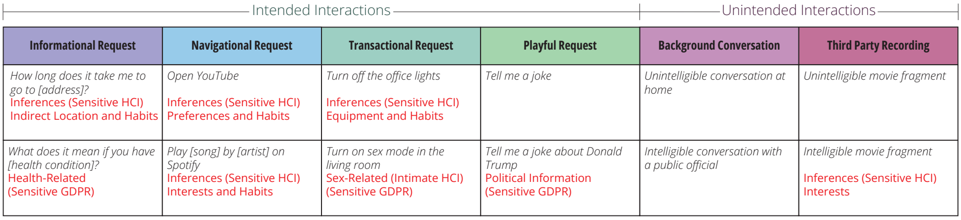
5.3.2. Initial Analysis: Familiarization and Classification

Throughout the data donation campaign, a co-author of this study, Jacky Bourgeois and I independently analyzed the donated speech records to gain insights into common queries and patterns within individual datasets. The purpose of this activity was twofold: (1) identify the characteristics of speech records and (2) recognize the relevant attributes to structure the interviews. In total, we received 8375 speech records (i.e., individual interactions with a Google Assistant), although the number of speech records obtained per donation varied widely. It depended on how long and how often donors interacted with their Google Assistant. The largest dataset contained 5766 speech records, and the smallest had 24. Although one donor contributed a significantly larger dataset, this did not influence how we used and analyzed speech records in this study, i.e., as prompt and support for qualitative exploration. We do not derived quantitative insights from the donated speech records; other than a descriptive overview. In addition, we found the same types of interactions and information in each of the datasets received. Therefore, we present the same type of prompts during the interviews, based on their personal datasets, to donors. Jacky and I systematically listened through each speech record and classified them as intended (i.e., the speaker used the activation command) or unintended (i.e., the speaker did not use the activation command). We then mapped the intended interactions into one of four categories. These came from Brode's web search taxonomy [35], integrating playfulness, common among voice assistant users [21, 111, 173]. The categories were: (1) informational request (e.g., "OK Google, what's the weather like?"), (2) transactional request (e.g., "OK Google, turn on the lights"), (3) navigational request (e.g., "OK Google, open Spotify"), and (4) playful request (e.g., "OK Google, tell me a joke"). Additionally, we mapped unintended interactions into one of two categories, according to the content of the audio recordings: (1) (background) conversation and (2) third-party audio (e.g., TV or radio). Finally, we determined if each speech record contained sensitive (as defined in the GDPR or understood in HCI) or intimate (as understood in HCI) information and annotated it. When sensitive or intimate information derives from inferences (e.g., "OK Google, How long does it take me to go to [address]?"), we annotated potential inferences (e.g., goes to [address] every Monday) considering similar interactions throughout the dataset and the state of the art on the different in-

---

[3] The GDPR applies to the population of the European Union. Yet, in practice, the right to data portability is available worldwide, since international companies rarely limit it by geography [31]. The Google Takeout dashboard (takeout.google.com) is available to users worldwide.

formation derived from interacting with personal voice assistants (e.g., [41]). Figure 5.2 illustrates this process.



Intended Interactions				Unintended Interactions	
Informational Request	**Navigational Request**	**Transactional Request**	**Playful Request**	**Background Conversation**	**Third Party Recording**
How long does it take me to go to [address]? Inferences (Sensitive HCI) Indirect Location and Habits	*Open YouTube* Inferences (Sensitive HCI) Preferences and Habits	*Turn off the office lights* Inferences (Sensitive HCI) Equipment and Habits	*Tell me a joke*	*Unintelligible conversation at home*	*Unintelligible movie fragment*
What does it mean if you have [health condition]? Health-Related (Sensitive GDPR)	*Play [song] by [artist] on Spotify* Inferences (Sensitive HCI) Interests and Habits	*Turn on sex mode in the living room* Sex-Related (Intimate HCI) (Sensitive GDPR)	*Tell me a joke about Donald Trump* Political Information (Sensitive GDPR)	*Intelligible conversation with a public official*	*Intelligible movie fragment* Inferences (Sensitive HCI) Interests

Figure 5.2: Example of initial analysis. The speech records (*italic*) are categorized and annotated (red) with respect to sensitivity and intimacy.

5.3.3. Semi-Structured Interviews: Contextualizing Data

I conducted semi-structured interviews with 17 donors, prompted by their donated speech records. The interviews aimed to facilitate donors' reflection and exploration of their speech records and to capture thick and nuanced insights about sharing, sensitivity, and intimacy. I conducted the interviews in English between June and July 2022. Interviews lasted between 35 and 55 minutes; 5 took place in person and 12 via Zoom. I conducted one interview with two members of a household who share a device ($D9_{a,b}$), the remaining 16 interviews were one-on-one as most donors were single-users of their Google Assistant. During the interview, I aimed to comprehensively explore speech records; hence I delved into individual speech records (perceived as not sensitive by users [111, 123]) as well as patterns and inferences derived from multiple speech records (not considered in previous literature [111, 123, 154]). I designed a *personal data canvas* for each interviewee (Fig. 5.3) containing individual interactions as audio clips (Fig. 5.3.a) and a visualization of the data (Fig. 5.3.b) to support the interviews.

Interview Protocol

The interview protocol draws from the theory of privacy as Contextual Integrity (CI) proposed by Helen Nissenbaum. It suggests understanding privacy in terms of the appropriateness of information flows according to social or cultural norms and grounded in specific contexts [137]. Information flows are described according to five parameters: (1) subject of the information, (2) sender of the information, (3) attribute, describing the type of information, (4) recipient of the information, and (5) transmission principle, stating the condition under which the information flow is permitted. For instance, a Google Assistant user (subject) might be comfortable with Google employees (recipient) reviewing the audio recordings (attribute) from her Google Assistant (sender) if she has opted-in for the collection and revision of voice and audio (transmission principle); but not with the police doing so (a different recipient and privacy violation).

CI is considered an appropriate framework to understand people's privacy norms and has been operationalized through large-scale surveys in different contexts [1, 11, 125, 178]. Broadly, CI surveys inquire about the acceptability of information flows, illustrated through scenario-based vignettes with varying parameters (e.g., a different recipient or transmission principle). For instance, Apthorpe and colleagues explored the acceptability of information flows in smart home IoT devices, including fitness trackers, thermostats, and personal assistants, among others [11]. While Abdi and colleagues explored the acceptability of information flows in smart home personal assistants and considered several types of data, including voice recordings [1].

Inspired by the CI scenario-based inquiry, the interviews revolved around six attributes introduced with examples from each interviewee's dataset and presented through the *personal data canvas*. Specifically, I focused on introducing the different types of single interactions present across the received datasets. These encompass a wide range of use cases and contexts: (1) a neutral and de-contextualized training interaction; (2) a simple yet telling playful interaction; (3) a common request interaction (informational, navigational, or transactional), illustrating common usage patterns and contexts of use; and (4) an unintended interaction that signals an unexpected device operation. In addition, I focused on introducing the main types of information that can derive from the aggregation of multiple interactions: (5) patterns and (6) inferences. These attributes are not mutually exclusive categories, but rather illustrative examples of various information grounded in the received datasets. I selected these attributes to provide a comprehensive overview of the speech records and the underlying information they contain, derived from the initial analysis. For each attribute, donors answered two questions:

Q1: How acceptable is it for you to share **<attribute>** with **<recipient>**?

Q2: How sensitive do you consider this **<attribute>** to be?

I introduced Q1 since sensitivity is often associated with information disclosure and governing information flows (e.g., [114, 168, 198, 208]). Hence, prompting participants to consider what information to disclose and with whom invites them to reflect on the characteristics of that information. For Q1, the recipients included: **partner(s), family, friends, colleagues, and researchers**. I selected this list to investigate how the acceptability to share varies within one's extended personal network and its implications regarding sensitivity and intimacy. Generally, CI scenario-based inquiries explore and vary five parameters: sender, attribute, recipient, and transmission principle. I vary only the attributes and recipients since in the wording of the question I emphasize people's agency over the transaction. Additionally, I wanted to reduce the complexity for donors by limiting the parameters and length of the question. Therefore the subject (i.e., user), the sender (i.e., Google Assistant prompted by its user), and the transmission principle (i.e., the user is informed and notified) remain constant. I introduced these at the beginning of the interview and brought them up in case further clarification was needed.

I invited the donors to answer Q1 and Q2 using a 5-point Likert-Scale[4]. I computed the average acceptability and sensitivity scores based on these [11, 12]. In addition, I invited donors to articulate the rationale behind their choice and elaborate on their responses. I did not want to impose the term 'intimacy'. Hence, I deliberately refrained from intro-

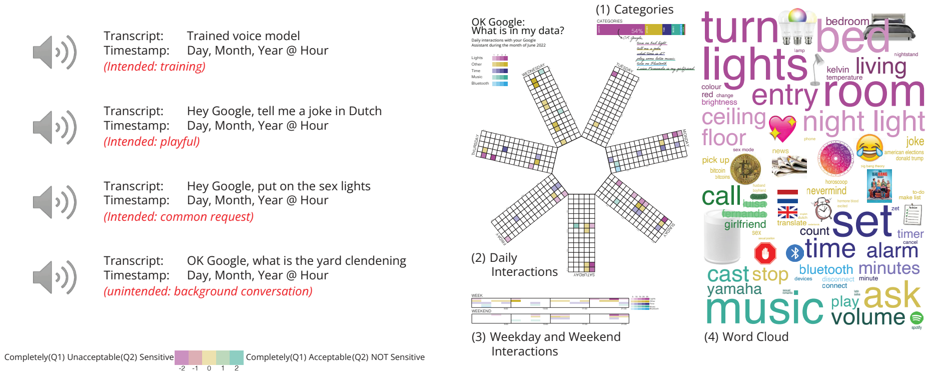
---

[4] Likert-Scale, Q1 from completely unacceptable (-2) to completely acceptable (+2), and Q2 from completely sensitive (-2) to completely not sensitive (+2)

ducing it and I only discussed it if and after it was introduced by donors. In addition, I deliberately left open the definitions of ‘sensitivity’ and ‘intimacy’ as I wanted donors to express themselves on their terms and bring their interpretations.

Personal Data Canvas

The *personal data canvas* is a representation of the donor’s data (Fig. 5.3). It is also described in Chapter 4. If you are already familiar with it, feel free to ignore this section. I designed it with the objective to comprehensively introduce the speech records and bring focus to the specific attributes. It consisted of two interactive views presented on a screen where donors could click and reproduce audio clips and zoom in and out of the different visualizations. If the interviews were in person, the visualization (Fig. 5.3.b) was also printed on A3 paper, inviting donors to explore it from multiple angles and annotate it. In the first view (Fig. 5.3.a), I focused on single interactions (e.g., “OK Google, turn on the lights”). Specifically, I presented the following interactions: (1) training (e.g., “OK Google”, generated when a user is first configuring their device and ‘training’ it to recognize her voice), (2) playful, (3) common request and (4) unintended. For these, I presented a clickable button with the audio recording (which is reproduced), and its transcript and timestamp in a human-readable format. In the second view (Fig. 5.3.b), I focused on multiple interactions, where the focus lies on the combination of timestamps and transcripts leading to potentially sensitive information, namely (5) patterns in time or sequence and (6) inferences. For these, I visualized the data and invited donors to explore the visualization, reconstruct the context of the data, and reflect on their patterns and potential inferences.



Transcript: Trained voice model
Timestamp: Day, Month, Year @ Hour
(Intended: training)

Transcript: Hey Google, tell me a joke in Dutch
Timestamp: Day, Month, Year @ Hour
(Intended: playful)

Transcript: Hey Google, put on the sex lights
Timestamp: Day, Month, Year @ Hour
(Intended: common request)

Transcript: OK Google, what is the yard clendening
Timestamp: Day, Month, Year @ Hour
(unintended: background conversation)

(a) Single interactions.

(b) Multiple interactions.

Figure 5.3: Example of a personal data canvas. Shown with permission of the donor

In the visualization (Fig. 5.3.b), I focused on conveying the information from the timestamps and transcripts of multiple interactions. Specifically, I identified common interactions for each dataset and grouped them into categories (e.g., weather, music, time). I visualized the distribution of these categories throughout the dataset with a bar graph

(Fig. 5.3.b(1)), and I represented each category with a different color throughout the visualization. Additionally, I presented the number of daily interactions for each category per hour of the day and day of the week with a heat map (Fig. 5.3.b(2)) where I focused on the 16 hours of the day with more interactions, the start and end times vary by donor. Similarly, I used a heat map to present the number of interactions of each category per hour of the day during the weekdays (Monday through Friday) and weekends (Saturday and Sunday) (Fig. 5.3.b(3)). Finally, I presented a word cloud (Fig. 5.3.b(4)) with the most frequent words grouped and color-coded by category, additional images were visually representing some of the terms. I added the images to make the interactions more prominent and easier to explore.

Reflexive Thematic Analysis

The interviews were audio recorded and transcribed. I made an initial transcript using MS Office 365, then manually reviewed and edited it. Jacky and I analyzed the transcripts using reflexive thematic analysis [32, 33], within a constructionist framework. We independently read through the transcripts to familiarize with the data and coded the entire dataset using ATLAS.ti. Through this process, we aimed to capture all the aspects of the data relevant to understanding how data subjects articulate sharing, sensitivity, and intimacy of their speech records. Jacky and I independently reviewed the codes and subsequently discussed and grouped them into tentative themes. I iteratively reviewed and refined the themes.

5.4. Results

5.4.1. Characteristics and Dimensions of Speech Records

Characteristics

Speech records contain brief interactions between a person and her Google Assistant, generally in the form of a request (e.g., “OK Google, set timer”) or a reply to a request (e.g., “thank you”). On average, the received speech records had 4.40 words per interaction – excluding the wake word; and lasted between 2 and 6 seconds. Donors interacted with their Google Assistant on average 3.91 times a day, with most interactions taking place in the early morning, late afternoon, and evening (Fig. 5.4(5)). The most frequent interactions were about turning on and off the lights, alarms, and reproducing music on a third-party app or device. These findings are consistent with previous research on the long-term use of voice assistants [21]. Jacky and I classified a small percentage (1.05%) of the interactions as unintended; most of these were recordings of background conversations. The majority of the interactions (98.95%) corresponded to intended interactions, as they were initiated from the wake word or were part of a series of intended interactions (Fig. 5.4(1)). From these, more than half (62.54%) corresponded to transactional requests, followed by navigational (16.86%), informational (15.05%), and playful requests (1.54%) (Fig. 5.4(2)). Furthermore, in all the datasets, more than one person (i.e., speaker) was present in the audio recordings.

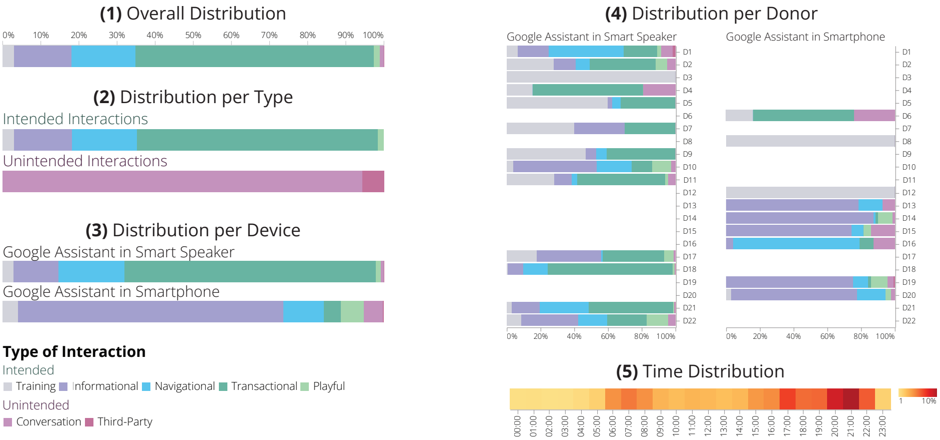


Figure 5.4: Distribution of **(1)** types of interaction across the received datasets **(2)** types of intended and unintended interactions **(3)** types of interaction when using the Google Assistant on a smart speaker and smartphone **(4)** types of interaction in the datasets received from each donor **(5)** interactions by the time of day over the entire data set.

Each type of request contains different layers of information about the person interacting with the Google Assistant. Transactional requests contain information about a person's set-up and the Google Assistant capabilities they use (e.g., "OK Google, set bed light to rainbow", "OK Google, change the blood pressure pill reminder time to 8 am"). Transactional requests potentially contain information about people's routines and habits from which information regarding a person's health, political interests, and sex life could be derived, considered sensitive under the GDPR. Navigational requests contain information about a person's digital routine and habits, including the digital content they consume and the third-party apps they interact with (e.g., "OK Google, play [song]", "OK Google, start instrument tuner"). These do not contain any information considered sensitive under the GDPR. Informational requests contain information about a person's interests and concerns; including questions about current affairs (e.g., "OK Google, who is the prime minister in [country]?"), relevant events and activities (e.g., "OK Google, when does the new electricity rate apply?", "What is the wheel to pray the rosary called?"), people's bodies (e.g., "OK Google, how do I know if I injured my rotator cuff?"), and even dreams (e.g., "OK Google, what does it mean to dream that someone dies?"). Several categories of sensitive information can be derived from informational requests (e.g., political opinions, religious beliefs, health- and sex-related). In addition, the low threshold of interacting with the Google Assistant means speech records explicitly include sensitive and intimate questions (e.g., "OK Google, why did I get so dizzy after [medical procedure]?", "OK Google, what is the sexual cowgirl position?").

The distribution of request types varies between people who use Google Assistant on a smart speaker (13 donors) and those who use it on a smartphone (9 donors) (Fig. 5.4(5)). In the first scenario, transactional requests are more frequent, as smart assistants are often integrated and connected with other smart appliances, while in the second, informational requests are more frequent (Fig. 5.4(4)).

Dimensions

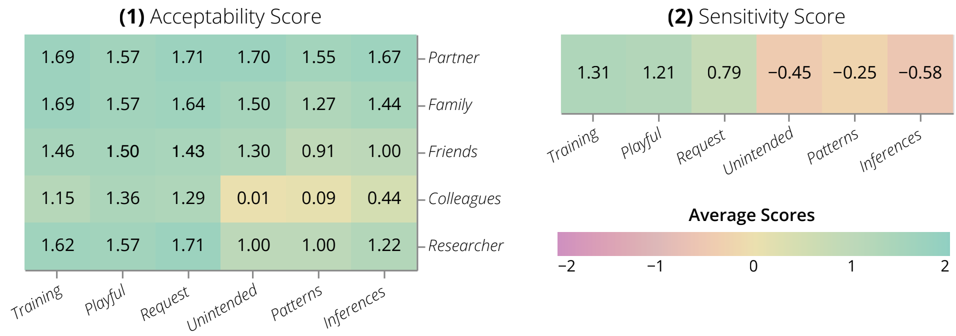
I introduce the three dimensions of speech records: *timestamp*, *transcript*, and *audio recording* by providing a short description and illustrating the information that can be derived from each.

- *Timestamp*, the date and time of an interaction (precision of milliseconds), describes when actions and interactions take place. For example, asking Google to "set an alarm for 8 am" late at night can indicate when a person goes to sleep and when she intends to wake up. When multiple interactions are combined, the timestamp can illustrate patterns in time (e.g., snoozes the 8 am alarm every weekday) and sequence (e.g., snoozes the 8 am alarm, plays a news podcast, asks about the weather). It can highlight specific aspects of a person's routine. Moreover, the timestamp enables data to be interpreted and abstracted according to different instants, such as the time of day (e.g., the middle of the night) and the month of the year (e.g., September). These can be associated with external factors (e.g., daylight, weather, public holidays) or situated within larger contexts (e.g., pandemic).
- *Transcript*, the content of the interaction, as interpreted by Google Assistant, describes what an interaction is about. Transcripts can reflect a person's worries (e.g., "OK Google, what should I do to protect myself from Corona?") and interests (e.g., "OK Google, how did [soccer team] score this season?"). Transcripts can also indicate a person's location (e.g., "OK Google, what's the weather like in [city]?", "OK Google, movies in [cinema], [city] today"). Moreover, transcripts illustrate people's relationships (e.g., "OK Google, [name] is my girlfriend", "OK Google, call my mom") and how people relate to their Google Assistant (e.g., "OK Google, set reminder for tomorrow morning, *darling*", "OK Google, you are so *stupid*"). The inferences derived from the transcripts are limited by and specific to the way people interact with their Google Assistant.
- *Audio Recording*, The sounds and speech of an interaction, as recorded by the Google Assistant in an audio file, describes who (i.e., speaker) and how (e.g., quiet room and sleepy voice, or loud background music and loud voice) interacts with the Google Assistant. Audio recordings can help differentiate intended and unintended interactions. The voice is a distinctive element in audio recordings, that identifies and distinguishes the speaker(s). The voice is rich and nuanced and communicates more than just words. Hence, interactions gain an extra layer when considering the nuances of the voice.

5.4.2. Sharing Speech Records

Average Acceptability Score

I calculate the acceptability score (Fig. 5.5(1)) by averaging the responses to Q1 (about sharing with different recipients) for each attribute and each recipient [11, 12]. Generally, donors considered it acceptable to share all attributes with most recipients, partner(s),



(1) Acceptability Score

	Training	Playful	Request	Unintended	Patterns	Inferences
Partner	1.69	1.57	1.71	1.70	1.55	1.67
Family	1.69	1.57	1.64	1.50	1.27	1.44
Friends	1.46	1.50	1.43	1.30	0.91	1.00
Colleagues	1.15	1.36	1.29	0.01	0.09	0.44
Researcher	1.62	1.57	1.71	1.00	1.00	1.22

(2) Sensitivity Score

Training	Playful	Request	Unintended	Patterns	Inferences
1.31	1.21	0.79	−0.45	−0.25	−0.58

Figure 5.5: Average **(1)** acceptability and **(2)** sensitivity scores

family, friends, and researchers ($acceptability\ score \geq 1$). The acceptability score is lower, tending toward neutral ($0 < acceptability\ score < 1$), for colleagues, especially for single unintended interactions and the patterns and inferences derived from multiple interactions.

Boundaries

A recurring theme in the analysis is the boundary between people's private and public lives (and spaces), which shape what is acceptable to share with whom. It aligns with the conceptualizations of privacy as boundary management proposed by [7] and [149] and it replicates some of their findings. I briefly introduce it as part of the results as it is highly relevant to the concepts of sensitivity and intimacy. D17 describes this boundary when referring to what can be inferred from her speech records:

> "Sometimes I ask pretty weird things to Google. So yeah, it shows a little bit more the weird parts of me, that I don't want to show everyone, other that, for example, my partner." (D17)

The acceptability to share personal data with a recipient decreases the further the recipient is from the context and the space in which data are generated. Sharing with recipients who are inside the private space is considered acceptable since they already inhabit that space and are familiar with what happens within, *"it is information that you usually talk* [about] *with your close people, that they already know. Your routine, your activities, if you are going somewhere, it is something that you share all the time"* (D16). In some cases, being inside this space means they are even part of the data, *"I mean he's there* [in the audio recording]*"* (D19).

Sharing with recipients who are in *the boundary* of the private space is considered less acceptable, *"friends, and colleagues, they don't have to know you are going* [somewhere]*"* (D10), *"they don't need to know what I do, what is my routine"* (D16). Because these recipients do not belong to the private space, they are unaware of its peculiarities *"they don't know the context"* (D18), and may misinterpret, misunderstand or judge what happens within, *"if I share that* [playful interaction] *with my colleagues, they can judge me, and*

I interact with them so the judgment is more real" (D11). The concern of being judged comes from the lack of control over how data, incomplete and decontextualized, is received by others.

Sharing with recipients who are outside the private space is generally considered not acceptable. Here, data might lead to an opportunity to access and violate the private space, *"that's something that people can abuse"* (D21). D17, who has smart lights linked to her Google Assistant, describes the possibility of harm when discussing these interactions:

> "[the light interactions] tell about my routines, when I'm home and when I am not home. So this might be used against me [...] If someone wants to threaten me or if they want to steal something in my house or something like that." (D17)

Researchers, who are outside the private space, are a notable exception. Donors highlighted the importance of contributing to science and advancing research, *"I believe in science, so they* [researchers] *may use my information"* (D17). Evidently, donors already shared their personal data with this research. For donors, researchers are responsible and carry values such as trust and discretion. As such, they are expected to make proper use of the data, *"handle this* [data] *with precaution"* (D1).

Characteristics of the Data

I developed three themes related to the characteristics of the data that are important to consider in the context of sharing, sensitivity, and intimacy. These are not specific to individual speech records or the information derived from them; hence, we use the encompassing term data.

- Data are *contextual.* They are generated and stored in a given context, which is lost when translated into a discrete event. For example, the interaction "OK Google, how does a guinea pig sound" occurs in a specific context, *"I use it to demonstrate how my guinea pigs react. And its purpose can be various. It can be for friends, it can be for yourself, it can be for your cats. It's extremely funny* [laughs]" (D10). Yet, it loses playfulness and social dynamics when recorded. Hence, data can *"always be misinterpreted"* (D18) if they are disassociated from their context. In addition, because of how the data are generated (e.g., through seamless voice interactions), it portrays a limited and specific image of a person that loses the nuance embedded in the context.
- Data are *relational.* They relate to the interaction between a person (or people) and a device, sharing a physical space (e.g., a room), as well as other people, not necessarily sharing a physical space (e.g., partners, family members, roommates, neighbors, visitors). In addition, they account for relationships with others, *"I'm putting information out there about our relationship"* (D21). In fact, more than one speaker was present in all received datasets, especially in multi-user environments, where more than one person shares space and device (e.g., partners, family members, roommates), but also in single-user environments where other people are occasionally around (e.g., neighbors, visitors).

- Data are *multiple*. A term introduced by Prainsack [156] referring to it being able to be and used in more places than one at the same time. They can be duplicated and shared, which became apparent during the data donation process when donors obtained and shared a 'copy' of their data. Multiplicity illustrates how data can be reused, and potentially misused, for more than one purpose and more than one entity. In addition, data when aggregated and combined becomes a combination of *"multiple fragments* [from which] *you can distill some meaning"* (D21).

5.4.3. Sensitivity of Speech Records

Donors interpreted the sensitivity of speech records in three ways. First, how personal or private individual speech records or the information derived from them are. Here, sensitive speech records contain personal information (from within the private space). Second, how disclosable individual speech records or the information derived from them are. Here, sensitive speech records are not to disclose; their disclosure would be considered a violation or potentially harmful. Third, to what extent can individual speech records or the information derived from them be expanded or used differently. Here, sensitive speech records can lead to inferences about personal information, from within the private space, and can be re-used outside the context in which they were generated. Starting from these interpretations, I describe the characteristics of sensitive speech records and map these into a sensitivity spectrum.

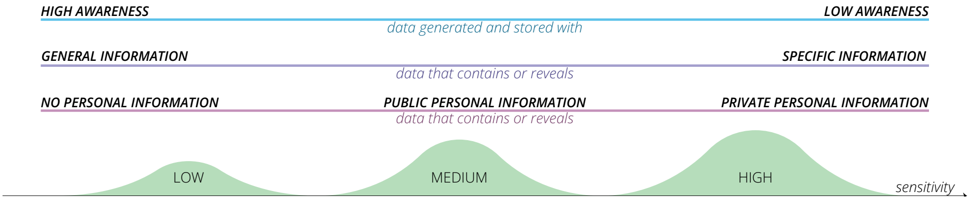


Figure 5.6: The spectrum of sensitive speech records.

Characteristics

I developed three themes describing the characteristics of sensitive data (Fig. 5.6): *intrusiveness*, *specificity*, and *(un)availability*. Sensitive data are *intrusive* as they are generated within people's private spaces where others are not necessarily welcome or invited. Intrusiveness unfolds through a shared physical space where, by capturing what happens within, others gain indirect access through data.

> "That I'm planning [activity] it's something that I believe should be in my private area. It's something that I don't feel comfortable seeing on that screen. Because it means someone has the opportunity to know what I'm planning, where I'm going, what I'm thinking, things that are worrying me, very, very, very personal things." (D14)

Intrusiveness is related to people's awareness of the device's data collection and storage practices and how they experience them. Hence, it manifests when the device captures more than expected.

> "It's like if you're naked you don't say «OK, Google» you first dress up, and then you say «OK Google», which is, I don't know if it is a lousy metaphor. But what I'm trying to say is that we feel aware that for a few seconds when we say «OK Google», it's not like, we're not exactly alone." ($D9_b$)

As expressed by $D9_b$ donors described a sense of awareness and control over their interactions with the Google Assistant that is lost when data is generated from unintended interactions or interactions that record *"a little bit more than my question"* (D1). Thus, the device intrudes further into a space by capturing data that reveals more information than people intended to disclose or were unaware of. Moreover, intrusiveness relates to the amount of data that is made available, as more data means more angles from which to access a person's private space and more possibilities of unawareness, *"anyone can have one digit of my pin. I'm happy to give everyone one digit of my pin. But not all four"* (D21).

Sensitive data are *specific* as they more narrowly record and reflect certain aspects or themes of a person's private space. D17 illustrates how sensitivity varies with more and less specific points when comparing two interactions, the first about her mom and the second about the office lights.

> "The [first] one attaches something to me as a person that is my mother. So, some information regarding something specific to an individual. This [the second] is really general; all kinds of people have office lights. When I ask about my mom, I know that the answer is going to have specific information in it, so that's why I consider it more sensitive." (D17)

Specificity relates to what is uniquely about a person, as opposed to generic information that can be attached to anyone or that relates to decontextualized activities.

> "It's about how unique is that data. Generalistic things that everyone does, that don't have any specific anchoring in space or time, for instance, the fact that I live in [city], it's on my profile. The fact that I live on that particular street, and like on Tuesday he does this, that is more sensitive, so it's about the specificity of the information." (D21)

In addition, specificity relates to data being spatiotemporal. This means data can reveal when and where certain interactions occur as they have a *"specific anchoring in space or time"* as described by D21. It gives an extra layer of specificity to the data, as they can be further situated by considering external factors such as the day of the week and the weather. Furthermore, specificity may derive from the aggregation of multiple data and sources of data, resulting in a distinct representation of certain aspects of a person's private space, *"they know, hey, that is* [name] *and combined with all the things you do on the internet. Well, it gets pretty sensitive"* (D10).

Sensitive data is *(un)available*, as they are generated within a person's private space where *"stuff is much more personal. So I want to guard it a little bit more"* (D11). For instance, the health-related questions that people ask Google before even discussing them with their family, friends, and doctors.

They are collected and stored, hence they are potentially available, yet they are generally not at someone's disposal. They reveal information that is not widespread or publicly available as opposed to information that people *"probably can find everywhere"* (D18). The increasing availability of personal data online and on social media means that widespread information from within the public space is not necessarily considered sensitive.

> "I have social media, so yeah, it's normal that these things are known by a lot of people that I don't know about" (D17).

The combination of multiple interactions over time means sensitive data, in the form of patterns and inferences, is sometimes (un)available even to the data subjects.

> "You know? It's weird. Like I think I'm not conscious of how much they know about me. Because I just ask random questions [to Google], but then, they put it together, and it's like more information that I'm actually conscious. I'm not conscious of that quantity" (D19)

Average Score

I calculated the sensitivity score (Fig. 5.5$_{(2)}$) by averaging the responses to Q2 (about perceived sensitivity) for each attribute [11, 12]. Single intended interactions (i.e., training, playful, and (common) requests) are generally considered not sensitive ($sensitivity\,score \geq 1$). While single unintended interactions and the patterns and inferences derived from multiple interactions are generally considered slightly sensitive ($sensitivity\,score < 0$).

Spectrum

I identified a spectrum of sensitivity of speech records (Fig. 5.6) based on the three characteristics, *intrusiveness*, *specificity*, and *(un)availability*. Low sensitivity is attributed to individual intended interactions; where data subjects are aware data was generated and stored. These relate to simple and mundane activities and reveal little or no information about a person. This is illustrated by D6, for whom asking Google to turn on Bluetooth *"it's just an order. I didn't say anything, like personal information, nothing like that. It's just some words. It doesn't mean anything."* Additionally, they contain information that is already known or widely available (e.g., through social media, on the internet), *"that everyone can look for"* (D19). Such as someone's relationship status or city of residence. For example, D20 uses his Google Assistant to navigate the city and has multiple interactions containing information about specific stores and addresses, including work and home, *"I mean, everybody knows that. What I do, where I go shopping. It's nothing special."*

Medium sensitivity is attributed to single intended interactions that contain traces of personal information and *"tell a bit more about me"* (D18), even when these correspond to simple and mundane activities. For example, the single interaction *"what's the weather like in* [city]*?"* (medium sensitivity) *"shows a little bit about my behavior or plans"* (D2) with respect to the interaction *"what's the weather like?"* (low sensitivity). These kinds of interactions contain public personal information, but not information that is specific about a person or her private space.

High sensitivity is attributed to single intended interactions when they reveal specific information about a person's private space, including health and well-being, alcohol consumption, political opinions, interests, musical preferences, and a person's *"weird parts"* (D17), including the *"kind of stupid things I said to Google"* (D18).

> "I guess more a window into your more personal life, yeah, which I guess is by definition more sensitive." (D6)

High sensitivity is also attributed to single interactions, occurring when the speaker is unaware. This includes unintended interactions, *"Google was recording information that I was not aware of, which I would say is sensitive"* (D16), and intended interactions where the speaker is not completely aware, *"if I'm sleepy or if I'm drunk, like I'm in a state* [where] *I'm not fully aware"* (D5). These interactions often generate a sense of discomfort and awkwardness. Unawareness extends to other people (e.g., partners, family members, roommates, neighbors, visitors) tangled in the data due to its relationality. Other people's interactions, whether intended or not, are considered highly sensitive and a violation of their privacy, *"it is my uncle's privacy, I don't want to compromise, someone else's privacy"* (D2). Moreover, single interactions gain an extra layer when considering the nuances stored in the audio recordings and communicated through the voice, *"that's also something I don't think about, right? Like how do I sound at 11:33 PM when I'm asking Google to do something for me?"* (D11). Finally, high sensitivity is attributed to the patterns and inferences that can be derived from multiple interactions.

> "Even the data that you normally would say it's not sensitive, but if you put it all together it paints quite a picture of who you are and what you're doing, when, how, why." (D18)

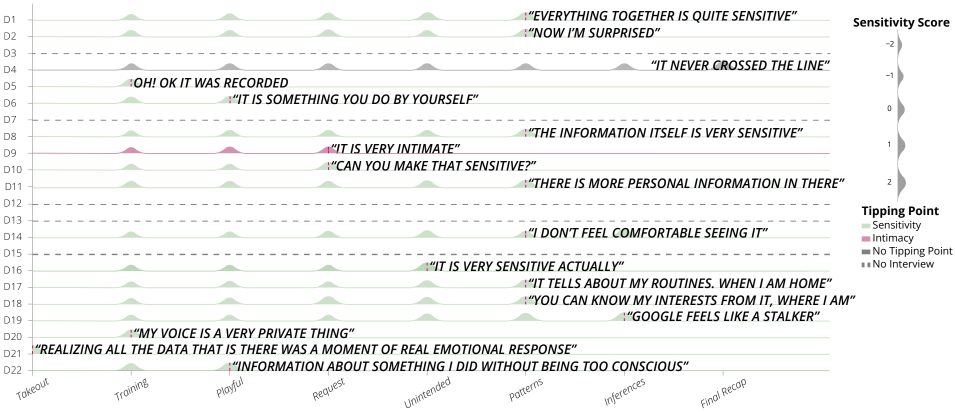


Figure 5.7: Tipping point in the course of data donation process and data-centric interviews.

5.4.4. Intimacy of Speech Records

Although I did not refer to it, the term 'intimacy' came up frequently during the interviews. Donors interpreted the intimacy of speech records in three ways. First, how speech records surface from closeness and physical proximity in a shared space. Here, intimate speech records reflect information from within that shared physical space. Second, how speech records capture intimate moments and activities. Here, speech records are derived from a device that is the viewer, or nearly participates, in intimate moments and activities. Third, how speech records capture the nuances of the daily life. Here, intimate speech records grasp what is often not material or explicit (e.g., playfulness, sleepiness, vulnerability). These three interpretations are reflected in a short excerpt where $D9_a$ explains what she means when she describes the speech records as *"very intimate"*:

> "We were listening to recordings of when we asked the alarm to turn off, and you can tell there is like a couple in bed, very sleepy, you know? And it feels like it [Google Assistant] was in bed with us, kind of, and so it's not the public voice I have when I'm online or with my friends. It's like a space that is really, really intimate, in the sense that no one else is usually allowed." ($D9_a$)

Given that the physical space is permanently shared and often corresponds to an intimate space within a person's home or close to her body, intimacy permeates through all three dimensions of speech records. Timestamp, by capturing instances of a time frame that belongs to the intimate private space, *"at 11:00 PM it's more like a private time, a more intimate time. But around the morning it's more like working* [time]*, so it doesn't feel like that intimate"* (D5); Transcript, or content, delving deeper into a person's private space by depicting her inner thoughts, ambitions, and vulnerabilities, which are often not visible to others, *"information about my thinking, my worries, and what I look for is critical. It seems very personal"* (D14); And audio recording, documenting the acoustic nuances of the space and the speaker, *"you can hear that the person doing the recording is actually sleepy, and that feels more intimate somehow"* (D2). Intimacy also derives from the interpretation and reconstruction of the data, where mundane interactions gain additional layers as they are situated in a specific context where intimate activities occur. This type of intimacy is not inherent in the data but unfolds through the active participation of the data subjects.

> "The other thing is, it's not represented here [the visualization], occasionally when my partner and I have sex, we turn on and off the lights, kind of like before to arrange things and then after to clean things up. So there's maybe a detectable pattern that reveals our sex life." (D11)

Intimate data shares the characteristics of sensitive data. That is, intimate data are *intrusive* (i.e., generated through close physical proximity to a bystander device), *specific* (i.e., uniquely about a person and inherent in who they are), and *(un)available* (i.e., capturing the nuances). Yet, unlike sensitivity, intimacy is not associated with information disclosure and potential privacy violations. It relates to the information itself and its materialization. Hence, capturing and reflecting what is often not expressed or made explicit; even if mundane.

5.4.5. Tipping Point

The data donation journey (Chapter 4) involved several opportunities for donors to reflect on their speech records, including downloading a copy, exploring the takeout file, and visualizing them. Through these activities, most donors experienced a tipping point where a seemingly innocuous dataset suddenly contained sensitive and intimate data about themselves. Of the 17 donors who participated in the data-centric interviews, 16 experienced a tipping point (Figure 5.7). For D5, D20, and D21 this point was early and came from the paradoxical realization that speech records were collected and stored. Paradoxical since donors were aware of this, but it became even more evident by listening to individual speech records. D21 manually explored the takeout file before donating his data; for him, this realization came before the data-centric interviews. For D5 and D20 this realization came early during the data-centric interviews when listening to the first speech record (i.e., single training interaction). For D6, D9, D10, and D22, this point came from realizing that traces of private personal information or nuanced intimate information were present in individual speech records. For D16 this point came from an unintended interaction; leading to a misalignment of expectations about the device's behavior and a breach of trust. Finally, for the remaining eight donors (D1, D2, D8, D11, D14, D17, D18, D19), this point came from gaining a better understanding of the information derived from the aggregation of multiple speech records; facilitated through the personal data canvas. Namely, patterns and inferences. The tipping point underlines how sensitivity and intimacy are not self-evident but characteristics of the data that need to be disentangled and explored.

5.5. Discussion

In this chapter, I investigated people's perspectives on sensitivity and intimacy with a focus on speech records. The findings illustrate three characteristics of 'sensitive' and 'intimate' speech records. First, they are intrusive as they enter a person's private space, capture what happens within and expose it to unwanted access, or 'the unknown' [177]. Second, they are specific, narrowly personal (i.e., about me and no one else) and depicting a situation through time and space. Third, they are (un)available, not widespread yet potentially available. I discuss my findings, discuss implications for data holders, policy makers, and researchers involved with sensitive and intimate data (i.e., speech records or the information derived from them), and provide recommendations for researchers aiming to support research participants disentangle the sensitivity and intimacy of their speech records.

5.5.1. Understanding Sensitivity and Intimacy

Previous research on single speech records (i.e., individual interactions with a voice assistant) concludes that these are not perceived as 'sensitive' information [111, 123]. My research partially aligns with their findings (i.e., individual interactions are generally considered not sensitive). However, it highlights the fundamental differences of comprehensively engaging with speech records. The interaction "OK Google, set an alarm", mundane and not considered sensitive in the GDPR [65], by HCI scholars [112, 114, 198, 208] (not accounting for inferences [168]), or according to previous research [111, 123], illustrates the nuances of people's perceptions of sensitive and intimate data (i.e., speech records or the information derived from them). This interaction is perceived as 'sensitive data' when it is generated from a place of playfulness (e.g., funny accent) or low consciousness (e.g., unaware, drunk). Multiple of these interactions are perceived as 'sensitive data' when information about people's behavior and routine can be inferred from them. Additionally, this interaction is perceived as 'intimate data' when it is generated from a place of vulnerability (e.g., sleepy) or it accounts for intimate contexts and activities (e.g., a couple in bed) occurring within a shared physical space.

Notions of 'sensitive data' as information that should not be disclosed [114, 198] or belonging to specific categories [65, 112, 168, 208] are limited. They do not account for the pervasive and dynamic nature of the data or potential inferences derived from aggregation. Sharing a physical space with a digital device, such as Google Assistant, means it captures various sensitive and intimate information over time, belonging to prescribed categories (e.g., political opinions) and beyond (e.g., subjectively defined private spaces). Yet, these are usually fuzzy, not evident on the surface. Sensitivity has been widely associated with private information [65, 112, 168, 208] and information disclosure [114, 198]. Therefore, it has been closely related to privacy; defining where sensitive information resides and where are the boundaries [7, 149] and selectively [145] and contextually [137] limiting disclosure. But, how can I control the disclosure of information I don't know is there? Sensitive speech records are not only a window for 'the unknown' [177] to access a person's private space but are also, to some extent, 'the unknown'.

From the three characteristics I identified, I propose to understand sensitive data as personal information that is absorbed and abstracted from a person's private space. Sensitive data are absorbed as they are generated through digital devices that intrude on a person's private physical space and transform specific instances into information that is situated and time-bounded but decontextualized. Thus, sensitive data retains information on a person's behavior that is often (un)available to her and others. They are abstracted as they detach a person from the context and nuances of her private space. Moreover, I propose to frame 'intimate data' as a subset of 'sensitive data' that relates to the absorption and abstraction of information about intimate contexts and activities within a shared physical space. This aligns with previous literature relating intimacy to physical togetherness [87, 146] and intimate contexts and activities [5, 6, 109]. This conceptualization, and the characteristics on which it is based, rely on the distinction and boundaries between private and public information [7, 149]. Yet, these boundaries become fuzzy around data collected as a byproduct of people's interactions with a digital product or service, such as speech records. The conceptualization of 'sensitive data' contributes to the privacy literature by underlining the need to disentangle sensitivity and intimacy from the data as a pre-requisite to defining the context-specific boundaries around it.

5.5.2. Implications for Data Holders, Policy Makers, and Researchers

> "I must say that I feel a bit naked. In the sense that this tells a lot about me, much more than I expected. And, in fact, it shows how technology actually is so close to us that it is able to give this amount of information." (D17 at the end of the interview)

Surprise and unexpectedness were common feelings among donors who experienced a tipping point. Yet, they played an active role in the generation of most speech records; when giving a command or asking a question. In addition, they were informed of the data collection practices of their Google Assistant at least in two instances; when setting up their device and when they came across this research. Feelings of surprise, unexpectedness, and even nakedness illustrate how data holders (e.g., Google) and researchers (e.g., me) often fail to adequately inform their users and participants. They know that personal data is being collected, but they often do not know how it looks, how it feels, or what kind of sensitive or intimate information it contains or can be inferred from it. I discuss implications for data holders, policy makers, and researchers, aimed at mitigating surprise and unexpectedness.

- **Implications for Data Holders:** The surprise of finding unintended interactions or unraveling sensitive information from the aggregation of multiple interactions diminishes trust in the data holders. For them, mitigating surprise is an opportunity to build better relationships with their users and build trust [31]. Data holders have the opportunity to build new functionalities around personal data and provide their users with new capabilities such as data exploration and curation. Currently, Google Assistant users can review a history of their data (i.e., a scrollable list) and manually delete specific interactions. They could provide alternative data representations that empower users to engage with data as a tool to reflect and gain personal insights [31], and even actively re-define their privacy boundaries [111, 145, 190, 217].

- **Implications for Policy Makers:** The donors in our study made use of the GDPR's 'rights to access and data portability' [65] to obtain a copy of their data from Google and reuse it for the purpose of contributing to this research. Although they obtained a copy of the data and could explore it autonomously, most donors found the multiple files and formats difficult to navigate. Previous research points out to similar limitations of data portability, as data is generally provided in frustrating formats and through difficult processes [4, 31]. It aligns with the purpose of 'portability' which attempts to increase end-users agency by decoupling the data from its holder. Policy makers could mitigate donors' surprise by developing policies encouraging personal data literacy [84, 212]. For example, they could require data holders to engage with end-users to create a set of data interactions that increase their understanding of the data.

- **Implications for Researchers:** My research is one of many that engage with personal data. During this process, I witnessed the discomfort and vulnerability that stem from surprise and the unexpected. I argue that the fuzzy and sensitive nature of personal data is an opportunity to develop a research process and agenda that integrates personal data literacy and exploration. First, it is an opportunity to facilitate the disentangling of the data through visualizations and creative sessions [71, 85, 154], and support participants in actively and iteratively defining personal boundaries. Second, it augments data with contextual information, reducing assumptions and misinterpretations for the researchers [51]. In addition, this research underlines the relational nature of personal data. They depict people's relationships with others [155, 156], which naturally occur when personal data is generated through multi-user digital devices in public or shared spaces, such as voice assistants [41, 111, 123, 154]. Thus, accounting for others, such as third parties and minors [41], presents an open challenge, involving and informing all parties or excluding them from data collection.

5.5.3. Recommendations: Disentangling Sensitivity and Intimacy from Speech Records

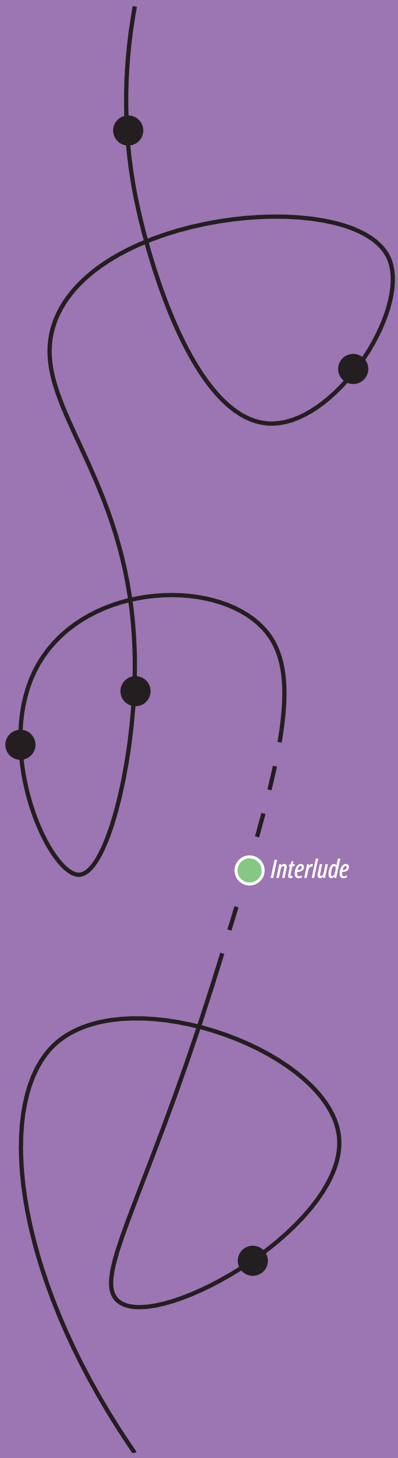
The *personal data canvas* was a first step in supporting users and research participants to disentangle sensitivity and intimacy from their speech records. It enabled donors to interact with their speech records in a less confusing way than the difficult-to-explore takeout file. Besides, it raised awareness about unintended interactions, patterns, and inferences. I translate my experience designing it into practical recommendations to design processes and tools that help disentangle the sensitivity and privacy of speech records. These are directed at researchers engaging with speech records and similar digital-trace data.

- **Provide a comprehensive overview:** I provided donors with multiple perspectives from which to approach their data. In contrast to previous research (e.g., [111, 123]), which focused on randomly selected single interactions, I introduced different types of single interactions and information derived through aggregation and inference. I recommend providing a comprehensive overview to progressively and incrementally approach and interrogate sensitivity and intimacy.
- **Draw attention to the unintended and the unexpected:** Donors often found sensitivity in situations of low awareness or vulnerability (e.g., unintended interactions). These are less frequent, especially in large datasets, but they are important to highlight. I recommend bringing them to the fore but acknowledging that they are rare.
- **Support structuring the data:** The *personal data canvas* depicted categories based on patterns of use, temporal patterns, and frequency, which donors highly appreciated. For some, the canvas was the highlight of the interview. In addition, they were helpful prompts for personal reflection and interrogation of the sensitive and intimate elements. I recommend facilitating the process of structuring the data to support interpretation and sensemaking.
- **Foster reconstructing the context of the data:** Speech records are limited and decontextualized. However, in some cases, sensitivity and intimacy derive only from the context in which they are generated. For instance, a donor recognized anomalies in her Friday routine that led to her identifying sensitive attributes in her data. I recommend creating activities that facilitate the reconstruction of the context of the data so that people understand how data relates to their day-to-day life and activities. It is an opportunity to encourage self-reflection through the data and disentangle sensitivity and intimacy.
- **Help navigate personal settings:** At the end of the interview, more than half of the donors were interested in receiving suggestions on how to configure their devices

to mitigate the collection of sensitive or intimate speech records. I recommend providing general guidelines (e.g., how to disable voice data collection) and space for case-specific suggestions.

5.6. Chapter Takeaways

I started this chapter with a question: What is sensitive about (sensitive) data? This question is framed to suggest there are (sensitive) data types – such as those defined in the GDPR – and sensitive elements, characteristics, or information in the data. The results of this chapter illustrate how sensitivity and intimacy are not inherent to data (types), but nuanced properties that are often unknown a priori. This has critical implications for sensitive data donation. Recognizing that sensitivity is not a static property of data (types) and that data might be sensitive in one context and not in another requires approaching (all) data as sensitive. This, in turn, suggests that interpreting, situating, and contextualizing data not only promotes Multiple Knowledge(s) **(P3)**, it also enables and empowers donors to uncover the sensitive and intimate elements of their data – mitigating surprise and the unexpected and allowing them to define meaningful boundaries around what (not) to disclose.


Interlude

6

Interlude: Dataslip

Abstract

In this chapter, I introduce dataslip. It is a provocative artifact that materializes the personal data trail in the form of a receipt. Dataslip aims to raise awareness over personal data collection and elicit creepiness, *"an emotional response to a sense of wrongness that is difficult to clearly articulate"* [177]. Dataslip responds to a critical challenge in data donation research, described in Chapters 3 and 5, donors are hardly aware of what their data contains. Thus, if and when they decide to donate them, they do so more-or-less blindly. I introduce the design considerations that led to dataslip and I describe how I used it to foster a conversation with local community members about the underlying challenges and potential alternatives to personal data collection and use. This conversation, together with a collaborative futuring exercise with children, led to a set of five speculative design scenarios that suggest hopeful and empowering alternatives to the personal data entanglements we are currently embedded in.

This chapter draws on the following publication: **Gómez Ortega, A.**, Noortman, R., Bourgeois, J., and Kortuem, G. (2024) Dataslip: Into the Present and Future(s) of Personal data. In Proceedings of the Eighteen International Conference on Tangible, Embedded, and Embodied Interaction (TEI '24).

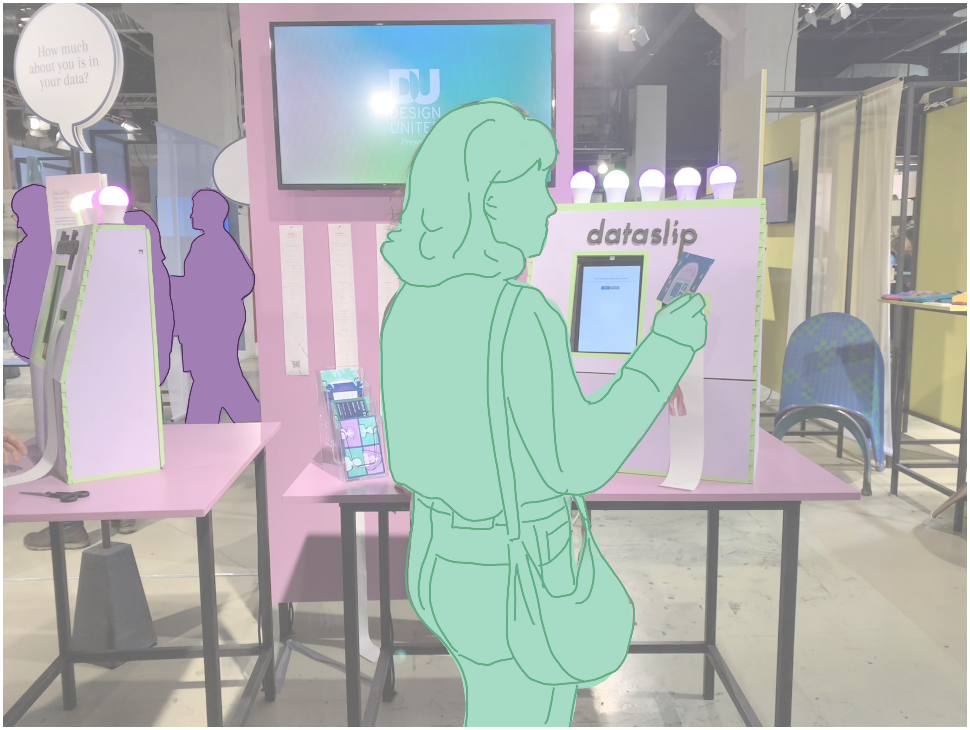


Figure 6.1: Dataslip installation.

6.1. Introduction

As I mentioned before, most of us routinely interact with products and services that collect and indefinitely store personal data; at the same time, we are (mostly) unaware of the nature and vastness of these data. For example, when a person registers for a loyalty card at the supermarket, she is aware that she is volunteering personal information, such as her name, address, and email [200, 209]. When she uses the card to get a discount on her purchases, it's less clear that personal data from this interaction is being collected and even more, exactly what (types of) data. Thus, there is a misalignment between people's understanding and expectations of their data and their actual collection and use by product and service providers. It hampers people's rational understanding of their data and even more, of what data feels like [56, 76, 176, 177]. What data feels like has been the focus of Data Epics, which aim to challenge how data are imagined and represented through fiction and speculation [19, 56].

Many factors contribute to people's lack of awareness, including the pervasive nature of data collection [58, 209], the abstract nature of data and the terminology commonly used to refer to them (e.g., the cloud) [71, 211], and the often unclear [119, 120] or misleading [27, 83] terms of service and privacy policies. With it, also comes a sense of disempowerment with respect to the power imbalance between individuals and private companies or public services collecting, storing, and benefiting from data about them [31, 155]. Shklovski and colleagues [176, 177] argue for eliciting visceral reactions such as creepiness and discomfort to underline these issues. Yet, these feelings that stem from increased awareness do not lead people to stop using the products and services that trigger them. Instead, people often choose to ignore or forget about them and move on, by continuing to engage and interact with products and services with creepy data collection practices. Hence, people tend to close themselves off in the face of the creepy powerlessness regarding personal data collection. Yet, being closed off means they often shy away from being involved in shaping the future, which further disempowers them.

In this chapter, I materialize the creepiness of personal data collection as a starting point to envision alternative futures. To do so, I designed dataslip, an interactive and provocative artifact that promotes awareness and invites people to feel and interrogate the creepiness of their personal data trails (i.e., the collection of data left behind from each interaction with a digital product or service) through a tangible representation in the form of a receipt. Together with a co-author of this study, Renee Noortman, we use dataslip in two activities. First, during a local community event, where the creepiness of the receipt prompted attendees to reflect on challenges and concerns around personal data collection and their current approaches to mitigate them. Second, during a workshop with primary school children, where we used dataslip as a co-speculator to co-create alternative and hopeful futures that stem from the challenges identified during the first activity. We propose five speculative design proposals co-created with community members that respond to their experiences with dataslip through hope and empowerment. I conclude this chapter by discussing our approach and reflecting on the benefits and limitations of moving from creepiness toward hopefulness and empowerment.

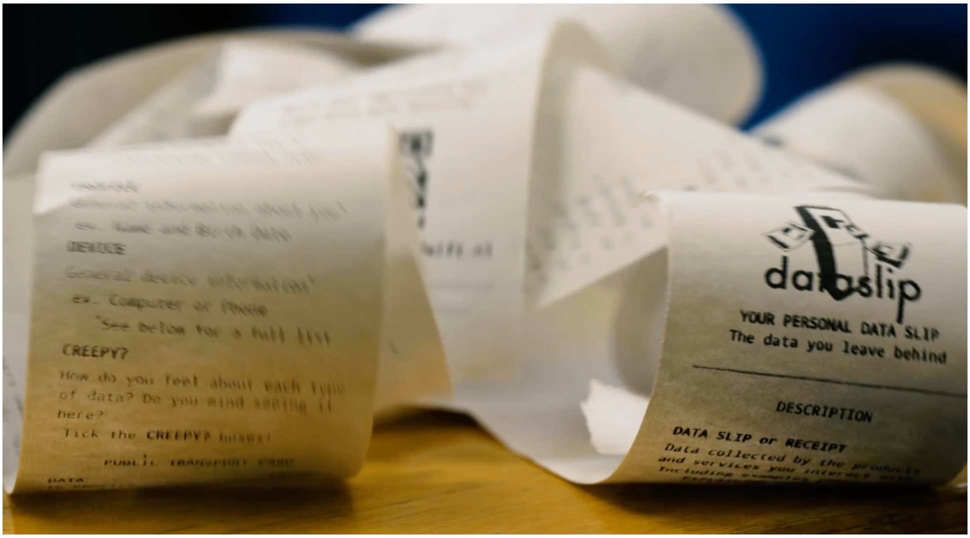


Figure 6.2: Receipt or personal dataslip materializing the personal data trail

6.2. Creepy Personal Data

Personal data are defined in the European General Data Protection Regulation (GDPR) as any information through which a person can be directly or indirectly identified ([65, Art. 4]). Examples include a person's phone number or email address, directly associated with her, as well as the WiFi access points she connects to through her mobile device, indirectly associated with her. Whether directly or indirectly, individuals have the right to be informed ([Art.13 - 14] [83]) about the collection and use of their personal data. Yet, they are often informed through lengthy documents, such as privacy policies, or short statements, such as cookie pop-ups, which are hardly effective [54, 202].

For this reason, when it comes to personal data collection and use, most of us "don't know what we don't know" [100]; meaning personal data collection and use are opaque. Moreover, personal data collection is so entangled with our day-to-day, and our interactions with digital products and services that data themselves are opaque and unknown [76, 177]. They contain several fragments and types of personal information, and it is difficult to account for their length and depth. Even if individuals go one step further, for instance, by requesting a copy of their data (e.g., browsing history logs) from a data controller (e.g., Google), data are returned in files and formats that are hard to manipulate and understand [4, 31]. Thus, even here, when data are in a person's hands (or device), they remain opaque and unknown.

In most cases, people are surprised when they become aware of the data collection practices of the products and services they interact with. This includes: (1) realizing that the data was being collected in the first place [74, 177], (2) understanding what types of data are being collected [74, 76], and (3) discovering all the information that can be inferred from the data [76, 154]. For this reason, becoming aware, and realizing data are, and reveal, more than they seem generates discomfort or creepiness [76, 177]. Yet, creepiness is temporary. Even if they are creeped out by the data collection practices of a specific product or service, most people continue to use it as they normally would. Shklovski and colleagues [177] argue that emotional visceral reactions, such as creepiness, point to important underlying issues and suggest provoking and confronting the creepy nature of digital technologies head-on.

6.3. Futuring and Design Fiction

Since Dunne and Raby's Speculative Everything [60] was published, 10 years ago, speculative design and futuring have become a common item in the toolkit of researchers and designers. They serve to elicit open discussion and debate about a wide range of topics and subjects [23, 89]. More recently, these approaches have been reflected on more extensively, leading to a distinction of the different modes of reflection in futuring [104]:

1. **Designerly formgiving, its specificity, and experiential qualities:** how a specific artifact to imagine one specific future can 'close down' on thinking about certain futures, while 'opening up' to think about and focus on one future in detail.
2. **Attending to temporal representations:** how engaging with temporality, as culturally situated can underline assumptions around the conceptualization of futures

(e.g., linearity of time).

3. **Positionality, futuring from somewhere:** how researchers and designers can more thoughtfully and intentionally consider their own positionality and privilege.
4. **Engaging with the real world and the public:** How design futuring projects can engage with real-world gaps, issues, and opportunities and foster more open debates.
5. **How design futuring generates new knowledge:** how design futuring projects can connect or build upon other(s') design futuring works.

Coulton and Lindley observed two common approaches to futuring: (1) that of Vapourworlds and (2) that of design fiction. Vapourworlds propose industry-driven technocratic utopias, while design fictions are produced as critical, technology-averse dystopias [46]. Previous research underlines a need to close the gap between utopian and dystopian futures thinking [23, 104, 196]. For this reason, the pioneers of speculative design practice are actively working towards closing the loop by taking speculative work from the art gallery into everyday life [89]. One approach to this could be to design for Ustopia - or the interaction between utopia and dystopia - instead [138]. In response, this chapter strives to critique current data practices while also providing a positive outlook towards the future, to emphasize that the future is not inevitable and that consumers can play a role in how they use their data and how it is used by others.

6.4. Method

6.4.1. Designing Dataslip

Dataslip is an interactive installation that emulates an Automated Teller Machine (ATM). It alludes to the "data is the new gold" narrative, where data is equivalated to a valuable currency, but valuable to whom? Similar to an ATM, it consists of a touchscreen and a printer. I designed dataslip with the following goals: (1) to promote awareness of personal data collection and invite people to reflect on the value of data to them, (2) to bring materiality to the abstract notion of personal data, and (3) to elicit anticipation, confrontation, and creepiness – an emotional response to a sense of wrongness that is difficult to clearly articulate [177], by enabling people to feel their data.

The interaction with the dataslip emulates that of an ATM. People interact with a touchscreen where they answer five simple yes/no questions, based on their routine interactions with digital products and services. Here, I focused on a range of products and services that people encounter daily: (1) personalized public transport cards, (2) supermarket loyalty cards, (3) credit and debit cards, (4) wearables, including smartwatches and smart rings, and (5) mobile apps, including weather, navigation, web browser, email, instant messaging, music, social media, dating, and period tracking apps. After answering the questions, people obtain a receipt, slowly generated by a thermal printer to build up anticipation and implicitly foster creepiness – on average it takes 1:30 minutes for the receipt to print. It contains a comprehensive list of the data that is collected and indefinitely stored as people interact with different products and services. It includes short but detailed examples to help people interpret the data and a checkbox that explicitly invites them to reflect upon and indicate the perceived creepiness. The receipt physicalizes the personal data trail. Its length conveys a type of data in itself, how much data is collected.

To populate the receipt with accurate information and examples, I made use of my rights of access and data portability in the GDPR [65] and requested a copy of my personal data from the different data controllers listed above. In total, I made 28 data portability requests. I reached out individually to each data controller, as indicated in their privacy policy. At the time, I was based in The Netherlands, hence four of the requests were made to companies operating primarily in this country; the public transport company (1 request), the supermarkets (2 requests), and the bank (1 request). The information on the receipt corresponding to these three entities is limited to the specific Dutch context and might not translate to other contexts and countries. The other twenty-four requests were made to companies operating in an international context.

For me, requesting and obtaining a copy of their data was a lengthy, confronting, and overwhelming process. Especially when the data was delivered digitally (i.e., a USB) and physically (i.e., printed files) by mail to her home address. The receipt aims to create a similar experience for dataslip users, by confronting them with the length and depth of their personal data.

6.4.2. Reacting to Dataslip

I demonstrated dataslip for two consecutive days at the Makers Faire in Delft, a local event open to community members of all ages. In doing so, I aimed to investigate individuals' practices, expectations, and concerns around personal data collection. During the event, I invited attendees who approached dataslip organically to interact with it and obtain their receipts. Then, I prompted them to record (1) their feelings or reactions to their receipts and (2) what they would like to do with data from the slip on a post-it note. Similar to [176], I found this set-up created conditions for discussion and exploration.

The event was advertised locally to residents of Delft and the surrounding area through flyers and social media posts. During the event, I informed attendees about our research and they consented verbally to participate by recording their thoughts on post-it notes. This activity was reviewed and approved by the Human Research Ethics Committee at TU Delft. Due to the set-up of the event, I did not collect any demographic or personally identifiable information from attendees.

About one hundred attendees obtained their receipts over the two-day event, and fifty-five left us post-it notes (A1-A55). After the event, Renee and I digitalized each post-it note and used clustering and visual mapping techniques to outline the main practices and challenges in the online whiteboard tool Miro.

Reactions to Dataslip

The majority of attendees expressed a sense of bewilderment while obtaining their receipt; *"OMG. It's too much!!! I feel uncomfortable because I don't know how to get out"* (A1). It was cultivated by the anticipation and waiting for a receipt that kept printing and printing. Similarly, a recurring reaction from attendees after inspecting their receipt was

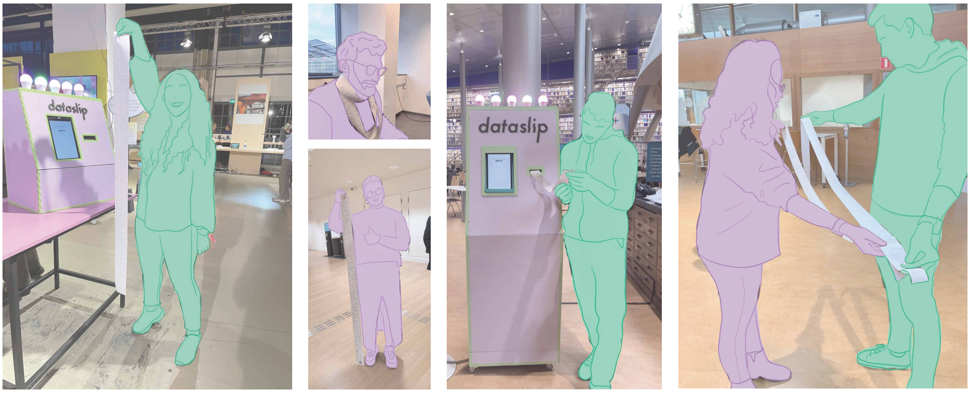


Figure 6.3: Interactions with dataslip.

concern about potentially exposing their hidden secrets.

For most attendees, the information on the receipt was new, crazy, and scary. *"My [receipt] is a scary way to look at how much information I am sharing with every click."* (A45) Nonetheless, some attendees were already familiar with the data collection and storage practices of digital product and service providers, yet they found the receipt format telling. *"It is not new to me, but seeing it in a receipt makes it different."* (A28).

Practices and Expectations

Although most of the attendees expressed concern about the length of their receipts, some saw it as valuable and reflected on the potential of using such data. For example, to learn more about themselves or to make their daily routine more convenient, *"I would like to use my data to automate my house, for instance, open my garage when I'm getting close."* (A32) In fact, some attendees already use some of the data from their receipts to gain personal insights and make data-informed decisions. *"I use my data to categorize my spending and get an overview"* (A2). Another potential application that attendees identified is to use personal data as proof or evidence that they were involved in certain activity during a specific time, *"I can use it as proof that something happened, that I was in the supermarket for instance"* (A12). In this way, the pervasive and timestamped nature of personal data is perceived as beneficial as it could serve as evidence in court or with the relevant authorities if necessary.

Some attendees also challenged the information on their receipts based on their current practices. For instance, *"I use fake birthdates and I have many profiles of myself"* (A40). Here, using different profiles and settings meant having different fragments of data associated with different versions of himself, and therefore, fragmented across multiple receipts.

Challenges and Concerns

Through clustering the different post-it notes Renee and I identified the following challenges and concerns:

- *Privacy trade-off:* Community members recognized a trade-off between (protecting) their privacy and using products and services that are convenient (e.g., personalized public transport card) or necessary for everyday activities (e.g., credit or debit cards). *"I try to protect my data but at the same time I use many apps, it is scary!"* (A8).
- *Transparency*: What do we say yes to when we agree to use a digital product or service? Why do digital product and service providers need personal data? Community members underlined the lack of adequate information about these aspects; essential for making informed decisions. *"I would like to know what I'm saying yes to, with things like cookies. What is a cookie?"* (A17).
- *Unbalanced Distribution of Benefits:* Who benefits from the collection of personal data? Mostly private companies and public services. However, community members expressed their willingness to benefit as well. *"I want to trade my data for money!!!"* (A6).
- *Temporality and Permanence:* Community members questioned the permanence of personal data. It is continuously collected, and indefinitely stored. *"Why do they need to keep my data forever?"* (A23) It is constantly growing and it is not necessarily easy to stop it from growing by deleting it or requesting to be forgotten.
- *Having a Choice:* Community members expressed a need for agency and autonomy when it comes to the distribution of their personal data. *"I would like to be able to choose what to share and what not."* (A39).

6.4.3. Speculating with Dataslip

To continue the conversations and debate that were triggered by the initial interactions with dataslip, Renee and I aimed to further contextualize those conversations and make a more hopeful and explicit contribution to the future practices around the collection and use of personal data. In doing so, we specifically respond to the first of Kozubaev et al's modes of reflection [104]: to imagine one specific future to close down other possible futures and open up conversations about that future in detail. We also respond to the fourth and fifth modes of reflection, by engaging with the current situation and further involving the community, and building upon earlier speculation, by involving dataslip as a co-speculator in the futuring exercise.

For this activity, we used the challenges and concerns identified above as prompts for further speculation. We involved primary school students, as their generation is an inherent stakeholder in the personal data practices of the future. Hence, it is valuable and important that they are involved in shaping that future. Additionally, we considered the unfiltered creativity of primary school children an inspiring resource for the further continuation of the project. We conducted two creative workshops (W1-2), each with 20 primary school students aged 10-12. The workshops were held in this context as the school was running a special curriculum on data, the internet, and online behavior at the time and invited us to participate with dataslip. Both the students and their parents or guardians consented and assented to their participation in the workshops. This activity was reviewed and approved by the Human Research Ethics Committee at TU Delft. The workshops were structured through three activities. First, we gave a short introduction to personal data and online behavior. Second, we invited students to obtain their receipts. Third, we invited students to form groups of 4-5 (G1-5) people and propose a solution to one of the five challenges identified above. For this activity, we provided the groups of students with an A3 paper sheet with a challenge statement, briefly summarizing each challenge, and creative material, including post-its, colors, and markers.

Some of the potential solutions proposed by the groups of students included:

- *"You can have an app that is safer and configures your privacy settings."* (W1, G1)
- *"Creating a warning that comes up whenever data is going to be collected about you."* (W1, G2)
- *"A one-day limit to personal data storage."* (W1, G5)
- *"Getting money every time you give websites your data."* (W2, G4)
- *"An automatic sorter for things people want and don't want to change."* (W2, G3)

After the workshop sessions, Renee and I analyzed the results and clustered the outcomes according to the topics that the children came up with. The goal of our analysis was to distill future prospects and positive, utopian scenarios to bring up to future dataslip users to counterbalance the creepy feeling they were left with after interacting with it. We wanted to help them understand that the future is not inevitable and that there is still a lot that we, as researchers, designers, and users of personal data technologies, can do to shape the future. For each cluster, we then selected one idea to further develop into a future scenario. The selection criteria included perceived feasibility and relevance to future practices around personal data collection and use. Additionally, we related each idea to the current practices that were described by community members in the earlier interactions with dataslip.

Hence, we further polished the ideas so that they conceptually responded to the themes that came out of the earlier discussions. To keep them easy to digest, each of the scenarios was translated into a postcard, with an image that illustrated the concept on the front, and a more detailed explanation of the scenario on the back. These postcards were then printed to be brought along to future dataslip events and exhibitions. To further close down the scenarios in order to open up thinking about them critically and to make a stronger connection to the initial dataslip interaction, each scenario also comes with a specific form that can be printed as a receipt through the dataslip machine and filled in by the participant to further personalize and contextualize their interaction.

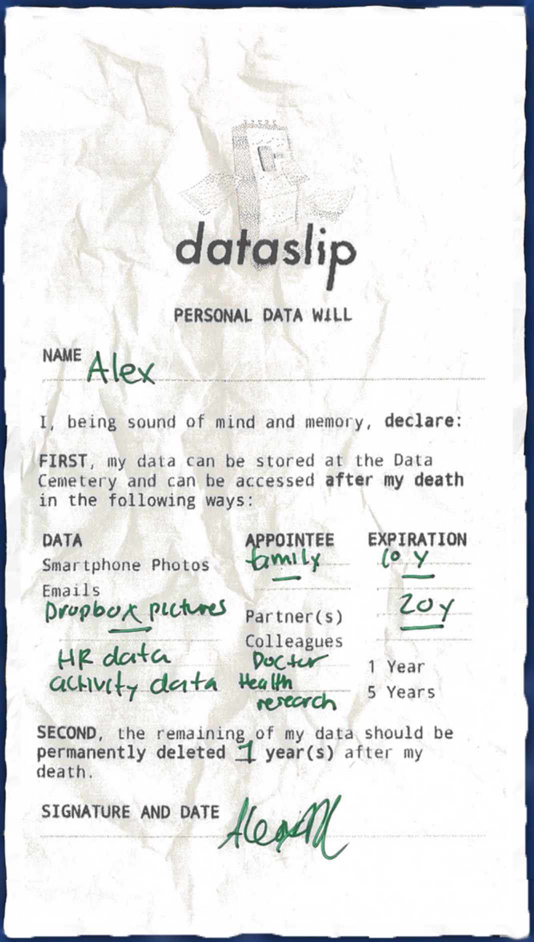
6.5. Speculative Scenarios

The five scenarios and their corresponding receipts are presented on the next pages.

Data Cemetery

A future where data doesn't have to live forever

In this future, when you die, your data dies with you, and your online identities are removed. Before you die, you can decide to leave part of your data in the Data Cemetery so it stays accessible for your loved ones. They can then visit the cemetery to bring back memories with your data and to look up important information. All the data is stored locally and securely at the cemetery so your data doesn't start living its own life when your life ends. Making use of the Data Cemetery is optional, in the end, you decide what you share and with whom!



dataslip

PERSONAL DATA WILL

NAME Alex

I, being sound of mind and memory, **declare:**

FIRST, my data can be stored at the Data Cemetery and can be accessed **after my death** in the following ways:

DATA	**APPOINTEE**	**EXPIRATION**
Smartphone Photos	family	10 y
Emails		
Dropbox pictures	Partner(s)	20 y
	Colleagues	
HR data	Doctor	1 Year
activity data	Health research	5 Years

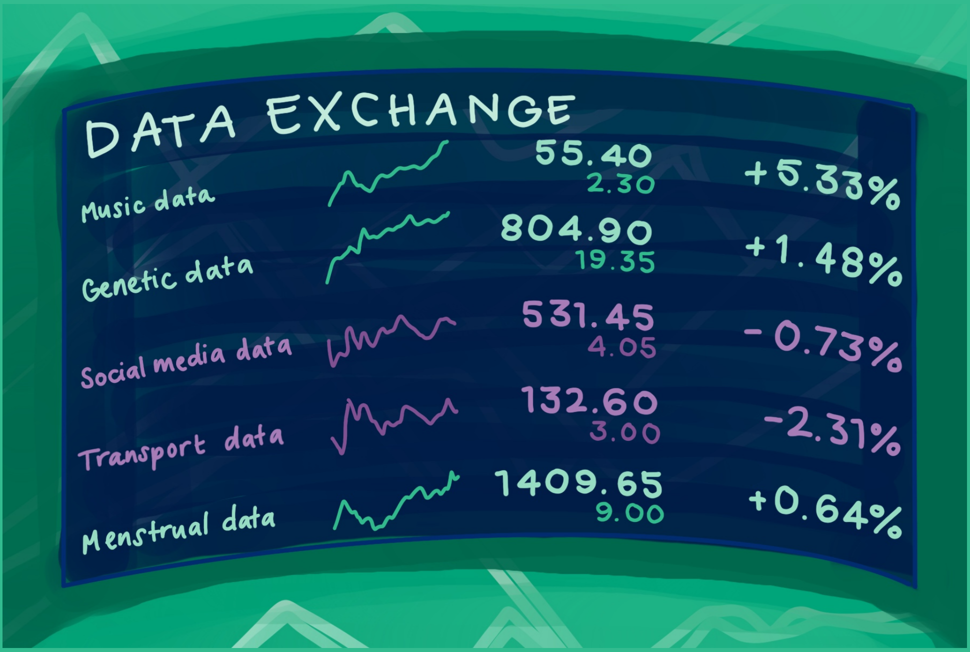
SECOND, the remaining of my data should be **permanently deleted** 1 **year(s)** after my death.

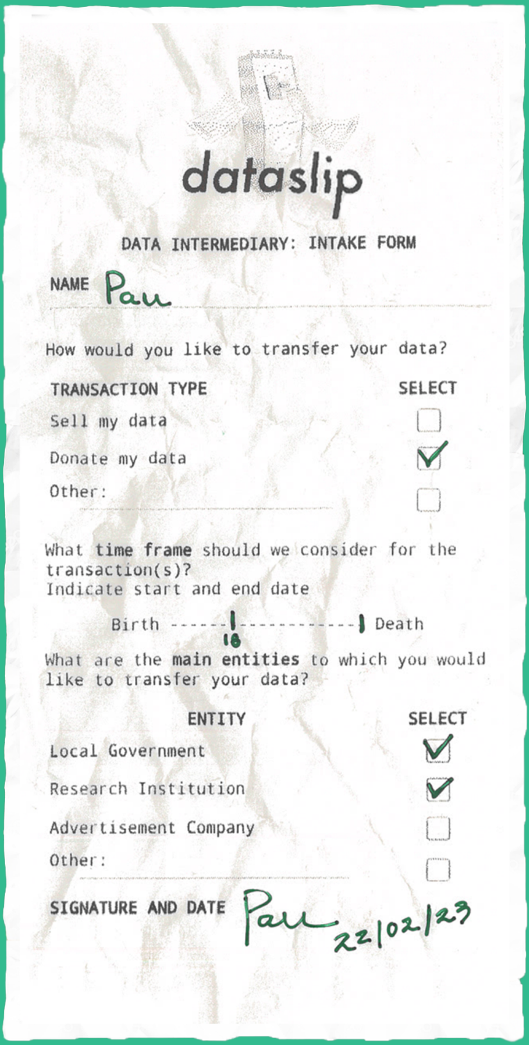
SIGNATURE AND DATE

Data Intermediary

A future where allowing access to your data brings you value.

In this future, you have your own data intermediary who manages who has access to your personal data. After an initial intake conversation where you explain your personal values with regard to data sharing, the intermediary invests with your data. Using this service means you can make some profit off your data and contribute it to the causes that you find important, such as clinical trials or marketing research. Your intermediary will update you regularly so you can decide whether you are still happy with the course of the investment.





dataslip

DATA INTERMEDIARY: INTAKE FORM

NAME Pau

How would you like to transfer your data?

TRANSACTION TYPE	SELECT
Sell my data	☐
Donate my data	☑
Other:	☐

What **time frame** should we consider for the transaction(s)?
Indicate start and end date

Birth ------|------------| Death
18

What are the **main entities** to which you would like to transfer your data?

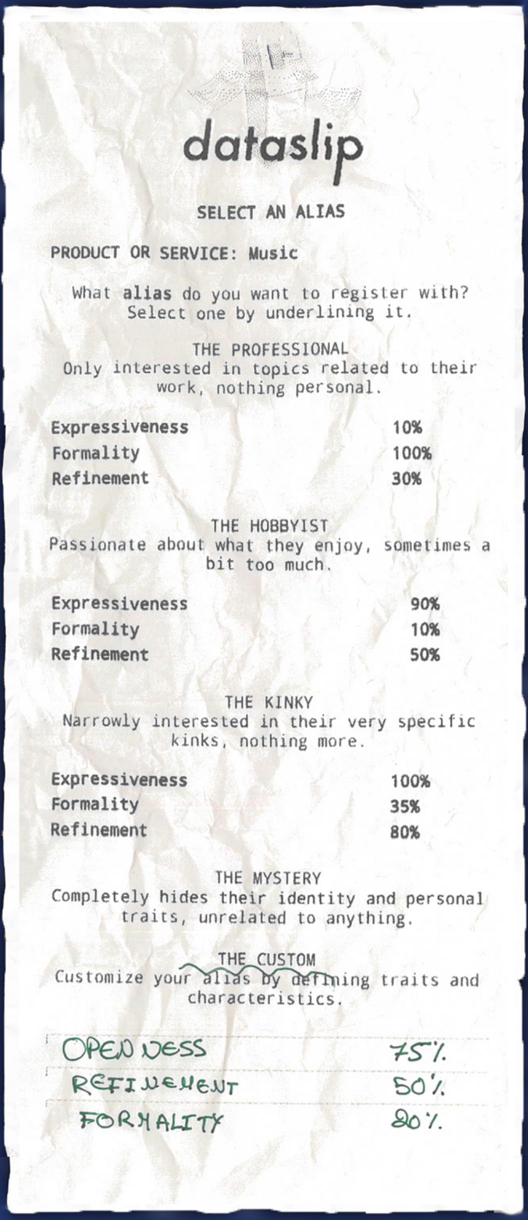
ENTITY	SELECT
Local Government	☑
Research Institution	☑
Advertisement Company	☐
Other:	☐

SIGNATURE AND DATE Pau 22/02/23

Data Alias

A future where you decide what data best suits you

In this future, you can switch between different social profiles in your online life as easily as in real life. You can use your fully personal profile while using a dating app, and switch to your professional profile when applying for a mortgage. They are completely separated, so you will no longer get ads for pregnancy tests while looking something up for work, and your colleagues don't have to know about your geeky gamer side if you don't want them to. A work-life balance dream come true!



dataslip

SELECT AN ALIAS

PRODUCT OR SERVICE: Music

What **alias** do you want to register with?
Select one by underlining it.

THE PROFESSIONAL
Only interested in topics related to their work, nothing personal.

Expressiveness	**10%**
Formality	**100%**
Refinement	**30%**

THE HOBBYIST
Passionate about what they enjoy, sometimes a bit too much.

Expressiveness	**90%**
Formality	**10%**
Refinement	**50%**

THE KINKY
Narrowly interested in their very specific kinks, nothing more.

Expressiveness	**100%**
Formality	**35%**
Refinement	**80%**

THE MYSTERY
Completely hides their identity and personal traits, unrelated to anything.

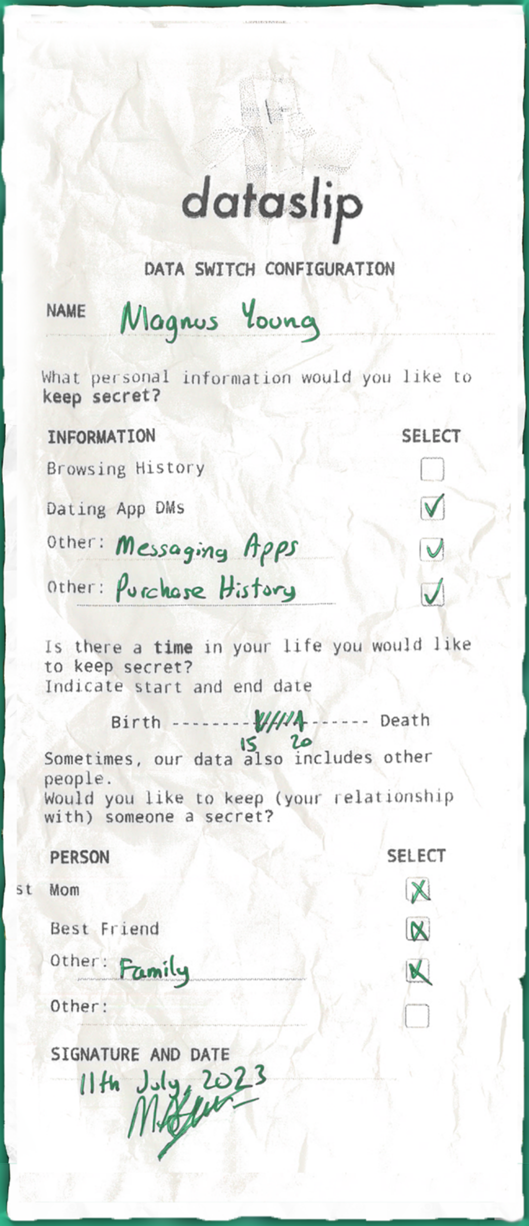
THE CUSTOM
Customize your alias by defining traits and characteristics.

OPENNESS	75%
REFINEMENT	50%
FORMALITY	80%

Data Switch

A future where you set boundaries to protect your data and secrets

In this future, setting your preferences with every product and service you use is something from the past. You set your personal boundaries about data collection and storage once and decide which secrets you never wish to reveal. Whenever you use a product or service, your preferences are configured automatically. If you try to use a product that collects data you would never share, you will receive a warning. The warning explains which of your boundaries is being violated and it is up to you to decide what to do: proceed with caution, change your values, or file a complaint?



dataslip

DATA SWITCH CONFIGURATION

NAME Magnus Young

What personal information would you like to **keep secret?**

INFORMATION	SELECT
Browsing History	☐
Dating App DMs	☑
Other: Messaging Apps	☑
Other: Purchase History	☑

Is there a **time** in your life you would like to keep secret?
Indicate start and end date

Birth --------////------ Death
15 20

Sometimes, our data also includes other people.
Would you like to keep (your relationship with) someone a secret?

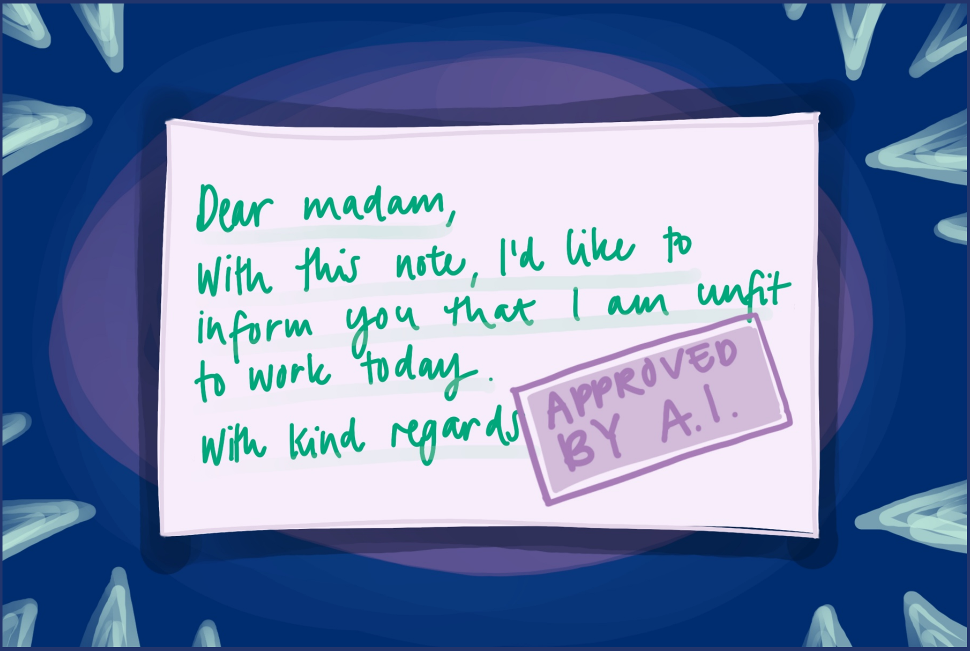
PERSON	SELECT
st Mom	☒
Best Friend	☒
Other: Family	☒
Other:	☐

SIGNATURE AND DATE
11th July, 2023

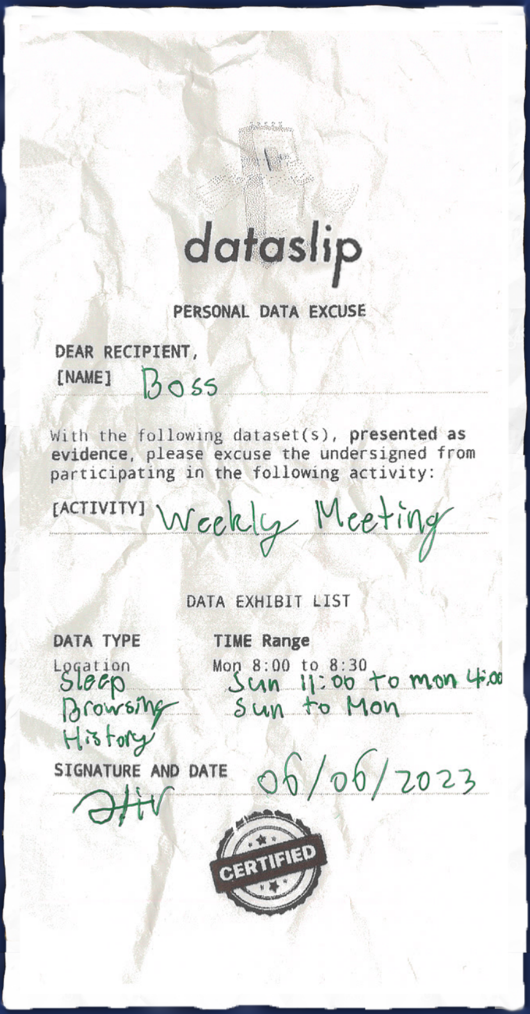
Data Jurisdiction

A future where data serves as evidence

In this future, you can use your personal data as official evidence to protect yourself in diverse situations. The Data Jurisdictor can interpret diverse sources of data and present them as evidence. For example, you might be having painful menstrual cramps but your boss doesn't believe that justifies you staying home. In this case, the Data Jurisdiction software can write a sick note for you, based on your menstrual tracking data. Start using your data to demand justice now!



Dear madam,
With this note, I'd like to inform you that I am unfit to work today.
With kind regards



dataslip

PERSONAL DATA EXCUSE

DEAR RECIPIENT,
[NAME] Boss

With the following dataset(s), **presented as evidence**, please excuse the undersigned from participating in the following activity:

[ACTIVITY] Weekly Meeting

DATA EXHIBIT LIST

DATA TYPE	TIME Range
Location	Mon 8:00 to 8:30
Sleep	Sun 11:00 to mon 4:00
Browsing History	Sun to Mon

SIGNATURE AND DATE 06/06/2023

CERTIFIED

6.6. Discussion

With the design of dataslip, we provided attendees with a tangible representation of their personal data trail. The tangibility of the receipt allowed attendees to engage with their personal data in different ways; inspecting every single line, sharing it with their friends and family, comparing the length with their own height or someone else's receipt, or wearing it as a hat or scarf. In the receipt, we added a checkbox, explicitly prompting participants to reflect on the creepiness of each type of data. Further, to implicitly foster creepiness, we used a thermal printer, that slowly printed every line on the receipt, building up anticipation and confrontation. Even if and when attendees were expecting their receipts to be long, they never expected it to be that long.

Similar to [177], we found that creepiness is powerful but temporary. "What now?", "What should I do then?" were some of the reactions of attendees during the community event. We engage with creepiness as a starting point for reflection and speculation. Yet, we move from creepiness towards empowerment and hopefulness. This is evidenced by the five future scenarios: (1) in the data cemetery, data could allow our loved ones to re-live us; (2) with the data intermediary, we could direct our data towards causes that matter because of our values or fair monetary compensation; (3) with the data alias, data could be used to conveniently inform automated tools – as long as they align with our preferred role; (4) with the data switch, data could serve as a defense mechanism to protect our privacy, by helping us define and control our personal boundaries; and (5) with the data jurisdiction, data could help us convince others.

We note that some of the challenges that our five scenarios respond to have been extensively approached and discussed in previous literature across various domains, including philosophy of technology, law, human-computer interaction, and computer science (e.g., underlining the power imbalances [51, 155], fostering transparency [54, 74], and supporting privacy and personal boundaries [49, 137, 150]). Although these challenges might not be considered "novel" they underline that, when it comes to personal data, individuals are the main parties involved; "data are people" [217]. They have valid and informed concerns and already rely on innovative practices to address them. Our participatory approach involving community members of all ages realizes the importance of engaging with the real world and the public in futuring and overall research. Thus, we invite and encourage researchers across these domains to creatively involve members of the public in their projects and activities.

Our research and practice can support individuals not only to become aware but also to envision and demand change and benefit from their data. In our case, by making the creepiness of personal data tangible and experiential we empowered individuals to reflect upon the potential value of their personal data. In line with the utopian and dystopian mindsets discussed in the introduction, we see this dual experience as a way to shed light on both sides of the story. The experience with the dataslip underlines and criticizes current practices, and the future scenarios suggest alternative ways in which we might overcome the current issues. This participatory approach may be relevant in other creepy contexts, such as climate change where the overwhelmedness and the creepiness of the current situation is often perceived as disempowering.

Through our research approach, we initiate and maintain a two-way conversation with community members; where they interacted with and responded to dataslip, and we, as design researchers, responded to their responses. Similarly, by presenting dataslip and the accompanying future scenarios in detail, we wish to respond to Kozubaev et al.'s [104] call for more continuous futuring in their fifth mode of reflection. First, by describing how we used attendees' responses to their interactions with dataslip in the workshops and future scenarios that followed, we present a process of iteration and call and response between the participants and ourselves. Secondly, we hope that by presenting dataslip and the scenarios in detail, others will also be able to continue building on these future scenarios and help us to mature the ideas represented by these.

6.7. Notes

The making of dataslip was a collaborative effort and a great excuse to spend time with my PhD colleagues doing something away from a screen. Through this process, I explored the many tools at the PMB with James Broadhead until we almost learned to follow the protocol, I received an amazing "how to build a box" tutorial from Wo Meijer, I deciphered the many mysteries of (thermal) printing with Vasilis Milias, and I spent some time gossiping and painting after work with Hosana Morales, Roos Teeuwen, Di Yan, Anotnio Manesco, and Marilia Silva. Since its debut at IDE in December 2022, I have exhibited dataslip at various locations and events, including the TU Delft Library (Delft), SURF "Privacy not Included" 2023 (Utrecht), ICT Open 2023 (Utrecht), Delft Maker Faire 2023 (Delft), CHI'23 (Hamburg), Interact'23 (York), MozFest House 2023 (Amsterdam), Dutch Design Week 2023 (Eindhoven), ThingsCon 2023 (Rotterdam), Universidad de Los Andes (Bogotá), and TEI'24 (Cork). Gabi, my partner, has spent countless weekends accompanying me and helping me transport and assemble dataslip – he is very good at explaining it by now. Antonio Manesco and Marilia Silva have also attended several dataslip events, keeping Gabi company and helping me take pictures. I want to acknowledge their help as it was crucial to making all members of the dataslip "family"[1] and I also want to show the many hands behind this project (Fig 6.4).

6.8. Chapter Takeaways

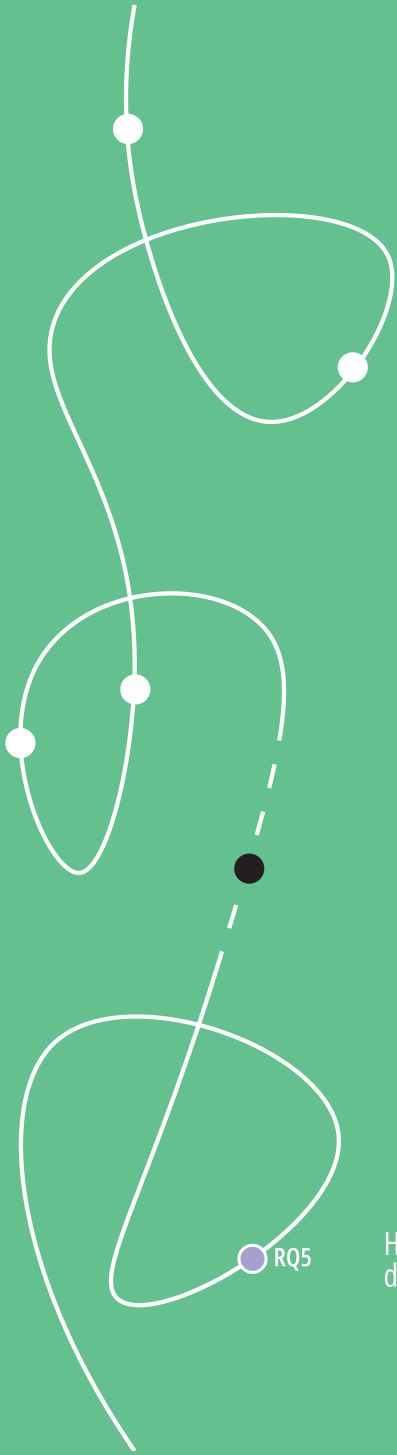
Although not directly integrated with the central Research through Design process, the design of dataslip explores some of the core challenges and themes relating to Sensitive Data Donation: awareness about the content and characteristics of data, data sensitivity, and empowerment through and from data. People's reactions to their interaction with dataslip, and the challenges and concerns we identified illustrate critical considerations for data donation research – and more broadly, activities around collecting and using personal data. Similarly, the five speculative scenarios, responding to these challenges, pose important questions for Sensitive Data Donation: What data should "die" at the end of a data donation project? How can donors benefit through and from their data? How

---

[1] James Broadhead and I made smaller versions of dataslip so they would be easier to carry to various events. He jokes that we built a whole family. It includes a Colombian uncle built by Andres Fernando Pedraza Hernandes and Alva Avila at Universidad de Los Andes.

Figure 6.4: Colleagues and friends helping with dataslip.

can donors set meaningful boundaries around themselves and their data? I will continue to explore these in the following chapters.



How can data donation promote different degrees of participation in research?

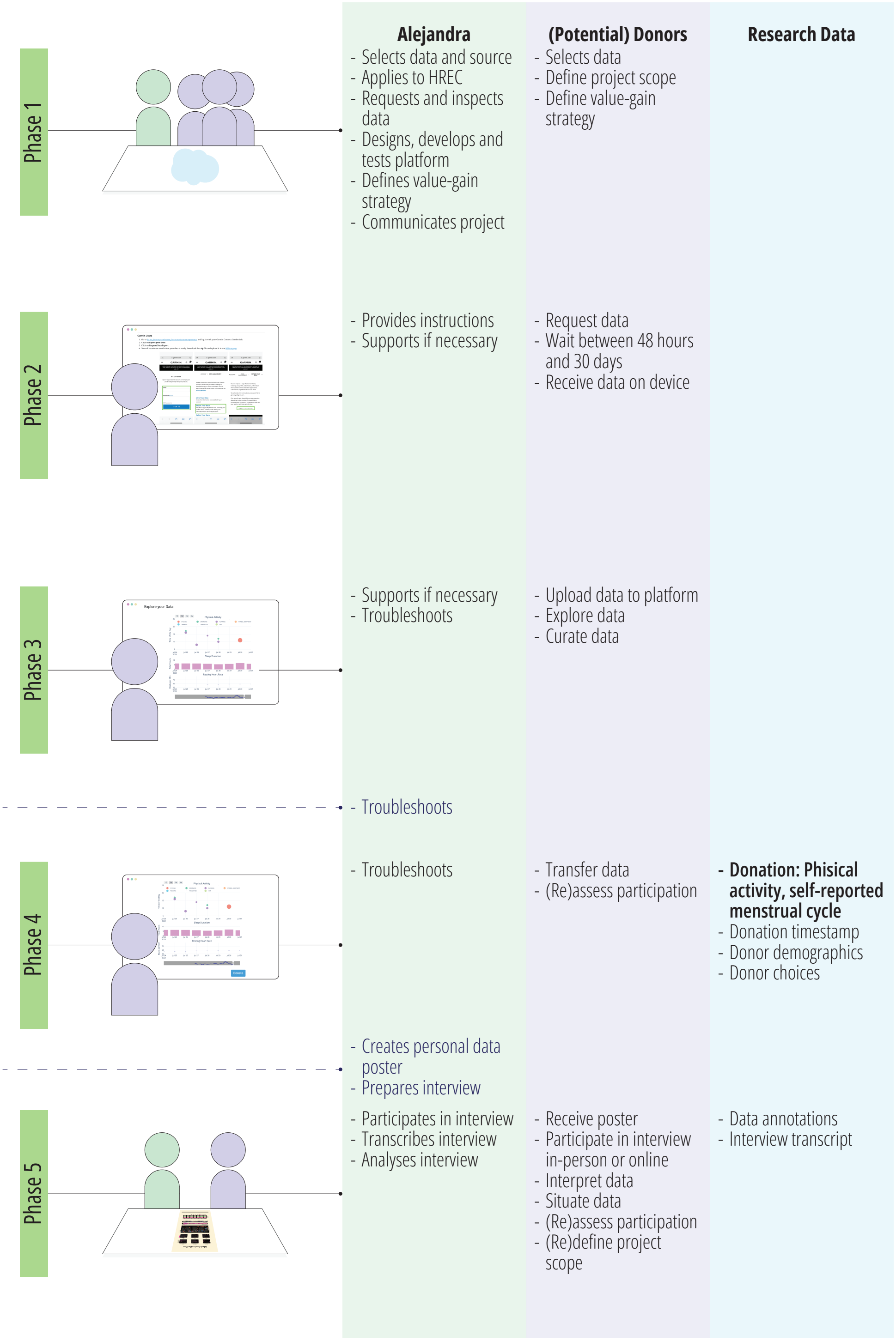
7

Promoting Different Degrees of Participation in Data Donation

Abstract

In this chapter, I investigate how data donation can promote different degrees of participation, specifically contributors, collaborators, and co-creators (RQ5). To do so, I design and develop the third iteration of a sensitive data donation approach embodied by a digital data donation platform. It synthesizes the lessons learned in the previous two iterations, described in Chapter 3 and Chapter 4. I apply it in a research project aiming to investigate athletes' perceptions of the impact of their menstrual cycle in sports through their physical activity data. I focus on physical activity from wearable devices, including all Garmin devices and all devices compatible with Apple Health. 20 athletes participated by donating their data, interpreting and contextualizing them and further contributing to the various research activities. I discuss their experience and conclude this chapter by discussing how promoting different forms of participation results in diverse and messy datasets which invite researchers to relate differently to the research process and the resulting data.

This chapter draws on the following publication: **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2024) 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).



	Alejandra	(Potential) Donors	Research Data
Phase 1	- Selects data and source - Applies to HREC - Requests and inspects data - Designs, develops and tests platform - Defines value-gain strategy - Communicates project	- Selects data - Define project scope - Define value-gain strategy	
Phase 2	- Provides instructions - Supports if necessary	- Request data - Wait between 48 hours and 30 days - Receive data on device	
Phase 3	- Supports if necessary - Troubleshoots	- Upload data to platform - Explore data - Curate data	
	- Troubleshoots		
Phase 4	- Troubleshoots	- Transfer data - (Re)assess participation	**- Donation: Phisical activity, self-reported menstrual cycle** - Donation timestamp - Donor demographics - Donor choices
	- Creates personal data poster - Prepares interview		
Phase 5	- Participates in interview - Transcribes interview - Analyses interview	- Receive poster - Participate in interview in-person or online - Interpret data - Situate data - (Re)assess participation - (Re)define project scope	- Data annotations - Interview transcript

Figure 7.1: Overview of the case study showing the involved people and activities mapped around the five phases of Sensitive Data Donation.

7.1. Introduction

This chapter investigates how data donation can promote different degrees of participation in scientific research through the third (and last) Research through Design (RtD) iteration. I design and develop an instance of the data donation approach described in Chapter 2 – embodied by a digital data donation platform. Additionally, I incorporate the insights gained from previous iterations (Chapter 3 and Chapter 4). I apply this instance of data donation at the intersection of sports performance and the menstrual cycle. In this context, athletes are often misrepresented [45, 62]. Thus, their participation can foreground their knowledge and experience(s) behind their digital-trace data. Aligned with the principle of Shared Goals **(P5)**, this iteration focuses on supporting different degrees of participation in data donation and inviting interested potential donors to relate to and shape the research project and goals from the start.

The principle of Shared Goals **(P5)** builds upon a framework for public participation in scientific research proposed by Shirk et al [175]. They define the degree of public participation in research processes as "the extent to which individuals are involved in the process of scientific research." They propose five types of projects based on the degree of participation: (1) *contractual*, where communities ask professional researchers to conduct a specific research project; (2) *contributory*, where researchers invite members of the public to contribute data; (3) *collaborative*, where researchers invite members of the public to contribute data, analyze data, and/or disseminate findings; (4) *co-created*, where researchers invite members of the public to design the research project and some members are actively involved in most aspects of the research process; and (5) *collegial*, where members of the public conduct research independently. I specifically focus on exploring how members of the public can contribute, collaborate, or co-create to scientific research projects through data donation.

As mentioned in Chapter 2.2.2, data donation research projects have been mainly *contributory*: donors contribute to a project by transferring their data and might further contribute by augmenting it or annotating it. The empirical studies described in Chapters 3 and 4 illustrate an attempt to invite donors to participate as *contributors* or *collaborators*, as donors contribute by transferring their data and are invited to actively participating in its interpretation and contextualization. To date, no projects have invited (potential) donors to participate as *co-creators*, limiting the perspectives included and accounted for around critical research-design decisions.

Overall, with respect to the data donation approaches described in Chapter 3 and Chapter 4, the data donation process described in this iteration comprises dedicated activities for each of the three degrees of participation:

- **Contributors:** Donors who decide to participate as contributors primarily request and receive a copy of their data, explore it, and decide what (not) to transfer through the data donation platform. Additionally, they participate in an interview about their data donation experience.
- **Collaborators:** Donors who decide to participate as collaborators, in addition to the above, participate in a one-on-one session where they are invited to interpret, contextualize, and analyze their data.

- **Co-Creators:** Donors who decide to participate as co-creators, in addition to the above, can opt to participate in a session to scope the research goals and activities and in meetings throughout the process to discuss the partial results and research direction.

These activities are described in Section 7.3. The resulting data donation process adopts the five principles of Sensitive Data Donation, namely Balanced Value **(P1)**, Sensitive Data **(P2)**, Multiple Knowledge(s) **(P3)**, Ongoing Consent **(P4)**, and Shared Goals **(P5)**. This chapter reports on donors' motivations and expectations throughout the data donation process according to their degree of participation. Additionally, it describes the characteristics of the donated data and how it is shaped by the degree of participation and other choices along the data donation process. This chapter concludes with a reflection on applying the five principles of Sensitive Data Donation.

7.2. Intimate Context: Sports and the Menstrual Cycle

I situate and apply data donation in the context of sports performance and the menstrual cycle. Athletes increasingly rely on digital technologies to track their physical activity and increase their performance [43]. These technologies play an instrumental role in measuring performance and feeding data back to athletes (e.g., resting heart rate, respiratory rate, sleep duration), and an experiential role in supporting and enhancing the sports experience, allowing athletes to have a closer connection to their bodies [193, 207]. Although digital technologies provide feedback and advice to athletes in various aspects, including recovery and training, one vital health indicator [64] is often overlooked; the menstrual cycle[1]. Likewise, most of the research in sports science – informing the design and development of these devices – has been conducted with men [45, 47]. Yet, findings are inappropriately applied to women athletes [62].

The impact of the menstrual cycle on physical performance is recognized as a key consideration for sports. Previous research demonstrates that hormonal fluctuations throughout the menstrual cycle affect athletes during training and competition (e.g., [39, 110, 128]). Similarly, physical activity impacts the menstrual cycle. For instance, amenorrhea (i.e., the absence of menstruation) is highly prevalent among athletes, yet, it is often not discussed or reported [206]. Elite sports organizations such as the Chelsea Football Club and the Red Bull High-Performance Division are using FITR Woman [66], a commercial smartphone app to track athletes' performance and bodily symptoms during the various phases of the menstrual cycle.[2]

The app invites athletes to self-report their physical activity and symptoms (e.g., cravings, sensitive breasts) and offers training and nutrition advice. Feminist HCI research has widely criticized these types of apps, arguing that by collecting intimate data and translating it into more or less accurate predictions they restrict other types of knowledge [38, 69, 95, 164, 185, 214]. I build on this call by exploring the perceived impact of the menstrual cycle on sports and collectively envisioning ways to generate new forms of

---

[1] Notably, the menstrual cycle was also disregarded in the initial release of Apple's Health app in 2014 [59]

[2] The menstrual cycle phases include, early follicular, late follicular, ovulatory, early luteal, mid-luteal and late luteal. It starts with menstruation during the early follicular phase [128].

knowledge about it. Following the Feminist tradition of positionality and reflexivity [15, 16, 52], I report that I am an amateur athlete who menstruates, with a strong motivation to undertake this research.

7.3. Data Donation Process

In this section, I describe how I apply a version of the Sensitive Data Donation method (Fig. 7.2) with a focus on promoting different degrees of participation, aligned with the principle of Shared Goals **(P5)**. I integrate the five principles of Sensitive Data Donation into the design of the data donation approach. The activities, participants, and data collected during each phase are summarized in Figure 7.1.

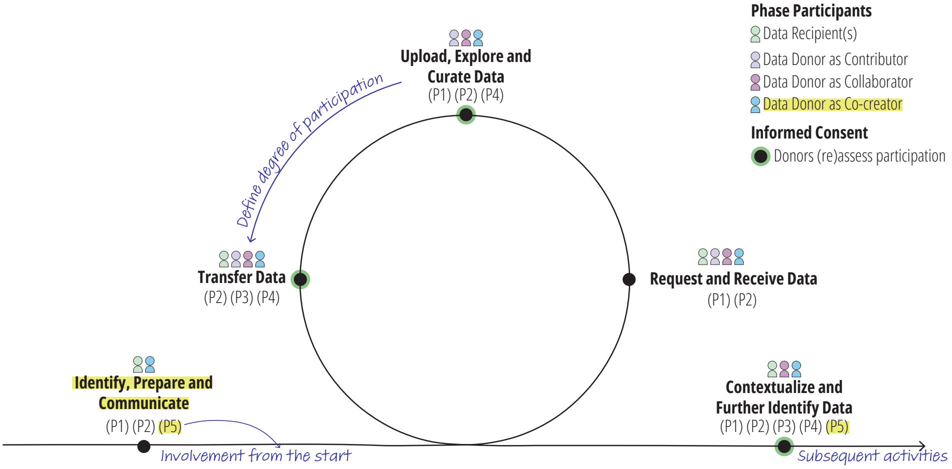


Figure 7.2: Simplification of the Sensitive Data Donation method for this iteration, highlighting in yellow the main focus of this chapter.

Phase 1: Identify, Prepare and Communicate

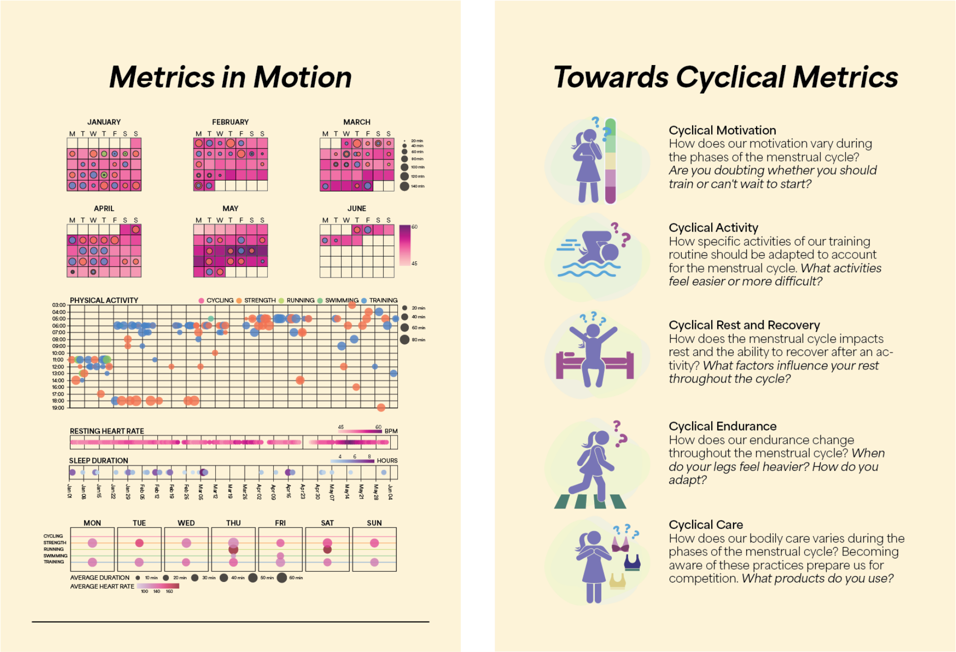
I scoped the research goals together with four athletes, who later became co-creators, and a sports gynecologist. I contacted them by posting flyers in local sports associations inviting athletes with vulvas to a creative session focused on menstruation in sports and I expressed my appreciation to them by inviting them to a Q&A with the sports gynecologists. They proposed to expand the focus: from investigating the perceived impact of the menstrual cycle in sports to envisioning new "metrics" or ways to generate knowledge about it. Together, we discussed relevant data sources and decided to focus on: (1) menstrual cycle, (2) sleep, (3) heart rate, and (4) physical activity, data as proxies recovery, and performance. Finally, we co-defined a value-gain strategy responding to athletes' wish to learn about their data and how to interpret it. It comprised a (printed or digital) poster representing their data (Fig. 7.3.a) with guiding questions on how to interpret them, as athletes often enjoy displaying their achievements and mementos (e.g., bib numbers and medals) [184], and a commitment to communicate our results throughout the process (Fig. 7.3.b). Thus, aligned with previous research [76, 155, 163], I aimed to honor donors' contributions and efforts by carefully engaging with their data, creating opportunities and artifacts for them to engage with their data, and building and maintaining a relationship. I intentionally decided not to incentivize donors financially due to the ethical precedent of not compensating research subjects as it could limit their ability to offer consent voluntarily [82, 122, 147]. I determined that voluntariness and autonomy should prevail around transactions of sensitive data.

I then selected data sources and determined how to offer donors meaningful options around disclosure that would allow me to conduct the research. I invited donors to self-report what they considered relevant about their menstrual cycle and to transfer retrospective data collected by Garmin, Apple Watch, or other devices compatible with the Apple Health ecosystem. Initially, I wanted to focus only on Garmin devices – designed with sport-specific features and most commonly used by athletes [135] – but I opted for more variety to lower the participation threshold and allow for participation by athletes who do not own a Garmin device. To prepare the data donation platform to parse and visualize the data, I made several requests to Garmin and Apple Health and reflected on this process through a research journal. Building upon the implementation on Chapter 4, in the platform, I designed an interactive visualization where potential donors could explore their data on a timeline (Fig. 7.4). The x-axis showed the month and year and the y-axis showed the value for sleep and heart rate data, and the time of the day for physical activity data. In the visualization, potential donors could filter the data by activity (e.g., running), and time (e.g., day, week, month, six months). I opted for this representation as previous research describes how viewing data arranged by time helps people reason about it [154].

To invite donors to participate in the case study, between June and July 2023, I reached out to athletes with vulvas who were active users of Garmin, Apple Watch, or other wearable devices compatible with Apple Health worldwide and I invited them to donate their data. I used purposive and snowball sampling by reaching out to sports associations, teams, and individual athletes with a strong social media presence, regardless of the discipline, which shared the 'call to donate' with their communities through social media, newsletters, and mailing lists. Additionally, I distributed flyers at local sports events.

Phase 2: Request and Receive Data

While reaching out to donors, I specified who was conducting the research, what the goal was, who could participate, and how. On the data donation platform, I provided potential donors with a statement describing how I intended to use the data and detailed visual instructions describing how to obtain their data from Garmin or Apple Health. Potential donors could contact me if they had any questions or needed additional information. They participated asynchronously and on their own time in this activity.



Cyclical Motivation
How does our motivation vary during the phases of the menstrual cycle? *Are you doubting whether you should train or can't wait to start?*

Cyclical Activity
How specific activities of our training routine should be adapted to account for the menstrual cycle. *What activities feel easier or more difficult?*

Cyclical Rest and Recovery
How does the menstrual cycle impacts rest and the ability to recover after an activity? *What factors influence your rest throughout the cycle?*

Cyclical Endurance
How does our endurance change throughout the menstrual cycle? *When do your legs feel heavier? How do you adapt?*

Cyclical Care
How does our bodily care varies during the phases of the menstrual cycle? Becoming aware of these practices prepare us for competition. *What products do you use?*

(a) Example of a personal data poster, excluding the menstrual cycle data.

(b) Example of one of the infographics with partial results returned to all donors.

Figure 7.3: Artifacts designed for donors to interact with their data and the research to honor their contribution.

Phase 3: Upload, Explore, and Curate Data

On the data donation platform, I invited donors to explore and visualize (Fig. 7.4) their data locally before they decide whether and what to donate. This aimed to promote their interaction and familiarization with the data and support them in identifying and defining initial boundaries around its sensitive elements. I explicitly prompted them to decide if and what types of data they wanted to transfer and for how long (i.e., the last three or six months). I intentionally chose not to collect all available retrospective data (i.e., from when donors started using a device to the date of donation) by bounding the time range. Thus, I adhered to the data minimization principle. Furthermore, potential donors could choose to transfer an ‘overview’ of their data, describing the summary and statistics of each day or activity, or the ‘specifics’ of the data, with timestamped logs for each variable based on the sampling frequency of the device. Finally, potential donors could remove any given activity or data point from the dataset through the interactive visualization. They participated asynchronously and on their own time in this activity. Giving donors various choices around sensitive information disclosure pushed me to consider how to approach partial, messy, or incomplete datasets.

Phase 4: Transfer Data

After exploring and curating their data, donors were prompted to provide initial informed consent through the data donation platform, transfer their data to researchers, and decide on their degree of participation (i.e., contributors, collaborators, collaborators). Donors participated asynchronously and on their own time in this activity. I automatically sent an email to donors who decided to participate as collaborators or co-creators inviting them to schedule a one-on-one session with the first author to further explore, situate, and contextualize their data. From this moment on, donors could modify or revoke their consent through the data donation platform and verbally in direct contact with the researchers. For example, by deciding to stop sharing a type of data or all data. If and when necessary, I recorded these changes through the platform.

Following this activity, I conducted semi-structured interviews with donors who decided to participate as contributors about their data donation experience; these were conducted to gain insights into its application from the contributors' perspective.

(a) Overview of the entire dataset. **(b)** Overview of a week of data.

Figure 7.4: Interactive data exploration visualization on the data donation platform.

Phase 5: Contextualize Data

I invited collaborators and co-creators to a one-on-one session to interpret and contextualize and reflect on their data donation experience. During the interview, I used the personal data poster (Fig. 4.4.a) to facilitate interpretation and reflection. I represented (1) the resting heart rate together with the physical activity types and duration in a calendar view; (2) the type and duration of physical activity for each hour of the day over time; (3) the resting heart rate and sleep duration in a timeline; and (4) the physical activity types and average duration for each day of the week. With the calendar view, I aimed to support donors in identifying patterns and outliers in their data. While with the timeline used to represent physical activity, resting heart rate, and sleep data, the goal was to help donors identify correlations between the different types of data and their menstrual cycle over time. During this session, I also discussed with co-creators their envisioned roles and ways forward for collaboration. Finally, I explicitly prompted donors to (re)assess their participation after the interviews and while sharing partial results with them.

Data Donation process

The resulting data donation process (Fig. 7.5) comprises six voluntary activities. It begins with (potential) donors receiving an invitation to shape the research project, data needs, and goals. It continues with (potential) donors receiving an invitation to participate in the research. Those inclined to participate are invited to follow the instructions to request and receive a copy of their data from Garmin or Apple Health and upload it into the data donation platform, where they are explicitly invited to explore and curate them locally on their device. Here, donors are also invited to decide if and how they want to participate in the subsequent activities. Donors who expressed a desire to participate as collaborators or co-creators are invited to participate in an individual session to interpret, situate, and contextualize their data, and further shape the research project and goals. All donors receive by email or post their personal data poster.

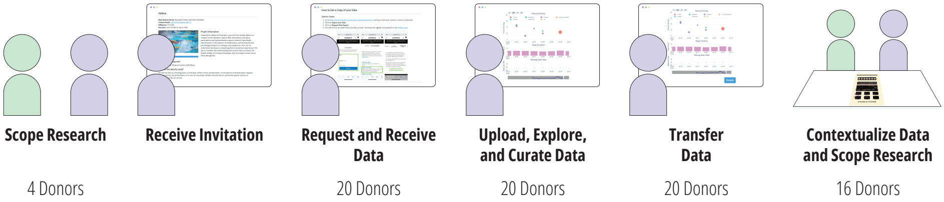


Figure 7.5: Data donation procedure illustrating the various activities, mediated by the data donation platform, in which potential donors participate and how many donors participate in each activity.

7.4. Applying and Reflecting on Data Donation

7.4.1. Participants

20 athletes, professional (n=2), semi-professional (n=8), and amateur (n=10), (referred to as donors D1-20) volunteered to participate in the research by donating their data. They were active users of a Garmin Wearable (n=13), an Apple Watch (n=5), or a third-party smartwatch synchronized with Apple Health (n=2). All donors identified as women, they ranged in age from 22 to 55 years (mean=29, median=32). Donors were located in the European Union (EU) and South America – obtaining a copy of the data, enabled by the GDPR, was also possible for donors outside the EU. Six donors participated in the case study as co-creators, ten donors participated as collaborators, and the remaining four participated as contributors.

7.4.2. Data and Analysis

I conducted semi-structured interviews with all donors around their expectations, information gained from autonomously exploring their data, and the perceived value derived from participation, and with the ten collaborators and six co-creators focused on collaboratively interpreting their data. Additionally, I had follow-up discussions with the six co-creators regarding the preliminary results and the directions of the case study. I anonymized and transcribed the interviews using MS Office and manually reviewed each transcript.

In sum, I generated and collected the following: (1) the researcher's journal entries and reflections, (2) the donor's choices on the data donation platform, including their preferred degree of participation, (3) the donated menstrual cycle, sleep, heart rate, and physical activity data (Fig. 7.7), (4) the donors' perceptions of their data donation experience, in the form of interview transcripts, and (5) the specific insights derived from the contextualization of the data, in the form of interview transcripts. I analyzed the interview transcripts using Reflexive-Thematic Analysis within a constructionist view [33]. I went through the transcripts to familiarize myself with the data and inductively coded the entire dataset using ATLAS.ti; reviewed the codes, and grouped them into tentative themes; and iteratively reviewed and refined the themes. The resulting themes and subthemes were discussed and refined with a co-author of the study, Gerd Kortuem. I report on these in the following sections.

7.5. Results

7.5.1. Motivations, Expectations, and Degrees of Participation

The main factor steering donor's willingness to contribute to the case study was interest in the specific research context. Interest derives from three motives; each carries underlying expectations. The first motive is relatedness, identifying oneself in the context and goals of the project and the community it is directed towards. Relatedness spans multiple experiences, such as being a woman, being an athlete, and having specific experiences (e.g., absent periods) and anxieties (e.g., leaking) with menstruation in sports contexts. Donors who relate to the research perceive it as highly relevant to their personal experience and are therefore eager to contribute. Their underlying expectation is to learn through the research process and from its outcomes and apply these learnings to their individual experiences.

> "This is a topic that is very, very interesting for me because I'm involved with the two subjects or topics that you are touching on within your research which are sports and menstruation. Because I have had during my whole life problems with that. So, it's very, very interesting to me to see what the relation is between these two and how my whole data, that is recorded in these gadgets that I use for sports, how can like help me to understand how my body works." (D4, co-creator)

The second motive is solidarity, recognizing oneself as part of a group and wanting to help others within that group; in this case, by contributing to research. Similar to relatedness, solidarity stems from feelings of sameness and belonging to the community the case study is directed towards (i.e., professional, semi-professional, and amateur athletes). Donors whose primary motivation is solidarity experience feelings of satisfaction or a 'warm glow' – an emotional reward from contributing to research. Their underlying expectation centers on the outcomes of the research and the impact they can have on the daily experience of other athletes.

> "Since I was a little girl I have been playing sports. (...) So, if I can help other generations to have more knowledge and to be part of a world with more sports, happy to help." (D3, co-creator)

The third motive is familiarity, belonging to a specific community (e.g., team, group of friends) already involved with the project and wanting to contribute due to the closeness with this community. Donors whose primary motivation is familiarity already trust the specific community they belong to, and as a result, have increased trust in the research. Their underlying expectation stems from trust and has to do with accountability; they expect researchers to communicate about the next steps and the outcomes of results of the research.

> "Well, [a friend] mentioned to me that you were doing this research and I think she already helped you before with it and she said it was good because she learned as well." (D11, co-creator)

The degrees of participation also carry expectations about what participation entails. The main goal of contributors is to support research; they expect to do so through data; being part of the data and increasing the amount of data. For instance, in response to the 'gender data gap' [50]: *"that's why I did it because I think there is not enough data"* (D13, contributor). Collaborators, in addition to supporting research, also expect to learn something new by interpreting and contextualizing their data; about the research, their data, or themselves, *"I was thinking on how like to learn about this data and how I can have those records to like be useful for my day by day"* (D2, collaborator). The intention to interpret and contextualize the data shapes how they engage in the other stages of the data donation process. For instance, several collaborators (D2, D4, D10, D12, D18) expressed that they did not explore their data on the platform because they knew they were going to do it with us and preferred our guidance. Co-creators, in addition to the above, expect to engage in a more active role throughout various research activities. It includes scoping future iterations of the project, *"I might have some, I don't know, like suggestions for you or like interesting topics that we would like to address in a second version of the project"* (D4, co-creator) and building a community for athletes to share experiences and discuss with others, *"I would love to know other women doing sports, maybe create a community and help each other"* (D3, co-creator).

The degrees of participation adopted by donors significantly shaped the direction of the research project. They demonstrated that (some) individuals can and want to be more actively involved in research. Donors were highly engaged in contributing to the project, inviting others to contribute, and continuing the research. Thus, promoting diverse forms of participation is beneficial to data donation projects as it creates bonds of trust between researchers and participants from a place of boundaries. These result in an engaged community committed to the (sustainability of the) research project.

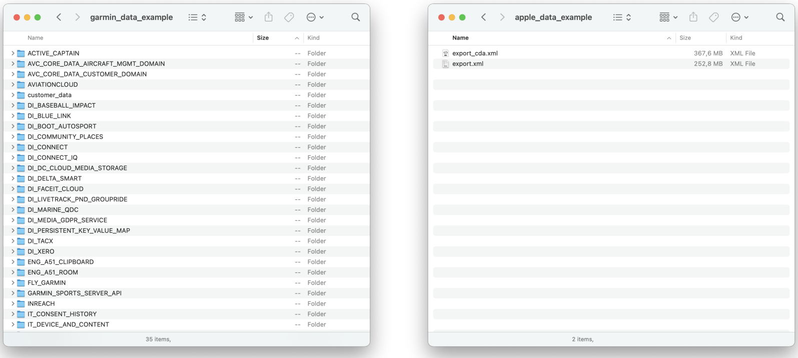
7.5.2. Obtaining, Exploring, Curating, and Contextualizing Data

For most donors, being part of a sensitive data donation process is a new, and almost accidental, experience. First, because of the process itself. Second, because of the decisions that the process invites.

> "I had never done it, and I never used the watch with any intention, like thinking it might be useful to someone, just for myself. And well, it's great that it can be useful for someone, for research." (D9, contributor)

The process starts with requesting and obtaining a copy of their data – an already unfamiliar procedure that most donors did not know was possible. The majority of donors perceived this process as easy and straightforward, *"it was like exactly the step by step that you described"* (D2, collaborator); although some ran into practical difficulties. For instance, forgetting their username or password, or having to wait a long time (i.e., more than six days) for their files. Throughout this process, donors remain largely in the dark regarding their data. Even if they technically 'have' a copy, it is opaque and meaningless as it is either in a large ZIP file containing multiple folders with no structure or guidance (Fig. 7.6), or in a large ZIP file containing files in a format donors are unfamiliar with (Fig. 7.6). Most donors did not open the ZIP files as they were not sure how to open, read or interpret them; those who did found them just as intimidating.

> "I saw in the beginning that the ZIP [file] has so many things. There are so many folders that I was like OK this is too much to look at. So no." (D15, contributor)



Name	Size	Kind
ACTIVE_CAPTAIN	--	Folder
AVC_CORE_DATA_AIRCRAFT_MGMT_DOMAIN	--	Folder
AVC_CORE_DATA_CUSTOMER_DOMAIN	--	Folder
AVIATIONCLOUD	--	Folder
customer_data	--	Folder
DI_BASEBALL_IMPACT	--	Folder
DI_BLUE_LINK	--	Folder
DI_BOOT_AUTOSPORT	--	Folder
DI_COMMUNITY_PLACES	--	Folder
DI_CONNECT	--	Folder
DI_CONNECT_IQ	--	Folder
DI_DC_CLOUD_MEDIA_STORAGE	--	Folder
DI_DELTA_SMART	--	Folder
DI_FACEIT_CLOUD	--	Folder
DI_LIVETRACK_PND_GROUPRIDE	--	Folder
DI_MARINE_QDC	--	Folder
DI_MEDIA_GDPR_SERVICE	--	Folder
DI_PERSISTENT_KEY_VALUE_MAP	--	Folder
DI_TACX	--	Folder
DI_XERO	--	Folder
ENG_A51_CLIPBOARD	--	Folder
ENG_A51_ROOM	--	Folder
FLY_GARMIN	--	Folder
GARMIN_SPORTS_SERVER_API	--	Folder
INREACH	--	Folder
IT_CONSENT_HISTORY	--	Folder
IT_DEVICE_AND_CONTENT	--	Folder

(a) Files returned by Garmin.

Name	Size	Kind
export_cda.xml	367,6 MB	XML File
export.xml	252,8 MB	XML File

(b) Files returned by Apple Health.

Figure 7.6: Files returned by Garmin when requesting the data.

Hence, when donors upload their (sensitive) data into the data donation platform they are doing so blindly. This is awkward and confronting, especially because of the type of information, which they considered "*very personal, sensitive, confidential, internal* information" (D2, D6, D8, D20). Privacy-concerned donors experienced this process as swimming against the tide as they are increasingly aware of the importance of data protection, *"I felt awkward, actually, because I know that my personal data should be pro-*

tected" (D3, co-creator). Here, two main factors helped mitigate their concerns: (1) the research context, and (2) the exploration and curation process built into the platform. The data exploration process helped donors see their data and better understand the types of data being transferred. It also invited them to challenge and have questions about their data; how it was encoded, *"I was really confused that all my rowing activities were generic activities"* (D8, contributor), and the reasons behind specific gaps, patterns, or outliers *"I saw like my training times, mornings and nights. I saw once that I trained like at midday and I was like what happened that day?"* (D5, co-creator). The data curation process prompted donors to pay attention to and care for the transaction being made. In addition, it explicitly invited donors to choose what data to donate and for how long, which fostered a sense of security and trust.

> "At first, I did it very automatically and then I saw that there was like a choice about the data and the amount of time. There I felt that I should pay more attention because suddenly there can be sensitive information. [...] So as I progressed, I became more alert because I was super calm at first. I don't know. It's like a form. When you are filling out a form and they need your name and surname, well you are calm, you give your name and surname, but then, when they have more fields, more information, like your phone number, or email address, then you begin to pay more attention. Like, wait, this is getting a bit long. Are they going to call me? What I am doing? That is how I felt it." (D13, contributor)

Nonetheless, when it came to curating their sensitive data and defining their personal boundaries most donors chose to donate as much data as possible. This decision stems from three motivations. First, full openness and not caring much about privacy considering the various entities that already have access to data about themselves, including Garmin or Apple, *"I was like, well, everything I'm seeing I'm willing to share. I'm always quite open"* (D8, contributor). Second, wanting to contribute to the research as much as possible, *"it was super specific, like until when, for what, which type. So I felt secure, and because of that, I tried to give my best. Like, they're doing a good job. I want to help too"* (D3, co-creator). Third, wanting to see and explore as much data as possible, or to compare it or relate it to a specific event (e.g., injury, competition). Especially, for collaborators and co-creators with the intention to participate in the contextualization interview. In fact, some of them had already a specific goal for this activity, *"I would like to learn how to interpret the data we have available in these phones and watches and decide, well, what interests me and what doesn't."* (D7, co-creator). Through the collaborative interpretation of their data, collaborators and co-creators (re)discovered patterns and events reflected by the data, confirmed their expectations using the retrospective data as a reference point, and challenged their assumptions from seeing the data over time. In addition, they gained more clarity about their data, awareness about their bodies, and new questions for the future.

> "I was very interested in the visualization of my data and identifying these patterns that we just discussed. (...) After the interview, now I have like these open questions. Yeah, I feel motivated also to help research. And I like the fact that right now I'm motivated to be part of it and not like OK, I donated

> my data, and I went to the interview. No, I'm like in a feedback process and I really want to be part of it." (D4, co-creator)

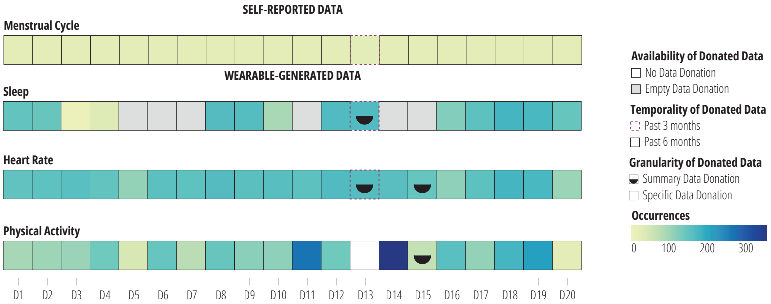


Figure 7.7: Summary of the data donated by each donor and how it is configured through their choices.

7.5.3. Types and Characteristics of the Donated Data

The sensitive data donation process results in different types of data with characteristics that are shaped by the degree of participation. The baseline is the contributor's data (Fig. 7.7). It is spatiotemporal and retrospective, going back up to six months from the date it was requested. It is decontextualized and its content and specificity depend on the choices made by donors. In turn, these are influenced by how donors use their devices, *"I donated everything I could except the sleep [data] because I don't sleep with my device"* (D5, co-creator). In the case of collaborators and co-creators, this baseline is augmented, annotated, situated, and contextualized through three strategies. The first strategy is relating it to other sources of data (e.g., calendar, menstrual tracking app) and to specific events *"I can check the actual date. Yeah, eight and nine. We had a weekend to, like, train for the competition"* (D17, contributor). The second strategy is to reflect on the experience lived around the data, here several factors come into account, including motivation, perceived effort, training preferences, and stress, *"this represents the* [event] *stress"* (D5, co-creator). The third strategy is to underline variability and temporality and describe the reasons behind it.

> "Those are the days I have the most time and I always train for an hour and a half, a little bit more. The other days of the week it is extremely restricted because I have one hour to train in the morning." (D16, co-creator)

Data is challenged and rectified. For instance, by noticing outliers and challenging these values, *"actually, sometimes it is earlier, like 7:00 a.m."* (D2, collaborator), or by differentiating activity types that were recognized poorly by the device or miscategorized, *"when you select like cardio in Garmin, in my case it can be either like cardio in the gym, like bike or a treadmill, or basketball because I also use it in that setting"* (D4, co-creator).

Moreover, through exploration, curation, and interpretation, data also plays different roles for donors. Here, it should be noted that before uploading the data to the platform, donors had never seen an overview of their data over an extended period of time. It is seen as evidence of their (lack of) consistency and compliance with a specific routine or training program, *"seeing the big picture. It's like I should be so consistent and should be attached to the plan"* (D3, co-creator). It is seen as a medal, commemorating the effort, *"this is also kind of proof to myself like you are doing really well."* (D20, collaborator). It provides satisfaction and it is appreciated. Finally, it is seen as an alarm, signaling disruption or change and motivating donors to care differently for their bodies or adapt their routines, *"that is then a wake-up call for myself. Like, oh, this isn't right. I need to change my habits"* (D8, contributor).

7.5.4. Into The Context: Tracking Practices and Envisioned Metrics

Tracking and Interacting with the Data

Donors use smartwatches and wearable devices to plan their day-to-day activities and monitor them as they go (e.g., to make sure they are running at the planned speed). Once an activity is over, they glance at the data to review the summary and their overall progress. For specific activities, such as intense training or competitions, they compare with their past performance – stored in the form of Personal Records (PRs) : *"I don't look back* [at the data]. *Not really. I just look for PRs"* (D3, co-creator). Outside specific activities, they engage with prescribed metrics (e.g., body battery[3], stress) throughout the day to better understand what is happening with their bodies. These metrics are sometimes a poor quantification of their experiences, which becomes a source of doubt and concern.

> "I don't know if it's an error, but sometimes I feel like I slept very well but my *body battery* doesn't recharge that much." (D9, contributor)

Other than these brief interactions, donors rarely look (back) at their data and never see an overview of their past data, *"it's like, oh, what did I do yesterday? What do I have for tomorrow? But I don't check anything else."* (D2, collaborator). For professional and semi-professional athletes, coaches and training staff have direct access to data and interpret them. They are assumed to have the knowledge and tools to correctly do so *"I like that because there are many things that I ignore or do not know how to interpret well"* (D7, co-creator). In these cases, the menstrual cycle is rarely discussed or considered, in part because the coaches of all athletes we interviewed were cisgender men.

Perceived Impact of the Menstrual Cycle in Physical Activity

Donors perceived that *physical*, *mental*, and *other* factors have an impact on sports (Fig 7.8). I categorize them in this way, as it is common for athletes to train and address them through different activities. Yet, these are interrelated. For instance, during the early follicular phase, with the onset of menstruation, water retention can lead to feeling bloated (i.e., *physical factors*). This, together with bleeding can increase tension or anxiety over physical appearance and spotting and decrease training motivation (i.e., *mental*

---

[3] Body Battery is a metric designed by Garmin to "show the effects of physical activity, stress, relaxation, and the restorative power of sleep together in a single place" [72]. It goes from 0 to 100.

factors), which in turn, might shape the choice of activity, clothes, or menstrual products (i.e., *other factors*).

> "I won't be swimming that day, but maybe in two days, you know? So, I will, like, reschedule my week to avoid the strongest day of my period being the day of swimming." (D4, co-creator)

Physical factors impact (1) the range and types of activities that athletes can perform – for instance, sensitive breasts make it difficult to perform activities that require running at high speeds or jumping; (2) how athletes perform and how they perceive their performance – for instance, a workout can be considered conditionally good or productive considering the phase of the cycle; and (3) the type of activity that athletes can participate in – for instance, whether a lightweight boxer will be allowed to compete or must modify her diet before competition to account for weight fluctuations.

> "When I'm training for a fight it is much harder because I need to keep checking my weight every day and then it's frustrating when you go and check and you did a hard training, and you were eating well, and it's like nothing changes, or even it [weight] goes up." (D11, co-creator)

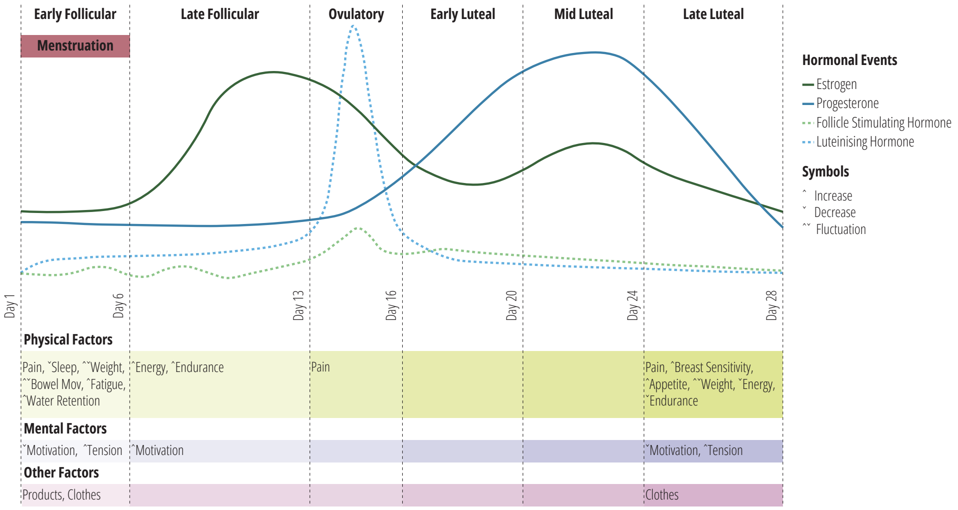


Figure 7.8: Factors that donors perceive have an impact on physical activity, training, and competition; represented on the phases of a eumenorrheic 28-day menstrual cycle. Adapted from Carmichael et al. [39].

Notably, a prominent physical factor is amenorrhea, or the absence of menstruation, especially for athletes competing in lightweight categories (e.g., boxing, rowing, weightlifting) or athletes who have recently lost weight. *Mental factors* impact (1) whether and how athletes approach sports – for instance, if they want to train and where *"it doesn't feel comfortable to go to the gym"* (D2, collaborator); and (2) their feelings before, during, and after sports. Notably, donors emphasized that performing any sport positively impacts how they feel after – independently of any other factor, *"I think it helps me feel* *energized"* (D18, contributor). *Other factors*, such as period products and clothes, affect the type of physical activity, training, and competition that athletes can and are comfortable doing.

> "I will never be able to do an Ironman [long-distance triathlon] with my period, never. Nothing will resist. Not even a tampon. So, when I was in my first Ironman I got an injection to stop it." (D3, co-creator)

Towards Cyclical 'Metrics'

Collaborators and co-creators were involved in crafting metrics that they would like to track as part of their routine. We use the term 'metric' loosely, as these don't aim to measure the body, but represent a subjective experience. These are not intended to (more or less accurately) predict but rather to foster understanding. Donors choose to craft these metrics on the cyclical nature of the menstrual cycle. They envisioned cyclically reflecting on them, through the menstrual cycle and after multiple cycles.

- **Cyclical Motivation:** How motivation to train varies through the cycle. It can be represented through a subjective ordinal scale in which athletes reflect on how much they look forward to a specific activity and the factors influencing this, *"I think that's really an indicator of how you are feeling well because if you are not feeling well, you are not looking forward to it. [...] For me, the low scale is «I'm doubting whether I should train» and then the high scale is «I can't wait to get started»"* (D12, co-creator)
- **Cyclical Activity:** How specific activities (e.g., flexibility, strength) and types of activity (e.g., running, swimming) from the training routine should adapt to account for the menstrual cycle. It can be represented through a qualitative reflection process that accounts for the experiences with different activities throughout the cycle and the factors influencing these; *"maybe it also has something to do with the body. Like when I have my period, running, for me it's very hard like I feel weak and not that motivated, but maybe doing some yoga will help me."* (D3, co-creator).
- **Cyclical Endurance:** How the menstrual cycle impacts endurance (i.e., the capacity to withstand a training of competition) and performance (i.e., how well athletes execute a specific training or competition). It can be represented quantitatively, for instance through a proxy for effort derived from existing data, *"you can see if I made more or less effort comparing if the heart rate was higher or lower, and at the same time it could be correlated with the speed"* (D5, co-creator); and qualitatively, for instance by reflecting on one's feelings: *"the data is not a measure of how good or how fit I'm feeling for exercising [...] but there is a clear difference that not every day I'm feeling the same."* (D4, co-creator).
- **Cyclical Rest and Recovery:** How the menstrual cycle impacts rest and recovery after sports. It goes beyond existing pre-defined metrics such as sleep duration and body battery and aims to represent athletes' bodily knowledge around sleep, rest, recovery, and fatigue. It can be represented through qualitative observations on specific factors, *"I feel that the rest is different [during menstruation], I think that*

my sleep is like heavier" (D2, collaborator) and documenting occurrences, such as pain: *"when the pain is very strong I wake up"* (D5, co-creator).

- **Cyclical Care:** How donors care for their bodies throughout the menstrual cycle, especially during training, to account for it during competitions. For instance, how they nurture their bodies, manage and train with pain, account for breast sensitivity by wearing 'the right' bra, or choose a menstrual product based on the type of activity and its duration. It can be represented through qualitative observations over time. *"Sometimes I was really in pain for my training, and the guys at the gym used to say yeah, have a day off. But if it* [pain] *happened on the day* [of the fight], *I can't have a day off, so I need it just to go anyway."* (D11, co-creator).

7.6. Discussion

7.6.1. Reflecting on Sensitive Data Donation

The Sensitive Data Donation method builds upon principles and practices described in the data donation literature. These include Jones' [100] incongruities around data awareness, underlining how donors *"don't know what they don't know"*, approaches inviting donors to interpret and contextualize data [74, 75], implementations of data donation platforms and systems [13, 25, 162], and legal and ethical considerations [34, 97, 140]. Contrasting with recent approaches, it stems from and advocates for a different way of thinking about data. When applying Sensitive Data Donation the goal is not to achieve ecological validity [162, 204] or generalizability [24, 190] but to foreground the content and characteristics of data and how it relates to individuals' knowledge and experiences in a given (intimate) context. In table 7.1, I summarize how the five principles of Sensitive Data Donation were integrated into the case study.

The method starts with a research question or goal that is scoped in collaboration with potential donors. Here, the initial research(ers) goals are present; yet they expand to incorporate people's interests, preferences, and experiences. This invites reconsidering how the data needs of a project are conceived: from fixed to flexible criteria encouraging choices. Similarly, it invites to understand successful participation in terms of agency and the relevance of the research instead of participation compliance, or the willingness to follow the procedure related to participation (e.g., [18, 24, 142]). For instance, D13 decided not to donate her physical activity data because she considered it sensitive information. This would have been considered a "consent error" [24] in other data donation approaches. For me, it was an opportunity to consider upfront and throughout the process how to harness partial or incomplete datasets.

Likewise, promoting different forms of participation results in diverse and messy datasets, as these are created from boundaries and exploration and account differently for contextual factors around data. I found how not all donors wish to participate in the same way and how expectations vary depending on the degree of participation. Consequently, the resulting data is not always "complete" nor situated and contextualized. This might also be inconvenient, and opposite to the tidy and clean datasets expected in most data science projects [51, 161]. Especially when donors participate as contributors the data

Principle	*Application*	*Example*
Balanced Value	Co-defining value-gain strategy and honoring donors effort and expectations.	D2 expected to learn from the research process and her data. Researchers delivered partial results and built a relationship.
Sensitive Data	Supporting data exploration before transaction and encouraging boundaries and choices.	D13 became aware of data sensitivity by exploring and curating it. Researchers requested and worked with messy, partial or incomplete data.
Multiple Knowledge	Incorporating voluntary activities where data is contextualised, enriched and augmented.	D11 contextualised her weight fluctuations with training and fights. Researchers account for contextual information when outlining metrics.
Ongoing Consent	Delaying initial consent until after exploration and inviting (re)assessment.	D15 blindly uploaded her data to the platform. Researchers promoted various forms of data exploration and invited donors to (re)assess their choices.
Shared Goals	Including donors in scoping research objectives and questions.	D4 is encouraged by the process and wants to stay in the loop. Researchers re-formulated research goals.

Table 7.1: Application of the sSD principles throughout the case study

has gaps and outliers, leaving researchers with questions, assumptions, and possible (mis)interpretations. I resisted the urge to clean up the data and instead chose to question them together with donors where possible and record my questions when not. Similar to D'Ignazio and Klein [51] who argue that the process of cleaning and tidying data *"can be a destructive rather than constructive act"*, I argue that Sensitive Data Donation requires embracing the messiness of data and determining what data is useful and when. For example, the decontextualized data from contributors can be used to assess the cyclical endurance metric, but this metric would not have been developed without situated and contextualized knowledge from contributors and co-creators.

Further, the research process defined by Sensitive Data Donation is intended to be useful beyond the resulting datasets or outcomes of the data analysis; by inviting individuals to engage with and explore their sensitive data. Previous research demonstrates how these activities are perceived as valuable and empowering and support donors to engage with sensitivity [74]. Sensitivity is a property of personal data [76]. Sensitive data belongs to a special category of personal information under the GDPR. It includes racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data; and data concerning a person's sex life or sexual orientation [65, Art. 9]. Outside the GDPR, it includes potentially sensitive information that is unknown in the data [76, 100] and sensitive private information that must be protected and not disclosed (e.g., mental illness [8], HIV status [208]). In this case, sensitivity does not belong to a predefined category but is built through interaction with data. The Sensitive Data Donation method accounts for prescribed and constructed sensitivity. First, through informed consent – a requirement for processing sensitive data under the GDPR. Second, through an incremental process of exploration and understanding of the data.

The results underline the importance of deliberately supporting people in rendering their opaque data visible as a prerequisite for transferring it. Thus, I echo previous research (e.g., [37, 101, 189]) advocating for continuous and dynamic consent processes in data donation and similar practices that entail transactions of personal data, where participants are supported in understanding their data and its implications and continuously invited to (re)assess their participation. I acknowledge that this process might lead to slowing down, and feelings of discomfort or creepiness [177], opposite to what is enabled through other data collection methods that are prioritized by researchers (e.g., crowdsourcing, web scrapping). Additionally, this process could be perceived as burdensome, yet, with the Sensitive Data Donation method, I don't necessarily aim to reduce effort but to increase attention and understanding. I position it as a slow method both for researchers and participants. For researchers, the data is not available in minutes or days but through a lengthy iterative process – on my part, the whole data collection process lasted several months. For participants, the incentive is not to complete a task as quickly as possible to get a (financial) reward instead they are invited to slow down, pay attention, and engage with their data. I join Gould [81] in questioning: Do we need speed? Further, referring back to the principles of Data Feminism [51], I question: Whose priorities are prioritized when prioritizing speed?

7.6.2. Practical Recommendations for Applying Sensitive Data Donation

Based on the challenges and experience applying the Sensitive Data Donation method, I provide the following practical recommendations for researchers aiming to apply it:

- **Become familiar with practicalities:** The donation of data requires individuals to request and receive their data from data controller(s). There are many ways in which this process can be delayed or go wrong. Additionally, the practicalities of this process could change throughout the project. Similar to van Driel et al. [204], I found that the content and file structures that potential donors requested from the same data controller changed over time. I recommend becoming familiar with this process and being alert to potential changes to support donors when necessary and anticipate challenges in the platforms and systems enabling data donation.
- **Balance trade-offs:** Are more data better data? Are more choices better choices? When defining the relevant sources of data there are choices involved entailing

practical trade-offs. In my case, supporting data from various devices (i.e., Garmin and the Apple Health ecosystem) increased the number of people available to donate, but it also increased the technical complexity of the platform and the data donation process. I recommend being aware of these trade-offs and balancing them with the needs and goals of the project.

- **Invite and expect messiness:** Data donation research projects and data needs should be flexible enough to enable various choices from donors, such as: how to participate and what information to disclose. We, as researchers, can gain valuable insights even from messy and incomplete datasets. I recommend anticipating messiness and incompleteness by considering upfront how to promote meaningful choices around information disclosure and how to use the resulting heterogeneous data.

7.7. Chapter Takeaways

In this third and final iteration, I explored the opportunity of contributing, collaborating, and co-creating with (potential) donors. Moreover, I applied the five principles of Sensitive Data Donation, namely: Sensitive Data **(P1)**, Balanced Value **(P2)**, Multiple Knowledge(s) **(P3)**, Ongoing Consent **(P4)**, and Shared Goals **(P5)**. The results of this chapter illustrate how inviting and supporting different degrees of participation is welcomed by donors as they have different expectations and needs. Additionally, they underline a need for flexibility in participatory data donation projects where researchers think differently about the data (needs) and prepare to embrace messy, dynamic, and somewhat incomplete datasets. Overall, promoting and supporting different choices and degrees of participation invited me to question the characteristics of data collection processes and methods prioritized by researchers. Similarly, previous research has challenged these practices through a critical perspective. Gould [81] illustrates several aspects that render data collection a *consumption experience* such as speed (i.e., being able to collect data quickly) and off-the-shelf tools (e.g., standard questionnaire tools). He invites us to reflect on the trade-offs we make around research design decisions. D'Ignazio and Klein [51] underline the power differentials in traditional data collection practices and invite us to challenge them. For example, by recognizing the people and labor involved in data practices. While Loukissas [115] argues for creating (data) interfaces that cause friction. I echo these perspectives and argue for a different mindset in data donation and similar practices around the collection and use of personal data for research. Instead of prioritizing speed, could we prioritize building and maintaining a relationship with interested people and communities? Instead of prioritizing the quantity and "heterogeneity" of data could we prioritize meaningful choices that stem from the supporting friction and interaction between people and their data? In Chapter 8 I further develop these questions, and I discuss the key findings and limitations of the research conducted in this thesis.


3 2021 03 01 00:15,21 2437 01 826.00, ...
4 2021-03 01,00 00,2021,35.50,22.00, ,......
5 2021 03 01 00:15,21 2437 01 826.00, ...
6 2021-03 01,00 00,2021,35.50,22.00, ,......
7 2021 03 01 00:15,21 2437 01 826.00, ...
8 2021-03 01,00 00,2021,35.50,22.00, ,......
9 2021 03 01 00:15,21 2437 01 826.00, ...
10 2021-03 01,00 00,2021,35.50,22.00, ,......
11 2021 03 01 00:15,21 2437 01 826.00, ...
12 2021-03 01,00 00,2021,35.50,22.00, ,......
13 2021 03 01 00:15,21 2437 01 826.00, ...
14 2021-03 01,00 00,2021,35.50,22.00, ,......
15 2021 03 01 00:15,21 2437 01 826.00, ...
16 2021-03 01,00 00,2021,35.50,22.00, ,......
17 2021 03 01 00:15,21 2437 01 826.00, ...
18 2021-03 01,00 00,2021,35.50,22.00, ,......
19 2021 03 01 00:15,21 2437 01 826.00, ...
20 2021-03 01,00 00,2021,35.50,22.00, ,......
21 2021 03 01 00:15,21 2437 01 826.00, ...
22 2021-03 01,00 00,2021,35.50,22.00, ,......
23 2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
Date,Time,avg(temp),avg(humidity),max(pi)
P1),max(tp1),min(tp2),max(tp2),min(tp3),
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
2021 03 01 00:15,21 2437 01 826.00, ...
2021-03 01,00 00,2021,35.50,22.00, ,......
BANK
1234567890123456
123

8

Discussion and Conclusion

8.1. Overview

Data is increasingly part of our daily lives and experiences. It accounts for our relationships with our bodies, as we self-report our bodily experiences through digital apps. It enters our homes and the various spaces we inhabit as we interact with connected products and services inside these spaces, such as voice assistants. It also accounts for our relationships with other people, including our friends, family, and colleagues. Additionally, it can support us in achieving a specific goal, such as running a marathon.

At the same time, data is increasingly valuable for scientific research. Because it is entangled with our lives, data can offer insight into specific aspects, such as our routine, eating habits, or the places we frequent. These can be used to answer research questions in multiple disciplines. Especially since multiple copies of the same data can be used in different projects and multiple copies of multiple types of data can be combined and aggregated. Additionally, because of its temporality, which anchors data to a specific time and date, data allows researchers to conduct research retrospectively and over time.

Thus, researchers from various disciplines, including myself, have been developing methods and tools to gain access to data. One of these, and the focus of this thesis, is **data donation**. Specifically, this thesis aimed to **develop a feminist approach to data donation in intimate contexts that empowers donors to intimately know their data and promotes different forms of knowledge and participation.** I engaged in an iterative and participatory Research through Design (RtD) process centered around the design and development of a data donation process, embodied by a data donation platform. I designed and developed three iterations of the data donation process and consequently, of the platform, and I applied them in three contexts, each focused on a different type of data: (1) menstrual tracking logs from the Clue app, (2) speech records from the Google Assistant, and (3) physical activity logs from Garmin and wearable devices compatible with Apple Health.

In this concluding chapter, I revisit the aim of this thesis and provide a summary of the answers to the research questions and contributions. Next, I discuss the implications for data holders, policy makers, researchers, and people affected by traces of data in modern society, effectively almost everybody. I conclude by reflecting on the limitations of the Sensitive Data Donation Method and discussing some avenues for future work.

8.2. Summary of Answers to Research Questions

RQ1: How is data donation defined, operationalized, and applied in scientific research?
With this question, I aimed to identify the relevant concepts and current practices in the data donation literature across various disciplines, including philosophy, psychology, health, social sciences and communication, and human-computer interaction. From the results in Chapter 2, I identified how data donation is defined around its characteristics, as a voluntary act with an altruistic nature and the purpose of contributing to scientific research, and in terms of its methodological strengths, where data donation is considered a powerful method for data collection as it enables researchers to collect private personal data at the individual level.

Data donation has been operationalized through the development of digital platforms and systems where individuals can donate a copy of their data, such as OSD2F [13] and PORT [25]. It has been applied in contributory research projects, where donors' primary contribution is their data, through two main approaches: (1) digital platforms where donors upload their data and (2) dedicated applications where donors consent to scrape data using their accounts. I synthesized these aspects into a conceptual framework (Fig. 8.1), illustrating the actors involved in data donation, the key concepts, and the relationships among them.

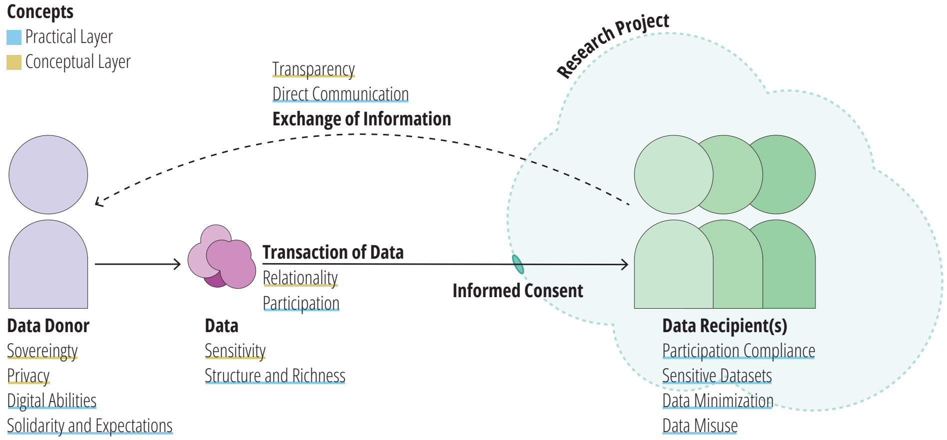


Figure 8.1: Conceptual framework of data donation

Additionally, I used this framework as a base to investigate how data donation *should* be approached in scientific research. I proposed a critique of data donation through the lens of Data Feminism [51]. It highlights power, value-gain, and knowledge differentials in data donation (e.g., between donors and researchers, between donors and their data), and limitations around the types of knowledge (e.g., digital-trace data) and the degrees of participation currently included in data donation research. I consolidate the insights derived from the critique into five principles of Sensitive Data Donation:

(P1) **Balanced Value:** This principle calls for recognizing and honoring donors' contributions and efforts by intentionally integrating activities into the data donation process that allows them to derive value. It invites the questions: **What contributions will donors make to the research project? What would be a fair benefit to donors?**

(P2) **Sensitive Data:** This principle calls for recognizing the sensitive nature of the data as a critical prerequisite for balanced and informed participation. It invites the question: **How can donors know their sensitive data and draw clear boundaries around its disclosure?**

(P3) **Multiple Knowledge(s):** This principle calls for involving donors in interpreting and contextualizing their data; prioritizing their embodied and situated knowledge. It invites the question: **How can donors participate in interpreting, contextualizing, and situating their data?**

(P4) **Ongoing Consent:** This principle calls for embedding informed consent as an ongoing incremental process that accounts for donors' preferences regarding participation and disclosure of sensitive information over time. It invites the question: **How can donors be empowered to increasingly know their sensitive data and identify whether and what information to disclose?**

(P5) **Shared Goals:** This principle calls for supporting different degrees of participation in data donation and inviting interested potential donors to relate to and shape the research project and goals from the start. It invites the question: **How could donors play a more active role in data donation research projects?**

These principles guided the design of the data donation process and were applied in this thesis. They motivate a re-framing of the actors involved in data donation (e.g., from contributing data donors to contributors, collaborators, and co-creators); the key concepts (e.g., from data to sensitive data); and the relationships among them (e.g., from unilateral exchange of information to bilateral exchange of information and balanced value). In Figure 8.2 I synthesize them into a conceptual framework of Sensitive Data Donation.

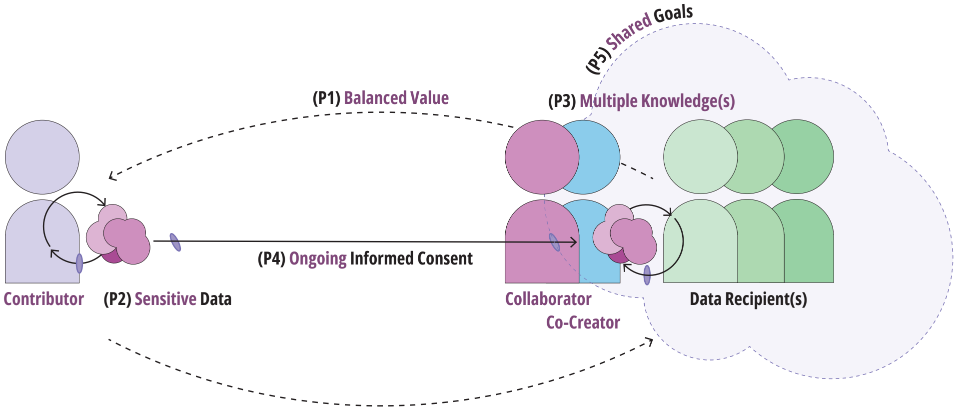


Figure 8.2: Conceptual framework of Sensitive Data Donation

RQ2: How can data donation integrate people's embodied and situated knowledge(s)? This question, aligned with the principle of Multiple Knowledge(s) **(P3)**, was aimed at exploring the possibility of integrating into data donation activities where donors could participate in the interpretation and contextualization of their data. Through these activities, they challenge and enrich their digital-trace data with their embodied and situated knowledge(s). In Chapter 2, I include these activities as part of the Sensitive Data

Donation process under **Phase 5:** *"Contextualize (and Further Identify) Data."* In Chapter 3, drawing from the participatory (data) design literature, I opted for these activities to take the form of semi-structured interpretative interviews prompted by the data. These are commonly used in the literature, for instance, as part of articulation work, proposed by Tolmie and colleagues [195], or the Data-Enabled Design Method [205].

Donors' involvement in these activities also meant their degree of participation shifted: from *contributors* to *collaborators*. In Chapter 3, when I first invited donors to participate in this activity, I was hesitant, as participating in a collaborative project was (more) time-consuming and involved donors' discussing their intimate data and experiences with a researcher. I was surprised when most donors (27/35) expressed interest in participating, illustrating how this approach was not only feasible but even interesting or desirable for them. Beyond feasibility or desirability, this activity allowed me, as researcher, to get closer to and familiarize with the data as I engaged with it and visually represented it in preparation to the session. In turn, this process invited me to come up with my own assumptions, (mis)interpretations and questions that I brought to the interviews.

More importantly, this activity allowed donors to delve into the experiences partially captured by the data, share important insights about their experiences, and even rectify the assumptions and (mis)interpretations I had made. In Chapter 4, I further explored this activity with a focus on the increased understanding of the data promoted by it and the resulting value-gain. From the results, I identified how donors gain a deeper understanding of their data through this process of familiarization and contextualization. Further, this understanding is perceived as valuable and empowering, even if somewhat uncomfortable and confronting.

Scientific research applying data donation can integrate multiple forms of knowledge, specifically people's embodied and situated knowledge, by designing activities that invite and encourage people to play a more active role in the research process centered around the collaborative interpretation of the data. Moreover, the knowledge(s) derived from these activities are not only beneficial to the research(ers), who gain access to new insights and mitigate their assumptions and (mis)interpretations. They can also support donors in knowing their data better. For instance, by becoming aware of what is discoverable (by others) in their data. Which, in turn, can motivate them to (re)define boundaries around the data and information that they disclose.

RQ3: How can data donation foster an incremental understanding of data that invites donors to (re)assess their participation?

With this question, I aimed to explore the nuances around what it means to be *informed* and the roles and responsibilities that we, as researchers, have around informing our research participants. This question originated from the practicalities of obtaining and requesting a copy of digital-trace data from any given data controller, a process enabled by the rights of access and data portability in the GDPR. In practice, although, through this process, we gain access to our data – in the sense that we obtain a digital copy of it – our data is not accessible to us – in the sense that we can hardly navigate or manipulate the returned files.

Previous research, and my own experience requesting multiple copies of my data from various data controllers in Chapter 6, demonstrates how this process results in files that are *"difficult to understand and impractical to use"* [31]. Thus, in most data donation approaches, including that in Chapter 3, when people donate or transfer their data to a researcher or research institution they do so blindly, they don't know what information it contains and consequently it is nearly impossible for them to define meaningful boundaries around its disclosure.

I addressed this question through the iteration described in Chapter 4 and applied in Chapters 4 and 5. I approached the five phases of the Sensitive Data Donation method as opportunities to invite donors to explore and familiarize with their data. First, I created an intermediate step between receiving and transferring a copy of the data in **Phase 3:** *"Upload, Explore and Curate Data"* where potential donors autonomously explored their data through the data donation platform *before*: (1) deciding if they wanted to donate it, and (2) deciding what (types, segments, points) they wanted to donate. Thus, delaying the transaction of data and the initial instance of informed consent until after donors could explore and engage with their data. Donors responded positively to this activity, increasing their trust in the research process and giving them a sense of control over what they contributed.

Second, I designed the activity in **Phase 5:** *"Contextualize (and Further Identify) Data"* to support donors in delving into different aspects of their data, including the different types of information it captures, its temporal distribution, as well as patterns and outliers. This activity resulted in feelings of discomfort that stemmed from sensitive information found in the data that was not apparent on its surface – nor during the initial exploration. Thus, reinforcing the need for dynamic consent practices where donors can translate these feelings into concrete boundaries and decisions regarding their participation. Notably, when inviting donors to (re)assess their participation after this activity, none (0/22) withdrew their participation from the research or their data from the data donation platform.

Scientific research applying data donation can encourage donors to (better) understand their data by designing tools and activities that invite data exploration and interpretation. These should promote different levels of reasoning about and from the data: what information it contains, how it relates to people and their behavior, and what information is discoverable. In turn, the understanding derived from these tools and activities can be translated into meaningful choices through explicit decision-making instances. Although inviting donors in Chapter 3 to (re)assess their participation did not influence their participation or disclosure, having a choice cemented in a true understanding of their data increased their trust in the research and sense of control.

RQ4: How do donors perceive 'sensitivity' and 'intimacy' when faced with a comprehensive view of their data?

With this question, I delve into the principle of Sensitive Data **(P2)** from the perspective of donors. Sensitive data is a well-defined term in the GDPR, as a special category of personal information that is subject to specific processing conditions, for instance, informed consent from the (data) subject. It includes personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, genetic data, biometric data processed solely to identify a human being, health-related data, and data concerning a person's sex life or sexual orientation [65]. People's perceptions of sensitive data partly align with this definition. Yet, they call for a more nuanced space.

In Chapter 5 I specifically explored donors' perceptions of 'sensitivity' and 'intimacy' regarding their speech records from Google Assistant. I accompanied donors on a journey into their data. It started with a focus on single interactions they had with their Google Assistant, such as *"OK Google, set an alarm"* and it concluded with an exploration of multiple interactions when aggregated, combined, and contextualized, such as their most frequent queries and how these distributed over time. Somewhere along this journey, most donors experienced a *tipping point*, where together with feelings of discomfort their perceptions changed: from their speech records being not sensitive to their speech records being very sensitive, creepy, and even intimate.

The results in Chapter 5, illustrate how in the context of digital-trace data 'sensitivity' and 'intimacy' are characteristics beyond prescribed categories. They are not inherent to data (types). They manifest when data intrudes the boundaries of a person's private space, capturing the nuances of what happens within, which were not exposed before, and exposing them to unwanted access. Recognizing that sensitivity is not a static property of data (types) and that data might be contextually sensitive requires approaching all data as potentially sensitive. Further, the *tipping point* underlines how these characteristics are not known a priori or through a superficial engagement with data, but they need to disentangle from it through an active process of exploration. Aligned with the answer to the previous research question, this process is incompatible with static consent practices and reinforces the need for dynamic consent processes.

RQ5: How can data donation promote different degrees of participation in research?
This question, aligned with the principle of Shared Goals **(P5)**, aimed to explore how to develop contributory, collaborative, and co-creative research projects through data donation. It builds upon a Framework for Public Participation in Scientific Research proposed by Shirk et al [175]. I approached this question in terms of feasibility (e.g., To what extent can data donation projects be other than contributory?), and of concrete activities as part of the Sensitive Data Donation approach that invite donors to participate as contributors, collaborators, and co-creators.

Chapters 3, 4 and 7 demonstrated the feasibility of promoting different degrees of participation, namely *contributors*, *collaborators*, and *co-creators*. In these chapters, the activities associated with each role included: (1) donating data (*contributors, collaborators, and co-creators*), (2) actively participating in the interpretation and contextualization of the data (*collaborators and co-creators*), and actively participating in scoping the research focus and goals (*co-creators*) – only explored in Chapter 7. Moreover, as described in Chapter 7, the degree of participation influences the resulting data, with the baseline being the data from *contributors* which is decontextualized.

Further, the results of Chapter 7, illustrate how donor's motivations and expectations change according to their chosen degree of participation. *Contributors*' primary goal is to support the research projects. *Collaborators*, in addition to supporting research, also expect to learn something new by interpreting and contextualizing their data; about the research, their data, or themselves. While, *co-creators*, in addition to the above, expect to engage in a more active role throughout various research activities, including scoping future iterations of the research project and building a community.

This demonstrates that (some) individuals can and want to be more actively involved in research. Donors were highly engaged in contributing to the research project, inviting others to contribute, and continuing the research. Their participation significantly shaped the direction of the research project. Thus, promoting diverse forms of participation is beneficial to data donation projects as it creates bonds of trust between researchers and participants from a place of boundaries: donors decide how to participate and what to donate; resulting in an engaged community (more or less) committed to the (sustainability of the) research project. Scientific research applying data donation can promote different degrees of participation by involving (potential) donors through various research activities and shaping the research accordingly.

8.3. Implications

Implications for Data Holders

This thesis underlines the gap between using, or interacting with, a connected product or service, knowing that said interaction results in the collection and indefinitely storage of personal digital-trace data, and even more knowing what information is in the data. This gap is illustrated by people's reactions to their personal dataslips in Chapter 6 and by the feelings of surprise when facing unintended interactions or disentangling sensitive information from the speech records collected by the Google Assistant in Chapter 5. It originates from a perceived lack of transparency and adequate information from data holders. Thus, when a person agrees to use a digital product or service it is not clear what they are agreeing to or what they can choose to agree to. This, in turn, diminishes trust in data holders. They could (re)build trust and address this gap by providing transparent and digestible information to their users, meaningful choices around the disclosure of personal information – beyond not using a specific product and service – and developing new functionalities that allow and invite their users not only to access their data, but to explore and familiarize with it.

Implications for Policy Makers

Through the empirical studies conducted as part of this thesis, donors and I made use of the right to access and right to data portability in the GDPR [65]. Together, from different parts of the world, we requested copies of our data from more than 15 data controllers. During this process, we experienced some of the inconveniences of exercising these rights. These included: forgotten login credentials when requests are handled through dedicated platforms, unexpected delays due not opening an email in due time, lengthy waiting times, and receiving data in surprising and not-so-machine-readable formats – such as the printed files delivered to my home when designing dataslip (Chapter 6). Above all these, the biggest obstacle remains how data is returned (i.e., unstructured files with no guidance, confusing formats) [4, 31]. Policy makers could develop policies that bridge the access gap: from obtaining a machine-readable copy of our data to being able to understand it and use it. Additionally, building upon the empirical characterization of 'sensitivity' and 'intimacy' developed in Chapter 5, policy makers could require data holders to implement dynamic consent practices that support individuals in defining meaningful and informed boundaries around the information that is collected and stored about them – for instance limiting the permanence of the data after a given time.

Implications for (Design) Researchers

Throughout this thesis there are several implications, considerations, and recommendations aimed at designers and researchers applying data donation and, more broadly, working in the interaction between people and their data. Here I will focus on five, aligned with the principles of Sensitive Data Donation. First, as we strive to include individuals and communities in our research activities, we should also strive to provide them with non-transactional value. Here, financial compensation should not be understood as value but as an incentive for participation. Second, working with sensitive data requires responsible and secure practices and systems. These should prioritize encouraging research participants to define boundaries around information disclosure, even when it might be inconvenient to the research(ers). Third, we should continue to acknowledge how data is shaped by the contexts where it is generated and shaped and strive to involve individuals and communities in its interpretation and contextualization. Fourth, we should further develop dynamic practices around informed consent that account for the opaque and dynamic nature of digital-trace data. We should also collaborate with Human Research Ethics committees so these practices become accepted. Fifth, when possible, we should closely co-create and collaborate our research questions, objectives, and processes with individuals and communities.

Implications for Individuals

In this thesis, I described some of the power and knowledge differentials around personal data collection and use: between individuals and companies, individuals and researchers, and individuals and their data. As individuals, we should acknowledge these differentials. We mostly exchange our data for products and services, which are convenient to us. Yet, our data is a fundamental part of these products and services, which is (disproportionately) convenient to product and service providers. To acknowledge our role in this process – recognizing that there is no data without our experiences and interactions – is the first step towards demanding change – if change is what we want. Our personal data, by definition, is related to us. It contains information about our behavior and experiences, the places we visit, people we interact with. Further, it is shaped by the products and services that we interact with and how we interact with them (e.g., how we configure our preferences or where we decide to use them). There is value for us in engaging with this data. It can help us reflect on our experiences and learn from our past behavior; it can also support us in achieving a specific goal. A key to unlocking this value is for us to get closer to our data, for instance by requesting a copy of it and seeing the information it contains and (potentially) reveals. We can also direct our data towards scientific research, as demonstrated in this thesis. Whatever we decide to do with our data, is up to us. We have a choice and through our choices, we have a role to play in shaping the future and demanding change.

8.4. Limitations

In this section, I acknowledge several limitations of the approach applied in this thesis. First, several factors limit people's willingness and ability to donate personal data. 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. Similarly, a group of donors might likely differ from one recruited through different means (e.g., crowdsourcing) or responding to different incentives (e.g., money); which might bias the research process and outcomes. As a result, the potential donors involved in the empirical studies (Chapters 3, 4, 5, and 7) tended to belong to highly educated strata of society and to have high digital skills. Second, given the time and resource constraints of a PhD project, the data donation processes (Chapters 3, 4, and 7) typically operated on short timescales. Data donation processes with longer timescales of collaboration should be explored to investigate critical aspects such as participation, bilateral communication, and accountability over time. Third, the specific contexts and data (types) included in the empirical studies are limited. Although I strived to explore different types of digital-trace data from diverse sources, several data types and characteristics remain unexplored.

8.5. Future Challenges

In this section, I discuss the future challenges of each phase of the Sensitive Data Donation method.

- **Identify, Prepare and Communicate:** Scoping the research questions, objectives, and data needs with potential donors requires their active involvement. Yet, not all individuals can or wish to participate in this process; which might limit the perspectives included in the early stages of the research – for example, by only including those of highly motivated participants. How can we best involve diverse perspectives and participants? Future research should explore how research projects can include and accommodate diverse voices and experiences early and throughout the process.
- **Request and Receive Data:** Several factors limit people's willingness and ability to donate personal data. 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. Thus, a group of donors might likely differ from one recruited through different means (e.g., crowdsourcing) or responding to different incentives (e.g., money); which might bias the research process and outcomes. How can we recognize these biases? Future research should propose tools for researchers and participants in collaborative projects to reflect on their positionality and biases.

- **Upload, Explore, and Curate Data:** Recent approaches to data donation represent data in tabular form [13, 25] and through interactive data visualizations [74, 75]. In all three projects, I opted for a timeline data visualization by following guidelines proposed by Pins et al. [154]. However, I did not evaluate its effectiveness or accuracy. I demonstrated how donors initially approach data donation blindly, underlining the importance of this activity. What are best practices when supporting autonomous data exploration? Future research should evaluate how to best (re)present data to donors.
- **Transfer Data:** I delayed the initial moment of informed consent until after potential donors were invited to explore their data. However, inviting them to explore and inspect their data does not mean they will. I found that some donors who participated as collaborators and co-creators did not explore their data on the platform as they preferred to do so with me. Does this limit their initial ability to assess and consent to their participation at the time of the transaction? Furthermore, I explicitly asked donors to (re)evaluate their participation, but none withdrew their consent or data. Future research should explore how to implement and disseminate dynamic consent modes.
- **Contextualise, and Further Identify Data:** What happens after data donation? I remained accountable and available to donors, yet, there are critical open challenges around these activities. For instance: What happens to donated data after the end of a project? How can donors guarantee that (all copies) of their data are deleted? Future research should propose considerations on how to close sensitive data donation projects and handle the private sensitive data that results from them.

8.6. Concluding Remarks

In this thesis, I proposed the Sensitive Data Donation around five substantive principles: **(P1)** Balanced Value, **(P2)** Sensitive Data, **(P3)** Multiple Knowledge(s), **(P4)** Ongoing Consent, and **(P5)** Shared Goals. I integrated these principles into a five-phase approach, where specific activities suggested an implementation of one or more principles. For instance, the principle of Multiple Knowledge(s) **(P3)**, which calls for involving donors in interpreting and contextualizing their data was approached through the activities of **Phase 5:** *Contextualize (and Further Identify) Data.* I now see the five principles as interrelated and enabling each other. The principle of Balanced Value **(P1)**, when integrated into the data donation process can derive from activities that support donors in engaging with their data **(P2)(P3)**, or from the findings and outcomes of the research **(P5)**. It can also be independent from the data donation process and be aligned with the needs and goals of (potential) donors **(P5)**. The principle of Sensitive Data **(P2)**, invites important considerations around agency and security. It can also be cultivated and further developed through activities that promote Multiple Knowledge(s) **(P3)** and that invite data exploration and interpretation. In turn, inviting donors to approach their data as sensitive enables and promotes Ongoing Consent **(P4)**. Finally, the principle of Shared Goals **(P5)**, can shape how data, knowledge, and value are conceived in a project **(P1)(P2)(P3)**. Ultimately, these principles invite adopting a different mindset around personal digital-trace data collection for research purposes and the (data) partnerships we have with individuals. For instance, prioritizing personal boundaries over the availability of data or the distribution of value over speed.

Epilogue: Unforeseen Challenges and Lessons Learned

In this thesis, I followed an iterative Research through Design (RtD) process where I designed, developed, and implemented three iterations of a Sensitive Data Donation approach – embodied by a digital data donation platform – and applied these in a case study. The first iteration, described in Chapter 3, focused on value creation and supporting data curation and interpretation of menstrual tracking logs from the Clue app. The second iteration, described in Chapters 4 and 5, focused on promoting incremental data exploration and understanding, and engaging with the sensitive and intimate elements and characteristics of Google Assistant speech records. The third iteration, described in Chapter 7, focused on exploring different forms of participation through physical activity data from Garmin and Apple Health devices. Aligned with the RtD approach, the process of 'designing' or 'making' a Sensitive Data Donation process and a digital platform for individuals to donate their data came with practical limitations and choices. It also involved some unforeseen challenges. These are often hidden between a sentence or two in the chapters and the papers they are based on. Hoping that my experience can help future (data donation) researchers, I describe some of the challenges I faced throughout this process and the consequent practical considerations in the form of reflective questions.

Ethical Approval: Dealing with Unusual Requests

A significant challenge in all three iterations, but especially in the first one, described in Chapter 3, was getting the research reviewed and approved by the Human Research Ethics Committee and Privacy Team at the TU Delft. 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 journey 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.

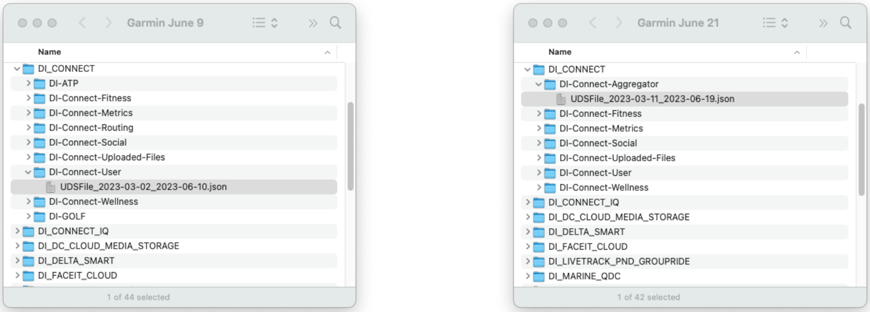
Parsing the Data: Changing Structures

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 iteration). 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 of many data-driven products and services. Thus, in the second iteration, described in Chapter 4, 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 (Fig. 2). 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 iteration, described in Chapter 7, 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 (Fig. 3). 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 (Fig. 1), van Driel and colleagues [204] 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 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 timely respond to unexpected changes.

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 iteration, described in Chapters 4 and 5, some donors obtained empty files or files with



- DI_CONNECT
  - DI-ATP
  - DI-Connect-Fitness
  - DI-Connect-Metrics
  - DI-Connect-Routing
  - DI-Connect-Social
  - DI-Connect-Uploaded-Files
  - DI-Connect-User
    - UDSFile_2023-03-02_2023-06-10.json
  - DI-Connect-Wellness
  - DI-GOLF
- DI_CONNECT_IQ
- DI_DC_CLOUD_MEDIA_STORAGE
- DI_DELTA_SMART
- DI_FACEIT_CLOUD

- DI_CONNECT
  - DI-Connect-Aggregator
    - UDSFile_2023-03-11_2023-06-19.json
  - DI-Connect-Fitness
  - DI-Connect-Metrics
  - DI-Connect-Social
  - DI-Connect-Uploaded-Files
  - DI-Connect-User
  - DI-Connect-Wellness
- DI_CONNECT_IQ
- DI_DC_CLOUD_MEDIA_STORAGE
- DI_DELTA_SMART
- DI_FACEIT_CLOUD
- DI_LIVETRACK_PND_GROUPRIDE
- DI_MARINE_QDC

(a) Garmin Takeout on June 9, 2023

(b) Garmin Takeout on June 21, 2023

Figure 1: Files obtained from Garmin by me while receiving donations for the case study described in Chapter 7. The structure of the folders in the file changes, changing the location of the JSON file titled USDFile, containing the heart rate data.

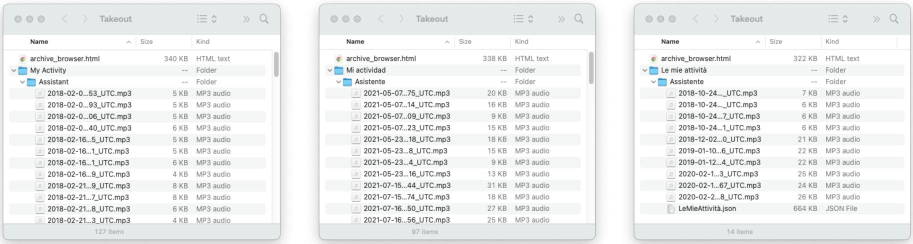
less data than expected due to the introduction of the GDPR in 2018. If they did not opt-in to voice data collection after Google's policy change in 2018, their data only contained speech records before that date, when some donors weren't even Google Assistant users. When the files were empty they could not upload them to the platform, which resulted in them expressing frustration with the whole process and giving up. While in the third iteration, described in Chapter 7, when the data was ready for download Garmin sent an email with a link that expired in three days. Some donors forgot to download their data within that time frame while others only opened the email after three 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 three 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 days[1]. Therefore, it is critical to remain accessible and available to support potential donors in navigating these procedures. **Considerations:** 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.

Exploring and Curating the Data: Sensitive Datasets and Best Practices

In the first and second iterations, the Takeout processes enabled donors to retrieve only their menstrual cycle logs (Chapter 3) and speech records (Chapters 4 and 5) respec-

---

[1] 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."*. Once people complete the Garmin Takeout process they see the following message: *"It takes approximately 48 hours 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."*



Name	Size	Kind
archive_browser.html	340 KB	HTML text
My Activity	--	Folder
Assistant	--	Folder
2018-02-0...53_UTC.mp3	5 KB	MP3 audio
2018-02-0...93_UTC.mp3	5 KB	MP3 audio
2018-02-0...06_UTC.mp3	5 KB	MP3 audio
2018-02-0...40_UTC.mp3	6 KB	MP3 audio
2018-02-16...5_UTC.mp3	5 KB	MP3 audio
2018-02-16...1_UTC.mp3	5 KB	MP3 audio
2018-02-16...1_UTC.mp3	6 KB	MP3 audio
2018-02-16...9_UTC.mp3	4 KB	MP3 audio
2018-02-21...9_UTC.mp3	8 KB	MP3 audio
2018-02-21...7_UTC.mp3	8 KB	MP3 audio
2018-02-21...8_UTC.mp3	6 KB	MP3 audio
2018-02-21...3_UTC.mp3	4 KB	MP3 audio

Name	Size	Kind
archive_browser.html	338 KB	HTML text
Mi actividad	--	Folder
Asistente	--	Folder
2021-05-07...75_UTC.mp3	20 KB	MP3 audio
2021-05-07...14_UTC.mp3	16 KB	MP3 audio
2021-05-07...09_UTC.mp3	9 KB	MP3 audio
2021-05-07...23_UTC.mp3	15 KB	MP3 audio
2021-05-23...18_UTC.mp3	18 KB	MP3 audio
2021-05-23...8_UTC.mp3	15 KB	MP3 audio
2021-05-23...4_UTC.mp3	9 KB	MP3 audio
2021-05-23...16_UTC.mp3	13 KB	MP3 audio
2021-07-15...44_UTC.mp3	31 KB	MP3 audio
2021-07-15...74_UTC.mp3	18 KB	MP3 audio
2021-07-16...50_UTC.mp3	27 KB	MP3 audio
2021-07-16...56_UTC.mp3	25 KB	MP3 audio

Name	Size	Kind
archive_browser.html	322 KB	HTML text
Le mie attività	--	Folder
Assistente	--	Folder
2018-10-24..._UTC.mp3	7 KB	MP3 audio
2018-10-24..._UTC.mp3	6 KB	MP3 audio
2018-10-24...7_UTC.mp3	6 KB	MP3 audio
2018-10-24...1_UTC.mp3	6 KB	MP3 audio
2018-12-02...0_UTC.mp3	21 KB	MP3 audio
2019-01-10...6_UTC.mp3	22 KB	MP3 audio
2019-01-12...4_UTC.mp3	22 KB	MP3 audio
2020-02-1...3_UTC.mp3	25 KB	MP3 audio
2020-02-1...67_UTC.mp3	24 KB	MP3 audio
2020-02-2...8_UTC.mp3	26 KB	MP3 audio
LeMieAttività.json	664 KB	JSON File

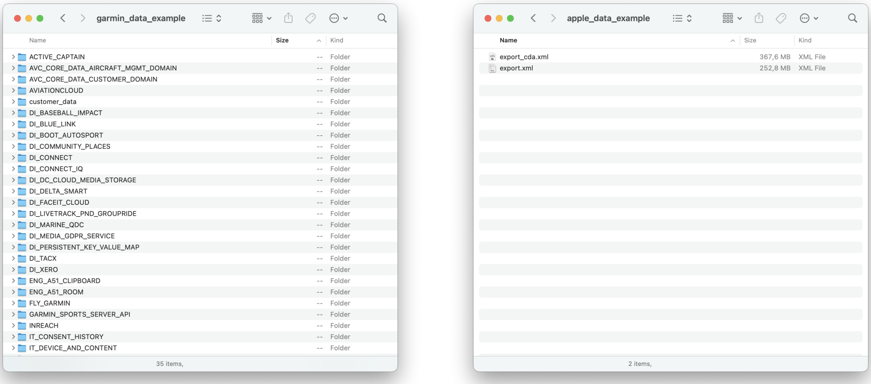
(a) Folders in English **(b)** Folders in Spanish **(c)** Folders in Italian

Figure 2: Files obtained from Google Takeout by potential donors in Chapter 4. The name of the folders in the file changes depending on the language in which the Google account was set up. Shared with permission from the donors.

tively. In the third iteration, described in Chapter 7, donors, specifically Garmin users (Fig. 3.a), obtained a large folder including different types of 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 process in all three iterations 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 Chapter 3; (2) What data points? With the majority of the donors donating all their data in Chapter 4; (3) What types and points of data? For how long? With the majority of the donors donating all their data within the last six months in Chapter 7. 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. Respect these boundaries.

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



Name	Size	Kind
ACTIVE_CAPTAIN	--	Folder
AVC_CORE_DATA_AIRCRAFT_MGMT_DOMAIN	--	Folder
AVC_CORE_DATA_CUSTOMER_DOMAIN	--	Folder
AVIATIONCLOUD	--	Folder
customer_data	--	Folder
DI_BASEBALL_IMPACT	--	Folder
DI_BLUE_LINK	--	Folder
DI_BOOT_AUTOSPORT	--	Folder
DI_COMMUNITY_PLACES	--	Folder
DI_CONNECT	--	Folder
DI_CONNECT_IQ	--	Folder
DI_DC_CLOUD_MEDIA_STORAGE	--	Folder
DI_DELTA_SMART	--	Folder
DI_FACEIT_CLOUD	--	Folder
DI_LIVETRACK_PND_GROUPRIDE	--	Folder
DI_MARINE_QDC	--	Folder
DI_MEDIA_GDPR_SERVICE	--	Folder
DI_PERSISTENT_KEY_VALUE_MAP	--	Folder
DI_TACX	--	Folder
DI_XERO	--	Folder
ENG_A51_CLIPBOARD	--	Folder
ENG_A51_ROOM	--	Folder
FLY_GARMIN	--	Folder
GARMIN_SPORTS_SERVER_API	--	Folder
INREACH	--	Folder
IT_CONSENT_HISTORY	--	Folder
IT_DEVICE_AND_CONTENT	--	Folder

(a) Garmin Takeout Overview

Name	Size	Kind
export_cda.xml	367,6 MB	XML File
export.xml	252,8 MB	XML File

(b) Apple Takeout Overview

Figure 3: Files obtained from Garmin and Apple Health by me while preparing the data donation platform to receive donations for Chapter 7. The structure and format are different.

these sessions, I relied on highly personal interactive data visualizations that I crafted for each individual donor. One important limitation is that crafting these visualizations is time-consuming and scales poorly. In the first iteration, described in Chapter 3, I was not able to carry out all the sessions because I did not have enough time.

Additionally, in all three iterations, 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 the second iteration, described in Chapters 4 and 5, 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 the third iteration, described in Chapter 7, 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. **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.

Practical Recommendations for Applying Sensitive Data Donation: A Checklist

I summarize the challenges and considerations described above into a checklist of practical aspects to consider when applying the five phases of Sensitive Data Donation.

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 to 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. 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. **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. **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. **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. **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.

Data and Code

Research data and code supporting the development of the data donation platform and the findings described in each Chapter of this thesis are available in **Zenodo** and **4TU.Centre for Research Data** at:

- **Data Donation Platform:**

  `https://doi.org/10.5281/zenodo.10908033`

- **Chapter 2:**

  `https://doi.org/10.4121/19976cc9-245f-46ce-b6a5-7836559adf36`

- **Chapter 3:**

  `https://doi.org/10.4121/e971ae20-c4a1-4ea8-87eb-987f2362bda4`

- **Chapter 4:**

  `https://doi.org/10.4121/520e9c5a-d137-4570-8e3f-d890d4b33521`

- **Chapter 5:**

  `https://doi.org/10.4121/0ebb1579-88ab-4bec-a866-1e9aff19581f`

- **Chapter 6:**

  `https://doi.org/10.4121/35a8648c-bca2-4566-a664-b415a12e176a`

- **Chapter 7:**

  `https://doi.org/10.4121/d78aea69-c1f9-4768-9eca-4af50f7f47ad`

Bibliography

[1] N. Abdi, X. Zhan, K. M. Ramokapane, and J. Such. “Privacy Norms for Smart Home Personal Assistants”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan: Association for Computing Machinery, 2021, pp. 1–14. ISBN: 9781450380966. DOI: 10.1145/3411764.3445122.

[2] Algorithm Watch. *DataSkop: Investigating YouTube's algorithm during Germany's election campaign*. 2021.

[3] AlgorithmWatch. *What does TikTok know about you? Data donations deliver answers!* 2023.

[4] F. Alizadeh, T. Jakobi, J. Boldt, and G. Stevens. “GDPR-Reality Check on the Right to Access Data”. In: *Proceedings of Mensch und Computer 2019*. New York, NY, USA: ACM, Sept. 2019, pp. 811–814. ISBN: 9781450371988. DOI: 10.1145/3340764.3344913.

[5] T. Almeida, R. Comber, and M. Balaam. “HCI and intimate care as an agenda for change in women's health”. In: *Conference on Human Factors in Computing Systems - Proceedings*. San Jose, CA, USA: Association for Computing Machinery (ACM), 2016, pp. 2599–2611. ISBN: 9781450333627. DOI: 10.1145/2858036.2858187.

[6] T. Almeida, R. Comber, G. Wood, D. Saraf, and M. Balaam. “On looking at the vagina through Labella”. In: *Conference on Human Factors in Computing Systems - Proceedings*. San Jose, CA, USA: Association for Computing Machinery (ACM), 2016, pp. 1810–1821. ISBN: 9781450333627. DOI: 10.1145/2858036.2858119.

[7] I. Altman. “The Environment and Social Behaviour: Privacy, Personal Space, Territory and Crowding”. In: ERIC, 1975.

[8] N. Andalibi, P. Ozturk, and A. Forte. “Sensitive self-disclosures, responses, and social support on instagram: The case of #depression”. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 1485–1500. ISBN: 9781450343350. DOI: 10.1145/2998181.2998243.

[9] A. J. Andreotta, N. Kirkham, and M. Rizzi. “AI, big data, and the future of consent”. In: *AI & SOCIETY* 37.2015 (Aug. 2021), pp. 1715–1728. ISSN: 0951-5666. DOI: 10.1007/s00146-021-01262-5.

[10] Apple App Store. *Search Result for 'Period Tracker' on Apple App Store*. 2021.

[11] N. Apthorpe, Y. Shvartzshnaider, A. Mathur, D. Reisman, and N. Feamster. “Discovering smart home internet of things privacy norms using contextual integrity”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.2 (2018), p. 23. ISSN: 2474-9567. DOI: 10.1145/3214262.

[12] N. Apthorpe, S. Varghese, and N. Feamster. “Evaluating the contextual integrity of privacy regulation: Parents' IoT toy privacy norms versus COPPA”. In: *Proceedings of the 28th USENIX Conference on Security Symposium*. Santa Clara, CA: USENIX Association, 2019, pp. 123–140. ISBN: 9781939133069.

[13] T. Araujo, J. Ausloos, W. van Atteveldt, F. Loecherbach, J. Moeller, J. Ohme, D. Trilling, B. van de Velde, C. de Vreese, and K. Wellbers. “OSD2F: An Open-Source Data Donation Framework”. In: *Computational Communication Research* 4.2 (2022). DOI: `10.5117/CCR2022.2.001.ARAU`.

[14] M. Balaam, R. Comber, E. Jenkins, S. Sutton, and A. Garbett. “Feedfinder: A location-mapping mobile application for breastfeeding women”. In: *Conference on Human Factors in Computing Systems - Proceedings* 2015-April (2015), pp. 1709–1718. DOI: `10.1145/2702123.2702328`.

[15] S. Bardzell. “Feminist HCI: Taking stock and outlining an agenda for design”. In: *Conference on Human Factors in Computing Systems - Proceedings* 2 (2010), pp. 1301–1310. DOI: `10.1145/1753326.1753521`.

[16] S. Bardzell and J. Bardzell. “Towards a feminist HCI methodology: Social science, feminism, and HCI”. In: *Conference on Human Factors in Computing Systems - Proceedings.* 2011, pp. 675–684. ISBN: 9781450302289. DOI: `10.1145/1978942.1979041`.

[17] S. Bardzell, J. Bardzell, J. Forlizzi, J. Zimmerman, and J. Antanitis. “Critical design and critical theory: The challenge of designing for provocation”. In: *Proceedings of the Designing Interactive Systems Conference, DIS ’12* (2012), pp. 288–297. DOI: `10.1145/2317956.2318001`.

[18] S. E. Baumgartner, S. R. Sumter, V. Petkevič, and W. Wiradhany. “A Novel iOS Data Donation Approach: Automatic Processing, Compliance, and Reactivity in a Longitudinal Study”. In: *Social Science Computer Review* 0.0 (2022), pp. 1–17. ISSN: 15528286. DOI: `10.1177/08944393211071068`.

[19] G. Benabdallah, M. A. Kaneko, and A. Desjardins. “A Notebook of Data Imaginaries”. In: (2023), pp. 431–445. DOI: `10.1145/3563657.3596025`.

[20] L. Bennett Moses. “Is your algorithm dangerous? [Leading edge]”. In: *IEEE Technology and Society Magazine* 37.3 (2018), pp. 20–21. ISSN: 1937416X. DOI: `10.1109/MTS.2018.2857618`.

[21] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge. “Understanding the Long-Term Use of Smart Speaker Assistants”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (Sept. 2018), pp. 1–24. ISSN: 2474-9567. DOI: `10.1145/3264901`.

[22] M. Bietz, K. Patrick, and C. Bloss. “Data Donation as a Model for Citizen Science Health Research”. In: *Citizen Science: Theory and Practice* 4.1 (Mar. 2019), pp. 1–11. ISSN: 2057-4991. DOI: `10.5334/cstp.178`.

[23] M. Blythe and E. Encinas. “The Co-ordinates of Design Fiction: Extrapolation, Irony, Ambiguity and Magic”. In: *Proceedings of the 2016 ACM International Conference on Supporting Group Work (GROUP ’16).* 2016.

[24] L. Boeschoten, J. Ausloos, J. E. Möller, T. Araujo, and D. L. Oberski. “A framework for privacy preserving digital trace data collection through data donation”. In: *Computational Communication Research* 4.2 (2022), pp. 388–423. DOI: `10.5117/ccr2022.2.002.boes`.

[25] L. Boeschoten, A. Mendrik, E. van der Veen, J. Vloothuis, H. Hu, R. Voorvaart, and D. L. Oberski. “Privacy-preserving local analysis of digital trace data: A proof-of-concept”. In: *Patterns* 3.3 (2022). ISSN: 26663899. DOI: `10.1016/j.patter.2022.100444`.

[26] S. Bogers, J. Frens, J. van Kollenburg, E. Deckers, and C. Hummels. “Connected Baby Bottle”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems.* New

York, NY, USA: ACM, June 2016, pp. 301–311. ISBN: 9781450340311. DOI: 10.1145/2901790.2901855.

[27] K. Bongard-Blanchy, A. Rossi, S. Rivas, S. Doublet, V. Koenig, and G. Lenzini. “I am Definitely Manipulated, Even When I am Aware of it. It s Ridiculous! – Dark Patterns from the End-User Perspective”. In: *Designing Interactive Systems Conference 2021*. Apr. 2021, pp. 763–776. DOI: 10.1145/3461778.3462086.

[28] J. Bourgeois, J. Van Der Linden, G. Kortuem, B. A. Price, and C. Rimmer. “Using participatory data analysis to understand social constraints and opportunities of electricity demand-shifting”. In: *ICT for Sustainability 2014, ICT4S 2014*. Stockholm, Sweden: Atlantis Press, 2014, pp. 392–401. ISBN: 9789462520226. DOI: 10.2991/ict4s-14.2014.49.

[29] J. Bourgeois, S. Foell, G. Kortuem, B. A. Price, J. Van Der Linden, E. Y. Elbanhawy, and C. Rimmer. “Harvesting green miles from my roof: An investigation into self-sufficient mobility with electric vehicles”. In: *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan: Association for Computing Machinery (ACM), 2015, pp. 1065–1076. ISBN: 9781450335744. DOI: 10.1145/2750858.2807546.

[30] J. Bourgeois, G. Kortuem, and F. Kawsar. “Trusted and GDPR-compliant research with the internet of things”. In: *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–8. ISBN: 9781450365642. DOI: 10.1145/3277593.3277604.

[31] A. Bowyer, J. Holt, J. Go Jefferies, R. Wilson, D. Kirk, and J. David Smeddinck. “Human-GDPR Interaction: Practical Experiences of Accessing Personal Data”. In: *CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2022, pp. 1–19. ISBN: 9781450391573. DOI: 10.1145/3491102.3501947.

[32] V. Braun and V. Clarke. “Using thematic analysis in psychology”. In: *Qualitative Research in Psychology* 3.2 (Jan. 2006), pp. 77–101. ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa.

[33] V. Braun and V. Clarke. *Successful Qualitative Research. A Practical Guide for Beginners*. London: SAGE Publications Ltd, 2013, p. 400. ISBN: 9781847875822.

[34] J. Breuer, Z. Kmetty, M. Haim, and S. Stier. “User-centric approaches for collecting Facebook data in the ‘post-API age’: experiences from two studies and recommendations for future research”. In: *Information, Communication & Society* (July 2022), pp. 1–20. ISSN: 1369-118X. DOI: 10.1080/1369118X.2022.2097015.

[35] A. Broder. “A Taxonomy of Web Search”. In: *SIGIR FORUM* 36.2 (2002), pp. 3–10. DOI: 10.1145/792550.792552.

[36] B. Brown, A. Weilenmann, D. McMillan, and A. Lampinen. “Five Provocations for Ethical HCI Research”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, May 2016, pp. 852–863. ISBN: 9781450333627. DOI: 10.1145/2858036.2858313.

[37] B. A. Brown, A. J. Sellen, and K. P. O’Hara. “A diary study of information capture in working life”. In: *Conference on Human Factors in Computing Systems - Proceedings* 2.1 (2000), pp. 438–445. DOI: 10.1145/332040.332472.

[38] N. Campo Woytuk, M. L. J. Søndergaard, M. Ciolfi Felice, and M. Balaam. “Touching and Being in Touch with the Menstruating Body”. In: *Conference on Human Factors in Computing Systems - Proceedings*. Honolulu, HI, USA: Association for Computing Machinery (ACM), 2020, pp. 1–14. ISBN: 9781450367080. DOI: 10.1145/3313831.3376471.

[39] M. A. Carmichael, R. L. Thomson, L. J. Moran, and T. P. Wycherley. “The impact of menstrual cycle phase on athletes’ performance: a narrative review”. In: *International Journal of Environmental Research and Public Health* 18.4 (2021), pp. 1–24. ISSN: 16604601. DOI: 10.3390/ijerph18041667.

[40] G. Chalhoub, M. J. Kraemer, N. Nthala, and I. Flechais. ““It did not give me an option to decline”: A Longitudinal Analysis of the User Experience of Security and Privacy in Smart Home Products”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, May 2021, pp. 1–16. ISBN: 9781450380966. DOI: 10.1145/3411764.3445691.

[41] P. Cheng and U. Roedig. “Personal Voice Assistant Security and Privacy - A Survey”. In: *Proceedings of the IEEE* 110.4 (2022), pp. 476–507. ISSN: 15582256. DOI: 10.1109/JPROC.2022.3153167.

[42] S. Chopra, R. Zehrung, and T. A. Shanmugam. “Living with uncertainty and stigma: Self-experimentation and support-seeking around polycystic ovary syndrome”. In: *Conference on Human Factors in Computing Systems - Proceedings*. Yokohama, Japan: Association for Computing Machinery (ACM), 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445706.

[43] T. Clegg, D. M. Greene, N. Beard, and J. Brunson. “Data Everyday: Data Literacy Practices in a Division I College Sports Context”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376153.

[44] Clue by Biowink. *Clue: The app to track your period, and so much more*. 2021.

[45] J. T. Costello, F. Bieuzen, and C. M. Bleakley. “Where are all the female participants in Sport and Exercise Medicine research?” In: *European Journal of Sport Science* 4.8 (2014), pp. 847–851.

[46] P. Coulton and J. Lindley. “Vapourworlds and Design Fiction: The Role of Intentionality”. In: *Design Journal* 20.sup1 (2017), S4632–S4642. ISSN: 14606925. DOI: 10.1080/14606925.2017.1352960.

[47] E. S. Cowley, A. A. Olenick, K. L. McNulty, and E. Z. Ross. “"Invisible Sportswomen": The Sex Data Gap in Sport and Exercise Science Research”. In: *Women in Sport and Physical Activity Journal* 29.2 (2019), pp. 146–151.

[48] A. Crabtree and R. Mortier. “Human Data Interaction: Historical Lessons from Social Studies and CSCW”. In: *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work*. Ed. by N. Boulus-Rodje, G. Ellignsen, T. Bratteteig, M. Aanestad, and P. Bjorn. September 2015. Oslo, Norway: Springer International Publishing, 2015, pp. 19–23. ISBN: 9783319204994. DOI: 10.1007/978-3-319-20499-4.

[49] A. Crabtree, P. Tolmie, and W. Knight. “Repacking ‘Privacy’ for a Networked World”. In: *Computer Supported Cooperative Work: CSCW: An International Journal* 26.4-6 (2017), pp. 453–488. ISSN: 15737551. DOI: 10.1007/s10606-017-9276-y.

[50] C. Criado Perez. *Invisible Women: Data Bias in a World Designed for Men*. Harry N. Abrams, 2019.

[51] C. D’Ignazio and L. F. Klein. *Data Feminism*. MIT Press, 2020, p. 328. ISBN: 9780262044004. DOI: https://doi.org/10.7551/mitpress/11805.001.0001.

[52] C. D’Ignazio, R. Michelson, A. Hope, J. Hoy, J. Roberts, and K. Krontiris. “"The Personal is Political": Hackathons as Feminist Consciousness Raising”. In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW2 (2020). ISSN: 25730142. DOI: 10.1145/3415221.

[53] P. D'Olivo, M. C. Rozendaal, and E. Giaccardi. "Making Tactful Objects for Sensitive Settings: A Research through Design Process". In: *ACM International Conference Proceeding Series* (2020). DOI: 10.1145/3419249.3420140.

[54] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz. "We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy". In: February (2019). DOI: 10.14722/ndss.2019.23378.

[55] A. Desjardins, J. McWhirter, J. Petelka, C. Simon, Y. Shin, R. K. Peven, and P. Widjaja. "On the Making of Alternative Data Encounters: The Odd Interpreters". In: *Conference on Human Factors in Computing Systems - Proceedings* (2023). DOI: 10.1145/3544548.3581323.

[56] A. Desjardins and H. R. Biggs. "Data Epics". In: *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. 2021, pp. 1–17. ISBN: 9781450380966. DOI: 10.1145/3411764.3445241.

[57] D. Diethei and J. Niess. "Sharing heartbeats: Motivations of citizen scientists in times of crises". In: *Conference on Human Factors in Computing Systems - Proceedings* 2020.April (2021), p. 15. DOI: 10.1145/3411764.3445665.

[58] P. Dourish. "What we talk about when we talk about context". In: *Personal and Ubiquitous Computing* 8.1 (2004), pp. 19–30. ISSN: 16174909. DOI: 10.1007/s00779-003-0253-8.

[59] A. Duhaime-Ross. *Apple promised an expansive health app, so why can't I track menstruation?* 2014.

[60] A. Dunne and F. Raby. *Speculative Everything: Design, Fiction and Social Dreaming*. Ed. by MIT Press. 2013.

[61] K. Ehrnberger, M. Räsänen, and S. Ilstedt. "Visualising gender norms in design: Meet the Mega Hurricane Mixer and the drill Dolphia". In: *International Journal of Design* 6.3 (2012), pp. 85–94. ISSN: 1994036X.

[62] S. Emmonds, O. Heyward, and B. Jones. "The Challenge of Applying and Undertaking Research in Female Sport". In: *Sports Medicine - Open* 5.1 (2019). ISSN: 21989761. DOI: 10.1186/s40798-019-0224-x.

[63] D. A. Epstein, N. B. Lee, J. H. Kang, E. Agapie, J. Schroeder, L. R. Pina, J. Fogarty, J. A. Kientz, and S. A. Munson. "Examining menstrual tracking to inform the design of personal informatics tools". In: *Conference on Human Factors in Computing Systems - Proceedings*. Vol. 2017-May. Denver, Colorado, USA: Association for Computing Machinery, May 2017, pp. 6876–6888. ISBN: 9781450346559. DOI: 10.1145/3025453.3025635.

[64] J. Eschler, A. Menking, S. Fox, and U. Backonja. "Defining Menstrual Literacy with the Aim of Evaluating Mobile Menstrual Tracking Applications". In: *CIN - Computers Informatics Nursing* 37.12 (2019), pp. 638–646. ISSN: 15389774. DOI: 10.1097/CIN.0000000000000559.

[65] European Parliament. *General Data Protection Regulation (GDPR)*. 2018.

[66] FITR Woman. *FITR Woman: Track Your Period & Train Smarter*. 2023.

[67] Flo Health. *About Flo*. 2021.

[68] L. Floridi. "Open Data, Data Protection, and Group Privacy". In: *Philosophy and Technology* 27.1 (2014), pp. 1–3. ISSN: 22105441. DOI: 10.1007/s13347-014-0157-8.

[69] S. Fox, N. Howell, R. Wong, and F. Spektor. "Vivewell: Speculating near-future menstrual tracking through current data practices". In: *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*. San Diego, CA, USA: Association for Computing Machinery (ACM), 2019, pp. 541–552. ISBN: 9781450358507. DOI: 10.1145/3322276.3323695.

[70] B. Friedman and P. H. Kahn. “Human Values, Ethics, and Design”. In: *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications.* L. Erlbaum Associates Inc., 2002, pp. 1177–1201. ISBN: 0805838384.

[71] S. Gabriele and S. Chiasson. “Understanding Fitness Tracker Users’ Security and Privacy Knowledge, Attitudes and Behaviours”. In: *Conference on Human Factors in Computing Systems - Proceedings.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12. ISBN: 9781450367080. DOI: `10.1145/3313831.3376651`.

[72] Garmin. *Body Battery™ Energy Monitoring.* 2021.

[73] E. Giaccardi and P. J. Stappers. *Research through Design.* 2017.

[74] A. Gomez Ortega, J. Bourgeois, W. T. Hutiri, and G. Kortuem. “Beyond data transactions: a framework for meaningfully informed data donation”. In: *AI and Society* 0123456789 (2023). ISSN: 14355655. DOI: `10.1007/s00146-023-01755-5`.

[75] A. Gómez Ortega, J. Bourgeois, and G. Kortuem. “Reconstructing Intimate Contexts through Data Donation: A Case Study in Menstrual Tracking Technologies”. In: *Nordic Human-Computer Interaction Conference.* New York, NY, USA: ACM, Oct. 2022, pp. 1–12. ISBN: 9781450396998. DOI: `10.1145/3546155.3546646`.

[76] A. Gómez Ortega, J. Bourgeois, and G. Kortuem. “What is Sensitive About (Sensitive) Data ? Characterizing Sensitivity and Intimacy with Google Assistant Users”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23).* Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: `10.1145/3544548.3581164`.

[77] E. Goodman. “Design and ethics in the era of big data”. In: *Interactions* 21.3 (2014), pp. 22–24. ISSN: 15583449. DOI: `10.1145/2598902`.

[78] Google Play. *Search Result for ’Period Tracker’ on Google Play.* 2021.

[79] K. Gorkovenko, D. J. Burnett, J. Thorp, D. Richards, and D. Murray-Rust. “Supporting Real-Time Contextual Inquiry Through Sensor Data Supporting Real-Time Contextual Inquiry Through Sensor Data”. In: *Ethnographic Praxis in Industry Conference Proceedings.* Edinburgh, UK, 2019, pp. 1–29.

[80] K. Gorkovenko, D. J. Burnett, J. K. Thorp, D. Richards, and D. Murray-Rust. “Exploring the Future of Data-Driven Product Design”. In: *Conference on Human Factors in Computing Systems - Proceedings.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–14. ISBN: 9781450367080. DOI: `10.1145/3313831.3376560`.

[81] S. J. J. Gould. “Consumption experiences in the research process”. In: *CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: ACM, Apr. 2022, pp. 1–17. ISBN: 9781450391573. DOI: `10.1145/3491102.3502001`.

[82] R. W. Grant and J. Sugarman. “Ethics in human subjects research: Do incentives matter?” In: *Journal of Medicine and Philosophy* 29.6 (2004), pp. 717–738. ISSN: 03605310. DOI: `10.1080/03605310490883046`.

[83] C. M. Gray, C. Santos, N. Bielova, M. Toth, and D. Clifford. “Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective”. In: (2021), pp. 1–18. DOI: `10.1145/3411764.3445779`.

[84] J. Gray, C. Gerlitz, and L. Bounegru. “Data infrastructure literacy”. In: *Big Data and Society* 5.2 (2018). ISSN: 20539517. DOI: `10.1177/2053951718786316`.

[85] B. Greshake Tzovaras, M. Angrist, K. Arvai, M. Dulaney, V. Estrada-Galiñanes, B. Gunderson, T. Head, D. Lewis, O. Nov, O. Shaer, A. Tzovara, J. Bobe, and M. Price Ball. “Open

Humans: A platform for participant-centered research and personal data exploration". In: *GigaScience* 8.6 (2019), pp. 1–13. ISSN: 2047217X. DOI: 10.1093/gigascience/giz076.

[86] P. Grindrod. "Beyond privacy and exposure: Ethical issues within citizen-facing analytics". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016), pp. 1–20. ISSN: 1364503X. DOI: 10.1098/rsta.2016.0132.

[87] K. Grivas. "Digital Selves: Devices for intimate communications between homes". In: *Personal and Ubiquitous Computing* 10.2-3 (2006), pp. 66–76. ISSN: 16174909. DOI: 10.1007/s00779-005-0003-1.

[88] E. Hafen. "Personal Data Cooperatives – A New Data Governance Framework for Data Donations and Precision Health". In: *The Ethics of Medical Data Donation. Philosophical Studies Series* 137 (2019).

[89] J. Hanna. *An Overview of Contemporary Speculative Practice.* 2019.

[90] D. Haraway. "Situated Knowledges : The Science Question in Feminism and the Privilege of Partial Perspective". In: *Feminist Studies* 14.3 (1988), pp. 575–599.

[91] M. Hassenzahl, S. Heidecker, K. Eckoldt, S. Diefenbach, and U. Hillmann. "All you need is love: Current strategies of mediating intimate relationships through technology". In: *ACM Transactions on Computer-Human Interaction* 19.4 (2012). ISSN: 10730516. DOI: 10.1145/2395131.2395137.

[92] K. Helms. "Do you have to pee? A design space for intimate and somatic data". In: *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference.* San Diego, CA, USA: Association for Computing Machinery (ACM), 2019, pp. 1209–1222. ISBN: 9781450358507. DOI: 10.1145/3322276.3322290.

[93] Y. Hirst, S. T. Stoffel, H. R. Brewer, L. Timotijevic, M. M. Raats, and J. M. Flanagan. "Understanding Public Attitudes and Willingness to Share Commercial Data for Health Research: A Survey Study in the United Kingdom (Preprint)". In: *JMIR Public Health and Surveillance* 9 (2022). ISSN: 23692960. DOI: 10.2196/40814.

[94] S. Homewood. "Designing for the changing body: A feminist exploration of self-tracking technologies". In: *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), pp. 1–4. DOI: 10.1145/3170427.3173031.

[95] S. Homewood. "Reframing Design Problems Within Women's Health". In: *DRS2018: Catalyst* 2 (2018), pp. 25–28. DOI: 10.21606/drs.2018.337.

[96] K. Hoök and J. Löwgren. "Strong concepts: Intermediate-level knowledge in interaction Design research". In: *ACM Transactions on Computer-Human Interaction* 19.3 (2012), pp. 1–18. ISSN: 10730516. DOI: 10.1145/2362364.2362371.

[97] P. Hummel, M. Braun, and P. Dabrock. "Data Donations as Exercises of Sovereignty". In: *Philosophical Studies Series.* Vol. 137. Cham: Springer International Publishing, 2019, pp. 23–54. ISBN: 9783030043636. DOI: 10.1007/978-3-030-04363-6{_}3.

[98] "Incentivizing data donation". In: *Nature Biotechnology* 33.9 (2015), p. 885. ISSN: 15461696. DOI: 10.1038/nbt.3341.

[99] Y. Jabareen. "Building a Conceptual Framework: Philosophy, Definitions, and Procedure". In: *International Journal of Qualitative Methods* 8.4 (2009), pp. 49–62. ISSN: 1609-4069. DOI: 10.1177/160940690900800406.

[100] K. H. Jones. "Incongruities and Dilemmas in Data Donation: Juggling Our 1s and 0s". In: *Philosophical Studies Series.* Vol. 137. 2019, pp. 75–93. ISBN: 9783030043636. DOI: 10.1007/978-3-030-04363-6{_}5.

[101] J. Kaye, E. A. Whitley, D. Lund, M. Morrison, H. Teare, and K. Melham. “Dynamic consent: a patient interface for twenty-first century research networks”. In: *European Journal of Human Genetics* 23.2 (Feb. 2015), pp. 141–146. ISSN: 1018-4813. DOI: `10.1038/ejhg.2014.71`.

[102] A. H. Khan, S. Snow, S. Heiner, and B. Matthews. “Disconnecting: Towards a semiotic framework for personal data trails”. In: *DIS 2020 - Proceedings of the 2020 ACM Designing Interactive Systems Conference* (2020), pp. 327–340. DOI: `10.1145/3357236.3395580`.

[103] I. Koskinen, J. Zimmerman, T. Binder, J. Redström, and S. Wensveen. *Design Research through Practice: From the Lab, Field, and Showroom.* Elsevier Inc., 2012. ISBN: 978-0-12-385502-2.

[104] S. Kozubaev, C. Elsden, N. Howell, M. L. J. Søndergaard, N. Merrill, B. Schulte, and R. Y. Wong. “Expanding Modes of Reflection in Design Futuring”. In: *Conference on Human Factors in Computing Systems - Proceedings.* 2020, pp. 1–15. ISBN: 9781450367080. DOI: `10.1145/3313831.3376526`.

[105] J. Krutzinna and L. Floridi. “Ethical Medical Data Donation: A Pressing Issue”. In: *Philosophical Studies Series* 137 (2019), pp. 1–6. ISSN: 25428349. DOI: `10.1007/978-3-030-04363-6{\_}1`.

[106] J. Krutzinna and L. Floridi. *The Ethics of Medical Data Donation.* 2019, pp. 9–22. ISBN: 9783030043629. DOI: `10.1007/978-3-030-04363-6{\_}2`.

[107] J. Krutzinna, M. Taddeo, and L. Floridi. “An Ethical Code for Posthumous Medical Data Donation”. In: *Philosophical Studies Series.* Vol. 137. 2019, pp. 181–195. ISBN: 9783030043636. DOI: `10.1007/978-3-030-04363-6{\_}12`.

[108] A. Kurze, A. Bischof, S. Totzauer, M. Storz, M. Eibl, M. Brereton, and A. Berger. “Guess the Data: Data Work to Understand How People Make Sense of and Use Simple Sensor Data from Homes”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: ACM, Apr. 2020, pp. 1–12. ISBN: 9781450367080. DOI: `10.1145/3313831.3376273`.

[109] H. Kwon, J. E. Fischer, M. Flintham, and J. Colley. “The Connected Shower”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.4 (Dec. 2018), pp. 1–22. ISSN: 2474-9567. DOI: `10.1145/3287054`.

[110] B. Larsen, K. Morris, K. Quinn, M. Osborne, and C. Minahan. “Practice does not make perfect: A brief view of athletes’ knowledge on the menstrual cycle and oral contraceptives”. In: *Journal of Science and Medicine in Sport* 23.8 (2020), pp. 690–694. ISSN: 18781861. DOI: `10.1016/j.jsams.2020.02.003`.

[111] J. Lau, B. Zimmerman, and F. Schaub. “Alexa, Are You Listening?” In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (Nov. 2018), pp. 1–31. ISSN: 2573-0142. DOI: `10.1145/3274371`.

[112] H. Li, P. Jarusriboonchai, H. Müller, E. Harjuniemi, and J. Häkkilä. *Emotional Communication between Remote Couples: Exploring the Design of Wearable Ambient Displays.* 2020. DOI: `10.1145/3419249.3420139`.

[113] Y. Li, N. Vishwamitra, H. Hu, and K. Caine. “Towards A Taxonomy of Content Sensitivity and Sharing Preferences for Photos”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–14. ISBN: 9781450367080. DOI: `10.1145/3313831.3376498`.

[114] Y. Liu, Z. Song, and E. Tilevich. “Querying invisible objects: Supporting data-driven, privacy-preserving distributed applications”. In: *Proceedings of the 14th International Conference*

on Managed Languages and Runtimes. Prague, Czech Republic: Association for Computing Machinery, 2017, pp. 60–72. ISBN: 9781450353403. DOI: 10.1145/3132190.3132206.

[115] Y. Loukissas. *All Data Are Local: Thinking Critically in a Data-Driven Society*. The MIT Press, 2022. ISBN: 9780262545174.

[116] S. M. C. Loureiro, E. Friedmann, M. Breazeale, and I. Middendorf. "How can brands encourage consumers to donate data to a data-driven social partnership? An examination of hedonic vs. functional categories". In: *Journal of Business Research* 164.April 2022 (2023), p. 113958. ISSN: 01482963. DOI: 10.1016/j.jbusres.2023.113958.

[117] S. M. C. Loureiro, E. Friedmann, M. Breazeale, and I. Middendorf. "How can brands encourage consumers to donate data to a data-driven social partnership? An examination of hedonic vs. functional categories". In: *Journal of Business Research* 164.April (2023), p. 113958. ISSN: 01482963. DOI: 10.1016/j.jbusres.2023.113958.

[118] D. M. Low, K. H. Bentley, and S. S. Ghosh. "Automated assessment of psychiatric disorders using speech: A systematic review". In: *Laryngoscope Investigative Otolaryngology* 5.1 (Feb. 2020), pp. 96–116. ISSN: 2378-8038. DOI: 10.1002/lio2.354.

[119] E. Luger. "Consent reconsidered; Reframing consent for ubiquitous computing systems". In: *UbiComp'12 - Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012), pp. 564–567. DOI: 10.1145/2370216.2370310.

[120] E. Luger, S. Moran, and T. Rodden. "Consent for all: Revealing the hidden complexity of terms and conditions". In: *Conference on Human Factors in Computing Systems - Proceedings* (2013), pp. 2687–2696. DOI: 10.1145/2470654.2481371.

[121] M. E. Luka and M. Millette. "(Re)framing Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research". In: *Social Media and Society* 4.2 (2018). ISSN: 20563051. DOI: 10.1177/2056305118768297.

[122] Y. Luo, Y.-H. Kim, B. Lee, N. Hassan, and E. K. Choe. "FoodScrap: Promoting Rich Data Capture and Reflective Food Journaling Through Speech Input". In: *Designing Interactive Systems Conference 2021*. New York, NY, USA: ACM, June 2021, pp. 606–618. ISBN: 9781450384766. DOI: 10.1145/3461778.3462074.

[123] N. Malkin, J. Deatrick, A. Tong, P. Wijesekera, S. Egelman, and D. Wagner. "Privacy Attitudes of Smart Speaker Users". In: *Proceedings on Privacy Enhancing Technologies* 2019.4 (Oct. 2019), pp. 250–271. ISSN: 2299-0984. DOI: 10.2478/popets-2019-0068.

[124] L. Manikonda, A. Deotale, and S. Kambhampati. "What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants". In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. New Orleans, LA, USA: Association for Computing Machinery, 2018, pp. 229–235. ISBN: 9781450360128. DOI: 10.1145/3278721.3278773.

[125] K. Martin and H. Nissenbaum. "Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables". In: *Columbia Science and Technology Law Review* (2015), pp. 1–40. DOI: 10.2139/ssrn.2709584.

[126] M. S. McCoy, A. L. Allen, K. Kopp, M. M. Mello, D. J. Patil, P. Ossorio, S. Joffe, and E. J. Emanuel. "Ethical Responsibilities for Companies That Process Personal Data". In: *American Journal of Bioethics* 23.11 (2023), pp. 11–23. ISSN: 15360075. DOI: 10.1080/15265161.2023.2209535.

[127] M. S. McCoy, A. L. Allen, K. Kopp, M. M. Mello, D. J. Patil, P. Ossorio, S. Joffe, and E. J. Emanuel. "Ethical Responsibilities for Companies That Process Personal Data". In: *Ameri-*

can Journal of Bioethics 23.11 (2023), pp. 11–23. ISSN: 15360075. DOI: 10.1080/15265161.2023.2209535.

[128] K. L. McNulty, K. J. Elliott-Sale, E. Dolan, P. A. Swinton, P. Ansdell, S. Goodall, K. Thomas, and K. M. Hicks. “The Effects of Menstrual Cycle Phase on Exercise Performance in Eumenorrheic Women: A Systematic Review and Meta-Analysis”. In: *Sports Medicine* 50.10 (2020), pp. 1813–1827. ISSN: 11792035. DOI: 10.1007/s40279-020-01319-3.

[129] M. Mehrnezhad and T. Almeida. “Caring for intimate data in fertility technologies”. In: *Conference on Human Factors in Computing Systems - Proceedings*. Yokohama, Japan: Association for Computing Machinery (ACM), 2021, p. 11. ISBN: 9781450380966. DOI: 10.1145/3411764.3445132.

[130] N. Meng, D. Kekülluöğlu, and K. Vaniea. “Owning and Sharing”. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021), pp. 1–29. DOI: 10.1145/3449119.

[131] B. D. Mittelstadt and L. Floridi. “The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts”. In: *Science and Engineering Ethics* 22.2 (2016), pp. 303–341. ISSN: 14715546. DOI: 10.1007/s11948-015-9652-2.

[132] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, and J. Crowcroft. “Human-Data Interaction: The Human Face of the Data-Driven Society”. In: *SSRN Electronic Journal* 10 (2014), pp. 1–11. ISSN: 1556-5068. DOI: 10.2139/ssrn.2508051.

[133] R. Neisse, G. Baldini, G. Steri, and V. Mahieu. “Informed consent in Internet of Things: The case study of cooperative intelligent transport systems”. In: *2016 23rd International Conference on Telecommunications (ICT)*. IEEE, May 2016, pp. 1–5. ISBN: 978-1-5090-1990-8. DOI: 10.1109/ICT.2016.7500480.

[134] A. Ng, A. M. Walker, L. Wakschlag, N. Alshurafa, and M. Reddy. “Understanding Self-Tracked Data from Bounded Situational Contexts”. In: (2022), pp. 1684–1697. DOI: 10.1145/3532106.3533498.

[135] N. Nguyen. “The Cult of Garmin : Why Athletes Stick With These Smartwatches Over Apple or Samsung”. In: *The Wall Street Journal* (2022).

[136] P. J. Nickel. “The Ethics of Uncertainty for Data Subjects”. In: *The Ethics of Medical Data Donation. Philosophical Studies Series.* 137 (2019).

[137] H. Nissenbaum. “Privacy as Contextual Integrity”. In: *Washington Law Review* 79.1 (2004), pp. 101–139. ISSN: 09700420.

[138] R. Noortman, M. Funk, K. Andersen, and B. Eggen. “What Would Margaret Atwood Do? Designing for Ustopia in HCI”. In: *ACM International Conference Proceeding Series*. 2021, pp. 72–80. ISBN: 9781450385145. DOI: 10.1145/3464327.3464344.

[139] Y. O’Connor, W. Rowan, L. Lynch, and C. Heavin. “Privacy by Design: Informed Consent and Internet of Things for Smart Health”. In: *Procedia Computer Science* 113 (2017), pp. 653–658. ISSN: 18770509. DOI: 10.1016/j.procs.2017.08.329.

[140] J. Ohme and T. Araujo. “Digital data donations: A quest for best practices”. In: *Patterns* 3.4 (Apr. 2022), p. 100467. ISSN: 26663899. DOI: 10.1016/j.patter.2022.100467.

[141] J. Ohme, T. Araujo, L. Boeschoten, D. Freelon, N. Ram, B. B. Reeves, and T. N. Robinson. “Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking”. In: *Communication Methods and Measures* 00.00 (2023), pp. 1–18. ISSN: 19312466. DOI: 10.1080/19312458.2023.2181319.

[142] J. Ohme, T. Araujo, C. H. de Vreese, and J. T. Piotrowski. "Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function". In: *Mobile Media and Communication* 9.2 (2021), pp. 293–313. ISSN: 20501587. DOI: 10.1177/2050157920959106.

[143] C. S. Ooi, K. P. Seng, L. M. Ang, and L. W. Chew. "A new approach of audio emotion recognition". In: *Expert Systems with Applications* 41.13 (2014), pp. 5858–5869. ISSN: 09574174. DOI: 10.1016/j.eswa.2014.03.026.

[144] A. Oulasvirta, A. Pihlajamaa, J. Perkiö, D. Ray, T. Vähäkangas, T. Hasu, N. Vainio, and P. Myllymäki. "Long-term effects of ubiquitous surveillance in the home". In: *UbiComp'12 - Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* Association for Computing Machinery, 2012, pp. 41–50. ISBN: 9781450312240. DOI: 10.1145/2370216.2370224.

[145] L. Palen and P. Dourish. "Unpacking "Privacy" for a Networked World". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 5. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 129–136. ISBN: 1581136307. DOI: 10.1145/642611.642635.

[146] E. Parviainen and M. L. J. Søndergaard. "Experiential Qualities of Whispering with Voice Assistants". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376187.

[147] J. A. Pater, R. Pfafman, and A. Coupe. "Standardizing reporting of participant compensation in hci: A systematic literature review and recommendations for the field". In: *Conference on Human Factors in Computing Systems - Proceedings* (2021). DOI: 10.1145/3411764.3445734.

[148] D. Pavel, V. Callaghan, F. Sepulveda, M. Gardner, and A. K. Dey. "The story of our lives: From sensors to stories in self-monitoring systems". In: *2012 4th Computer Science and Electronic Engineering Conference, CEEC 2012 - Conference Proceedings* (2012), pp. 71–76. DOI: 10.1109/CEEC.2012.6375381.

[149] S. Petronio. *Boundaries of privacy: Dialectics of disclosure.* Albany: State University of New York Press, 2002. ISBN: 0791455157.

[150] S. Petronio. "Road to Developing Communication Privacy Management Theory: Narrative in Progress, Please Stand By". In: *Journal of Family Communication* 4.3-4 (2004), pp. 193–207. ISSN: 1526-7431. DOI: 10.1080/15267431.2004.9670131.

[151] N. Pfiffner and T. N. Friemel. "Leveraging Data Donations for Communication Research: Exploring Drivers Behind the Willingness to Donate". In: *Communication Methods and Measures* 00.00 (2023), pp. 1–23. ISSN: 19312466. DOI: 10.1080/19312458.2023.2176474.

[152] K. Pilgrim and S. Bohnet-Joschko. "Donating Health Data to Research: Influential Characteristics of Individuals Engaging in Self-Tracking". In: *International Journal of Environmental Research and Public Health* 19.15 (2022). ISSN: 16604601. DOI: 10.3390/ijerph19159454.

[153] K. Pilgrim and S. Bohnet-Joschko. "Effectiveness of Digital Forced-Choice Nudges for Voluntary Data Donation by Health Self-trackers in Germany: Web-Based Experiment". In: *Journal of Medical Internet Research* 24.2 (2022), e31363. ISSN: 14388871. DOI: 10.2196/31363.

[154] D. Pins, T. Jakobi, A. Boden, F. Alizadeh, and V. Wulf. "Alexa, We Need to Talk: A Data Literacy Approach on Voice Assistants". In: *Designing Interactive Systems Conference 2021.* New York, NY, USA: ACM, June 2021, pp. 495–507. ISBN: 9781450384766. DOI: 10.1145/3461778.3462001.

[155] B. Prainsack. "Data Donation: How to Resist the iLeviathan". In: *Philosophical Studies Series*. Vol. 137. 2019, pp. 9–22. ISBN: 9783030043636. DOI: 10.1007/978-3-030-04363-6{\$}{_}2.

[156] B. Prainsack. "Logged out: Ownership, exclusion and public value in the digital data and information commons". In: *Big Data & Society* 6.1 (Jan. 2019), p. 205395171982977. ISSN: 2053-9517. DOI: 10.1177/2053951719829773.

[157] B. Prainsack and S. El-Sayed. "Beyond Individual Rights: How Data Solidarity Gives People Meaningful Control over Data". In: *American Journal of Bioethics* 23.11 (2023), pp. 36–39. ISSN: 15360075. DOI: 10.1080/15265161.2023.2256267.

[158] R. Pranckutė. "Web of Science (WoS) and Scopus: the titans of bibliographic information in today's academic world". In: *Publications* 9.1 (2021). ISSN: 23046775. DOI: 10.3390/publications9010012.

[159] C. Qu, L. Yang, W. B. Croft, J. R. Trippas, Y. Zhang, and M. Qiu. "Analyzing and characterizing user intent in information-seeking conversations". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ann Arbor, MI, USA: Association for Computing Machinery, 2018, pp. 989–992. ISBN: 9781450356572. DOI: 10.1145/3209978.3210124.

[160] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury. "BodyBeat: A mobile system for sensing non-speech body sounds". In: *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*. Bretton Woods, New Hampshire, USA: Association for Computing Machinery, 2014, pp. 2–13. ISBN: 9781450327930. DOI: 10.1145/2594368.2594386.

[161] K. RAWSON and T. MUÑOZ. "Against Cleaning". In: *Debates in the Digital Humanities 2019* (2019), pp. 279–292. DOI: 10.5749/j.ctvg251hk.26.

[162] A. Razi, A. Alsoubai, S. Kim, N. Naher, S. Ali, G. Stringhini, M. De Choudhury, and P. J. Wisniewski. "Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection". In: *Conference on Human Factors in Computing Systems - Proceedings*. 2022, pp. 1–18. ISBN: 9781450391566. DOI: 10.1145/3491101.3503569.

[163] J. C. Read, D. Fitton, and M. Horton. "Giving ideas an equal chance: Inclusion and representation in participatory design with children". In: *ACM International Conference Proceeding Series* (2014), pp. 105–114. DOI: 10.1145/2593968.2593986.

[164] L. Reime, V. Tsaknaki, and M. L. Cohn. "Walking Through Normativities of Reproductive Bodies: A Method for Critical Analysis of Tracking Applications". In: *Conference on Human Factors in Computing Systems - Proceedings* (2023). DOI: 10.1145/3544548.3581450.

[165] G. Richter, C. Borzikowsky, B. F. Hoyer, M. Laudes, and M. Krawczak. "Secondary research use of personal medical data: patient attitudes towards data donation". In: *BMC Medical Ethics* 22.1 (2021), pp. 1–10. ISSN: 14726939. DOI: 10.1186/s12910-021-00728-x.

[166] Robert Koch Institut. *Corona-Datenspende*. 2020.

[167] B. Roeber, O. Rehse, R. Knorrek, and B. Thomsen. "Personal data: how context shapes consumers' data sharing with organizations from various sectors". In: *Electronic Markets* 25.2 (2015), pp. 95–108. ISSN: 14228890. DOI: 10.1007/s12525-015-0183-0.

[168] A. Rudnicka, A. L. Cox, and S. J. J. Gould. "Why Do You Need This?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, May 2019, pp. 1–11. ISBN: 9781450359702. DOI: 10.1145/3290605.3300622.

[169] K. Saha, J. Seybolt, and S. M. Mattingly. “What life events are disclosed on social media, how, when, and by whom?” In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan: Association for Computing Machinery, 2021, p. 22. ISBN: 9781450380966. DOI: `10.1145/3411764.3445405`.

[170] E. Sanders and P. Stappers. “Co-creation and the new landscapes of design”. In: *CoDesign* 4.1 (2008), pp. 5–18. DOI: `10.1080/15710880701875068`.

[171] B. Schulte and E. Hornecker. “Full frontal intimacy - On HCI, design & intimacy”. In: *DIS 2020 Companion - Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (2020), pp. 123–129. DOI: `10.1145/3393914.3395889`.

[172] A. Sciuto, A. Saini, J. Forlizzi, and J. I. Hong. “"Hey Alexa, What’s Up?"” In: *Proceedings of the 2018 Designing Interactive Systems Conference*. New York, NY, USA: ACM, June 2018, pp. 857–868. ISBN: 9781450351980. DOI: `10.1145/3196709.3196772`.

[173] C. Shani, A. Libov, S. Tolmach, and L. Lewin-eytan. ““ Alexa , Do You Want to Build a Snowman ?” Characterizing Playful Requests to Conversational Agents”. In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI ’22 Extended Abstracts*. New Orleans, LA, USA: Association for Computing Machinery (ACM), 2022, pp. 1–7. ISBN: 9781450391566. DOI: `10.1145/3491101.3519870`.

[174] D. M. Shaw, J. V. Gross, and T. C. Erren. “Data donation after death”. In: *EMBO reports* 17.1 (2016), pp. 14–17. ISSN: 1469-221X. DOI: `10.15252/embr.201541802`.

[175] J. L. Shirk, H. L. Ballard, C. C. Wilderman, T. Phillips, A. Wiggins, R. Jordan, E. McCallie, M. Minarchek, B. V. Lewenstein, M. E. Krasny, and R. Bonney. “Public participation in scientific research: A framework for deliberate design”. In: *Ecology and Society* 17.2 (2012). ISSN: 17083087. DOI: `10.5751/ES-04705-170229`.

[176] I. Shklovski and E. Grönvall. “CreepyLeaks: Participatory Speculation Through Demos”. In: *ACM International Conference Proceeding Series* (2020). DOI: `10.1145/3419249.3420168`.

[177] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. “Leakiness and creepiness in app space”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 2347–2356. ISBN: 9781450324731. DOI: `10.1145/2556288.2557421`.

[178] Y. Shvartzshnaider, T. Wies, P. Kift, and H. Nissenbaum. “Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms”. In: *Proceedings of the Fourth AAAI Conference on Human Computation and Crowdsourcing* 4.1 (2016), pp. 209–218. ISSN: 0888-8892.

[179] A. Skatova and J. Goulding. “Psychology of personal data donation”. In: *PLOS ONE* 14.11 (Nov. 2019). Ed. by V. Capraro, e0224240. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0224240`.

[180] A. Skatova, E. Ng, and J. Goulding. “Data Donation: Sharing Personal Data for Public Good?” In: *Digital Economy All Hands Meeting*. December. 2014, pp. 1–3. DOI: `10.13140/2.1.2567.8405`.

[181] A. Skatova, K. Shiells, and A. Boyd. “Attitudes towards transactional data donation and linkage in a longitudinal population study: Evidence from the Avon Longitudinal Study of Parents and Children”. In: *Wellcome Open Research* 4 (2019), pp. 1–16. ISSN: 2398502X. DOI: `10.12688/wellcomeopenres.15557.1`.

[182] S. Skoog Waller, M. Eriksson, and P. Sörqvist. “Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age”. In: *Frontiers in Psychology* 6.July (2015), pp. 1–11. ISSN: 16641078. DOI: `10.3389/fpsyg.2015.00978`.

[183] L. Sloan, C. Jessop, T. Al Baghal, and M. Williams. “Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving”. In: *Journal of Empirical Research on Human Research Ethics* 15.1-2 (Feb. 2020), pp. 63–76. ISSN: 1556-2646. DOI: `10.1177/1556264619853447`.

[184] L. R. Smith and J. Sanderson. “I’m Going to Instagram It! An Analysis of Athlete Self-Presentation on Instagram”. In: *Journal of Broadcasting and Electronic Media* 59.2 (2015), pp. 342–358. ISSN: 15506878. DOI: `10.1080/08838151.2015.1029125`.

[185] M. L. J. Søndergaard. “Troubling design: A design program for designing with women’s health”. In: *ACM Transactions on Computer-Human Interaction* 27.4 (2020). ISSN: 15577325. DOI: `10.1145/3397199`.

[186] M. L. J. Søndergaard. *Staying with the Trouble through Design: Critical-feminist Design of Intimate Technology*, p. 2018. ISBN: 978-87-7507-436-5. DOI: `10.7146/aul.289.203`.

[187] M. L. J. Sondergaard, M. C. Felice, and M. Balaam. “Designing menstrual technologies with adolescents”. In: *Conference on Human Factors in Computing Systems - Proceedings*. Yokohama, Japan: Association for Computing Machinery (ACM), 2021, p. 14. ISBN: 9781450380966. DOI: `10.1145/3411764.3445471`.

[188] C. Speed and J. Oberlander. “Designing from, with and by Data: Introducing the ablative framework”. In: *Proceedings of DRS 2016 International Conference: Future–Focused Thinking*. Brighton, UK: Design Research Society, 2016, pp. 2991–3004. DOI: `https://doi.org/10.21606/drs.2016.433`.

[189] Y. Strengers, J. Sadowski, Z. Li, and A. Shimshak. “What can hci learn from sexual consent? a feminist process of embodied consent for interactions with emerging technologies”. In: *Conference on Human Factors in Computing Systems - Proceedings* (2021). DOI: `10.1145/3411764.3445107`.

[190] V. Strotbaum, M. Pobiruchin, B. Schreiweis, M. Wiesner, and B. Strahwald. “Your data is gold – Data donation for better healthcare?” In: *it - Information Technology* 61.5-6 (Oct. 2019), pp. 219–229. ISSN: 2196-7032. DOI: `10.1515/itit-2019-0024`.

[191] C. A. Sula. “Research Ethics in an Age of Big Data”. In: *Bulletin of the Association for Information Science and Technology* 42.2 (2016), pp. 17–21. ISSN: 2373-9223. DOI: `10.1002/bul2.2016.1720420207`.

[192] M. Taddeo and L. Floridi. “What is Data Ethics ? Subject Areas : Author for correspondence :” in: *Philosophical transactions. Series A* 374(2083):Dec 28 (2016), pp. 1–5.

[193] J. Tholander and S. Nylander. “Snot, sweat, pain, mud, and snow - performance and experience in the use of sports watches”. In: *Conference on Human Factors in Computing Systems - Proceedings* 2015-April (2015), pp. 2913–2922. DOI: `10.1145/2702123.2702482`.

[194] P. Tolmie and A. Crabtree. “The practical politics of sharing personal data”. In: *Personal and Ubiquitous Computing* 22.2 (Apr. 2018), pp. 293–315. ISSN: 16174909. DOI: `10.1007/s00779-017-1071-8`.

[195] P. Tolmie, A. Crabtree, T. Rodden, J. Colley, and E. Luger. ““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 & Social Computing*. CSCW ’16. New York, NY, USA: ACM, Feb. 2016, pp. 491–502. ISBN: 9781450335928. DOI: `10.1145/2818048.2819992`.

[196] C. Tonkinwise. “How We Intend to Future: Review of Anthony Dunne and Fiona Raby, Speculative Everything: Design, Fiction, and Social Dreaming”. In: *Design Philosophy Papers* 12.2 (2014), pp. 169–187. DOI: `10.2752/144871314x14159818597676`.

[197] A. Tricco, E. L. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D.J. Peters, T. Horsley, L. Weeks, and S. Hem. "PRISMA Extension for Scoping Reviews (PRISMA-ScR)". In: *Annals of Internal Medicine* (2018), pp. 467–473. DOI: `doi.org/10.7326/M18-0850`.

[198] V. Tsoukas, A. Gkogkidis, and A. Kakarountas. "A Survey on Mobile User Perceptions of Sensitive Data and Authentication Methods". In: *24th Pan-Hellenic Conference on Informatics*. Athens, Greece: Association for Computing Machinery, 2020, pp. 346–349. ISBN: 9781450388979. DOI: `10.1145/3437120.3437337`.

[199] M. Turilli and L. Floridi. "The ethics of information transparency". In: *Ethics and Information Technology* 11.2 (2009), pp. 105–112. ISSN: 13881957. DOI: `10.1007/s10676-009-9187-9`.

[200] U.S. Chambers of Commerce Foundation. *The Future of Data-Driven Innovation*. Tech. rep. U.S. Chambers of Commerce Foundation, 2014, pp. 1–90.

[201] U.S. Food and Drug Administration. *Payment and Reimbursement to Research Subjects*. 2018.

[202] C. Utz, S. Becker, T. Schnitzler, M. Florian, F. F. Herbert, and L. Schaewitz. "Apps against the spread: Privacy implications and user acceptance of covid-19-related smartphone apps on three continents". In: *Conference on Human Factors in Computing Systems - Proceedings*. February. Yokohama, Japan.: ACM, 2021, p. 22. ISBN: 9781450380966. DOI: `10.1145/3411764.3445517`.

[203] J. Van Kollenburg, S. Bogers, H. Rutjes, E. Deckers, J. Frens, and C. Hummels. "Exploring the value of parent-tracked baby data in interactions with healthcare professionals: A data-enabled design exploration". In: *Conference on Human Factors in Computing Systems - Proceedings*. Vol. 2018-April. Association for Computing Machinery, Apr. 2018. ISBN: 9781450356206. DOI: `10.1145/3173574.3173871`.

[204] I. I. van Driel, A. Giachanou, J. L. Pouwels, L. Boeschoten, I. Beyens, and P. M. Valkenburg. "Promises and Pitfalls of Social Media Data Donations". In: *Communication Methods and Measures* 16.4 (2022), pp. 266–282. ISSN: 19312466. DOI: `10.1080/19312458.2022.2109608`.

[205] J. van Kollenburg and S. Bogers. *Data-enabled design : a situated design approach that uses data as creative material when designing for intelligent ecosystems*. Tech. rep. 2019. Eindhoven, The Netherlands: Eindhoven University of Technology, 2019.

[206] S. J. Verhoef, M. C. Wielink, E. A. Achterberg, M. Y. Bongers, and S. M. Goossens. "Absence of menstruation in female athletes: why they do not seek help". In: *BMC Sports Science, Medicine and Rehabilitation* 13.1 (2021), pp. 1–11. ISSN: 20521847. DOI: `10.1186/s13102-021-00372-3`.

[207] L. T. Vidal, H. Zhu, A. Waern, and E. M. Segura. "The design space ofwearables for sports and fitness practices". In: *Conference on Human Factors in Computing Systems - Proceedings* (2021). DOI: `10.1145/3411764.3445700`.

[208] M. Warner, A. Kitkowska, J. Gibbs, J. F. Maestre, and A. Blandford. "Evaluating 'Prefer not to say' Around Sensitive Disclosures". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: `10.1145/3313831.3376150`.

[209] J. Wiese, S. Das, J. I. Hong, and J. Zimmerman. "Evolving the Ecosystem of Personal Behavioral Data". In: *Human-Computer Interaction* 32.5-6 (2017), pp. 447–510. ISSN: 07370024. DOI: `10.1080/07370024.2017.1295857`.

[210] J. Wilbanks and S. H. Friend. "First, design for data sharing". In: *Nature Biotechnology* 34.4 (2016), pp. 377–379. ISSN: 15461696. DOI: `10.1038/nbt.3516`.

[211] K. Williams, A. Bigelow, and K. E. Isaacs. "Data Abstraction Elephants: The Initial Diversity of Data Representations and Mental Models". In: *Conference on Human Factors in Computing Systems - Proceedings*. 2023. ISBN: 9781450394215. DOI: `10.1145/3544548.3580669`.

[212] A. Wolff, D. Gooch, J. Cavero Montaner, U. Rashid, and G. Kortuem. "Creating an understanding of data literacy for a data-driven society". In: *Journal of Community Informatics* 12.3 (2017), (In press). ISSN: 1712-4441.

[213] World Economic Forum. *Personal Data : The Emergence of a New Asset Class An Initiative of the World Economic Forum*. 2011. ISBN: 1000142405274.

[214] N. C. Woytuk, J. Y. Park, J. Maslik, M. C. Felice, and M. Balaam. "Tactful Feminist Sensing : Designing for Touching Vaginal Fluids". In: *In Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. 2023, pp. 2642–2656. ISBN: 9781450398930. DOI: `10.1145/3563657.3595966`.

[215] S. Zannettou, O.-N. Nemeth, O. Ayalon, A. Goetzen, K. P. Gummadi, E. M. Redmiles, and F. Roesner. "Leveraging Rights of Data Subjects for Social Media Analysis: Studying TikTok via Data Donations". In: *ArXiv* (2023).

[216] J. Zimmerman, J. Forlizzi, and S. Evenson. "Research through design as a method for interaction design research in HCI". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2007, pp. 493–502. ISBN: 9781595935939. DOI: `10.1145/1240624.1240704`.

[217] M. Zook, S. Barocas, D. Boyd, K. Crawford, E. Keller, S. P. Gangadharan, A. Goodman, R. Hollander, B. A. Koenig, J. Metcalf, A. Narayanan, A. Nelson, and F. Pasquale. "Ten simple rules for responsible big data research". In: *PLoS Computational Biology* 13.3 (2017), pp. 1–10. ISSN: 15537358. DOI: `10.1371/journal.pcbi.1005399`.

List of Publications

Conference and Journal Publications

- **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2024) 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). doi: 10.1145/3643834.3661524
- **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2024) 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). doi: 10.1145/3643834.3661503
- Wundsam A., **Gómez Ortega, A.**, and Cila, N. (2024) Intimate Data as a Design Material: Designing Tracking Practices for Menstruating Athletes. In DRS2024.
- **Gómez Ortega, A.**, Noortman, R., Bourgeois, J., and Kortuem, G. (2024) Dataslip: Into the Present and Future(s) of Personal data. In Proceedings of the Eighteen International Conference on Tangible, Embedded, and Embodied Interaction (TEI '24). doi: 10.1145/3623509.3633388
- **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2023) Personal Data Comics: A Data Storytelling Approach Supporting Personal Data Literacy. In XI Latin American Conference on Human-Computer Interaction (CLIHC 2023). doi: 10.1145/3630970.3630982
- **Gómez Ortega, A.**, Bourgeois, J., Hutiri, W.T., and Kortuem, G. (2023) Beyond Data Transactions: A Framework for Meaningfully Informed Data Donation. AI & Society. doi: 10.1007/s00146-023-01755-5
- **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2023) Understanding the Challenges Around Design Activities that Incorporate Behavioral Data. Proceedings of the Design Society, 3, 3711-3720. doi: 10.1017/pds.2023.372
- **Gómez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2023) What is Sensitive about (Sensitive) Data? Characterizing Sensitivity and Intimacy with Google Assistant Users In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). doi: 10.1145/3544548.3581164
- **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 (NordiCHI '22). doi: 10.1145/3546155.3546646
- Lu, J., **Gómez Ortega, A.**, Gonçalves, M., and Bourgeois, J. (2021). The Impact of Data on the Role of Designers and their Process. Proceedings of the Design Society. doi: 10.1017/pds.2021.563

Workshop Papers and Demonstration Papers

- **Gómez Ortega, A.**, Milias, V., Broadhead, J., van der Valk, C, and Bourgeois, J. (2023). Dataslip:

How far does your personal data go? (2023) Human-Computer Interaction – INTERACT 2023. Lecture Notes in Computer Science, vol 14145. doi : 10.1007/978-3-031-42293-5_25

- **Gomez Ortega, A.**, Bourgeois, J., and Kortuem, G. (2021) Towards Designerly Data Donation. In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '21) doi: 10.1145/3460418.3479362

Special Interest Group Papers

- **Gomez Ortega, A.**, Lovei, P., Noortman, R., Toebosch, R., Bowyer, A., Kurze, A., Funk, M., Gould, S., Huron, S., and Bourgeois, J. (2023) SIG on Data as Human-Centered Design Material. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23). doi: 10.1145/3544549.3583180
- **Gomez Ortega, A.**, van Kollenburg, J., Shen, Y., Murray-Rust, D., Nedić, D., Jimenez, J., Meijer, W., Chaudhary, P., and Bourgeois, J. (2022) SIG on Data as Human-Centered Design Material. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). doi: 10.1145/3491101.3516403

Curriculum Vitæ

Photo: Marcel Krijger

Alejandra Gómez Ortega is a Colombian Human-Computer Interaction researcher. From 2012 to 2016 she attended Universidad de Los Andes (Bogotá, Colombia) where she completed her bachelor's degree in Electronic Engineering and a minor in creative writing. Between 2016 and 2018 she completed a M.S.c in Biomedical Engineering at Politecnico di Milano (Milan, Italy). In Milan, she worked as a Junior Researcher developing wearables for activity monitoring at Ab.Acus – a research-driven company in the healthcare domain – until 2020. She conducted this research from 2020 to 2024 at Delft University of Technology (Delft, The Netherlands) in the Sustainable Design Engineering department in the faculty of Industrial Design Engineering.


BANK

TU Delft