ACTA

C

TECHNICA

Andy Alorwu

USER PERCEPTIONS OF PERSONAL DATA IN HEALTHCARE: ETHICS, REUSE, AND VALUATION

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

ANDY ALORWU

USER PERCEPTIONS OF PERSONAL DATA IN HEALTHCARE: ETHICS, REUSE, AND VALUATION

Academic dissertation to be presented with the assent of the Doctoral Programme Committee of Information Technology and Electrical Engineering of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 4 August 2023, at 12 noon

Alorwu, Andy, User perceptions of personal data in healthcare: ethics, reuse, and valuation.
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering
Acta Univ. Oul. C 887, 2023
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

There has been global concern about how personal data are collected and managed. A focused examination of people's perceptions of their personal data collection and use is relevant given the growing importance of personal data in the use of modern technologies. This thesis focuses on people's perceptions of their personal data particularly within the health domain. The exploration is carried out from multiple perspectives: ownership; ethics; reuse; and valuation. Future health services that rely on personal data may be developed considering several data ownership, reuse, privacy, and ethical issues. This thesis is conducted with two core objectives: 1) to develop research probes for collecting personal data in the health domain; and 2) to conduct a series of case studies on the perception of personal data using the research probes.

The findings of this thesis demonstrate people's willingness to donate personal data of varying sensitivity levels both for monetary benefits and for social good. Further, it presents insights into personal data management based on empirical investigations using purpose-built research probes complemented by online experiments. Throughout the thesis, I highlight opportunities and challenges that people consider critical to them regarding the collection, storage, processing, and management of their personal data. In this thesis, I demonstrate that gaining an understanding of user perceptions of personal data can benefit digital services that rely on such data. This thesis also highlights the important role that crowdsourcing marketplaces can play by serving as human-subject pools to contribute a vast amount of data.

Towards the end of this thesis, I revisit the research questions and highlights how they were answered. The thesis also discusses the implications of personal data for future digital health services that rely on such data. I then take a look into the future with a potential paradigm for managing personal data in a more granular way. I conclude the thesis by restating the thesis's aim and objectives, and consider the opportunities and challenges of using personal data and possible directions for future research.

Keywords: data reuse, data valuation, ethics, MyData, personal data, privacy

Alorwu, Andy, Henkilökohtaisen tiedon uudelleenkäytön mahdollisuudet, eettisyys ja arvo loppukäyttäjien näkökulmasta terveydenhuollon sovelluksissa.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähkötekniikan tiedekunta
Acta Univ. Oul. C 887, 2023
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Henkilötietojen keräämiseen ja hallintaan liittyy maailmanlaajuinen huoli. On merkityksellistä tarkastella ihmisten käsityksiä henkilötietojensa keräämisestä ja käytöstä, ottaen huomioon henkilötietojen kasvavan merkityksen modernien teknologioiden käytössä. Tämä väitöskirja keskittyy ihmisten käsityksiin heidän henkilötiedoistaan erityisesti terveysalalla. Tarkastelussa on mukana useita näkökulmia: omistajuus; etiikka; uudelleenkäyttö; ja arvon määrittäminen. Tämän väitöskirjan kaksi ydintavoitetta ovat: 1) kehittää tutkimussovelluksia henkilötietojen keräämiseksi terveysalalla; ja 2) suorittaa sarja tapaustutkimuksia henkilötietojen käsityksestä käyttäen tutkimussovelluksia.

Tämän väitöskirjan tulokset osoittavat ihmisten halukkuuden lahjoittaa vaihtelevan arkaluontoisia henkilötietoja sekä rahallisten etujen että yhteiskunnallisen hyvän vuoksi. Lisäksi väitöskirja tarjoaa näkemyksiä henkilötietojen hallintaan perustuen empiirisiin tutkimuksiin, joissa käytetääntarkoitukseen kehitettyjä tutkimussovelluksia sekä täydentävää verkkotutkimusta. Väitöskirja korostaa mahdollisuuksia ja haasteita, joita ihmiset pitävät kriittisinä liittyen heidän henkilötietojensa keräämiseen, säilytykseen, käsittelyyn ja hallintaan. Väitöskirja osoittaa, että käyttäjien henkilötietoihin liittyvien käsitysten ymmärtäminen voi hyödyttää digitaalisia palveluita, jotka nojaavat tällaisiin tietoihin. Väitöskirja myös korostaa kaupallisten joukkoistamisalustojen roolia tutkitun tiedon tuottamisessa.

Väitöskirjan lopussa palaan tutkimuskysymyksiin ja näiden vastauksiin. Pohdin henkilötietojen merkitystä tulevaisuuden digipalveluille, jotka ovat näistä tiedoista riippuvaisia. Luonuo katsauksen tulevaisuuteen, joka tarjoaa yksityiskohtaisemman tavan hallita henkilötietoja. Päätän väitöskirjan toistamalla sen tavoitteen ja pohdin henkilötietojen käytön mahdollisuuksia ja haasteita sekä mahdollisia suuntia tulevalle tutkimukselle.

Asiasanat: datan arvostus, datan saatavuus, etiikka, henkilötieto, omadata, yksityisyys

To all unsung heroes out there!

Acknowledgements

First and foremost, I give thanks to God for this moment. I am grateful for this opportunity and I am thankful that He has brought me this far. I would like to express my gratitude towards my supervisor, Assoc. Prof. Simo Hosio. Our fun-filled talks over coffee have been of great inspiration and the work presented in this thesis greatly benefited from them. Without his guidance and belief in my capabilities, critical and yet heartful comments, unwavering support and commitment to goals, this thesis would not have been possible. Both our serious and ridiculous conversations were important to me over the years.

I am also grateful to the members of my follow-up group, Dr. Karin Väyrynen, Dr. Dorina Rajanen and Dr. Juha Partala for providing valuable feedback over the years. I would also like to thank Dr. Denzil Ferreira and Dr. Aku Visuri for their roles as supporting supervisors during the conduct of this thesis. I thank all my co-authors, and extend a special appreciation to Assoc. Prof. Niels van Berkel, without whom much of the work done in this thesis would not have been possible. Thank you for all the help along the way. This work was conducted in the Crowd Computing group at the Center for Ubiquitous Computing and I am grateful for all the help I received from members of my group as well as the research unit. A big thank you to Mr. Ville Paananen and Dr. Paula Alavesa for helping out with the Finnish translations.

I thank my pre-examiners Assoc. Prof. Mohamed Khamis (University of Glasgow) and Assist. Prof. Kevin Doherty (University College Dublin) for their valuable comments and suggestions. In addition, I would like to thank Prof. Barry Brown (Stockholm University) for agreeing to act as the official opponent in the thesis defence.

My infinite thanks goes to my family, my father Mr Felix Alorwu, my mother Ms Grace Odani, siblings Ms Sheila Alorwu, Ms Patricia Alorwu, and Mr. Elvis Alorwu for all the support and love throughout my life. I also extend sincere gratitude to Prof. Jemmy Takrama, Prof. Nii Cudjoe, Mr. Alan Dowokpor, Mr. Janne Rasanen, Ms Maria Velasco, Rev. Daniel Yirenkyi, and a host of others who in one way or the other contributed immensely to me being here at this point.

My sincere appreciation goes to Ms Eunice E.Y.F. Agyei and Mr. Yaw Bram-Larbi for your support and ever presence in my times of need. You rock! Now to the Ghanaian community in Oulu, I would not have been able to do this without you! Your support was unwavering and I enjoyed our weekends and game nights together. I would also like to thank my friends back home. A special thanks to Mr. Juan Chiaradia, Ms Sara Eggert,

List of abbreviations

ANOVA   *Analysis of variance*
AWS   *Amazon Web Services*
CS   *Crowdsourcing*
CSV   *comma-separated values*
e.g.   *exempli gratia*
etc.   *et cetera*
et al.   *et alia*
GDPR   *General Data Protection Regulation*
HCI   *Human-Computer Interaction*
HIT   *Human Intelligence Task*
HSD   *honest significant difference*
i.e.   *id est*
IoT   *Internet of Things*
LBP   *Low Back Pain*
LSM   *Least Square Means*
m-Health   *Mobile Health*
MTurk   *Amazon Mechanical Turk*
OHD   *Open Health Data*
PD   *Public Display*
RQ   *Research Question*
SD   *Standard Deviation*
SDK   *Software Development Kit*

List of original publications

The following original publications serve as the foundation of this thesis. They are identified in the text by their Roman numerals (I-V).

I   Alorwu A, van Berkel N, Goncalves J, Oppenlaender J, López M B, Seetharaman M, & Hosio S (2020). Crowdsourcing sensitive data using public displays—opportunities, challenges, and considerations. *Personal and Ubiquitous Computing*, 1-16. doi: 10.1007/s00779-020-01375-6

II   Alorwu A, Kheirinejad S, van Berkel N, Kinnula M, Ferreira D, Visuri A, & Hosio S (2021). Assessing MyData Scenarios: Ethics, Concerns, and the Promise. *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-11). doi: 10.1145/3411764.3445213

III   Alorwu A, Visuri A, van Berkel N, & Hosio S J (2021). (Re) using Crowdsourced Health Data: Perceptions of Data Contributors. *IEEE Software*, 39(1). doi: 10.1109/MS.2021.3117684

IV   Kheirinejad S, Alorwu A, Visuri A, & Hosio S J (2022). Contrasting the Expectations and Experiences of Mobile Health Use on Chronic Pain: A Questionnaire Study. *JMIR Human Factors*, 9(3), doi: 10.2196/38265

V   Alorwu, A, van Berkel, N, Visuri, A, Suryanarayana, S, Yoshihiro, T, & Hosio, S (under review, 202X). Monetary Valuation of Personal Health Data in the Wild. *IJHCS*

Contents

1   Introduction

1.1   Motivation

New generations of networked computing systems and mobile data collection mechanisms allow new types of end-user applications to be created that rely on increasing amounts of user data, and that are already playing an important role in people's lives. In this thesis, we follow the definition of Edwards (2018), and consider personal data as any information that is related to an identified or identifiable individual. Personal data have gained much attention globally, with a lot of emphasis on the policy level. Personal data initiatives such as the General Data Protection Regulation (GDPR) have been implemented to help in the management of data (European Commission, 2016). GDPR is based on the notion of privacy as a fundamental human right as set out in the Charter of EU Rights (Goddard, 2017).

Social media platforms such as Facebook, Instagram, and Twitter rely heavily on user data to provide users with personalized content. E-commerce websites use user data to provide recommendations and suggest products that the user might be interested in. Health and fitness applications use user data to create personalized workout plans and provide insights into the user's health. These applications are already playing an important role in people's lives by making daily tasks easier and more efficient. Similarly, they generate tremendous amounts of personal information such as demographic and location data, social connections, and health information on a daily basis (Beigi, Shu, Zhang, & Liu, 2018; Eke, 2011). Such data have afforded researchers the opportunity to conduct various forms of research, including on election forecasting, interactions and behaviour analysis, fake news, and political polarization. With the increasing reliance on user data, it is important to ensure that such data is collected and used in an ethical and responsible manner to protect user privacy and prevent the misuse of people's personal information.

In a similar vein, mobile and smart devices are increasingly becoming pervasive (Janeček, 2018), fostering and facilitating the delivery of healthcare, and also generating and collecting personal data on a large scale. The control and management of such data pose serious ethical (Van den Hoven, 2013) and legal (Drexl et al., 2016) concerns. At the centre of the issues surrounding data management and control is the ownership of personal data and its effects on the digital economy (Farkas, 2017; Janeček, 2018). Mobile and smart health-focused devices allow vast amounts of data to be generated and

collected, enabling the capture of user-generated health data that will eclipse the amount of captured data in clinical settings (Kish & Topol, 2015).

In recent years, there has been a global concern about how personal data are managed. These concerns are related to data collection, sharing, ownership, and data management practices. The role of personal data in modern technology use calls for a focused investigation of the perception of people about their own personal data. With personal data becoming a key enabler of modern-day technologies, attention must be paid to ethical, consensual data collection, and data management practices. This thesis seeks to explore user perceptions of their personal data from multiple perspectives: ownership; reuse; ethics; and valuation. Central to this investigation is MyData, a personal data management approach and set of principles (Kuikkaniemi, Poikola, & Honko, 2015) that is gaining international attention.

1.2   Overview of the thesis

This thesis consists of five original articles which are referred to in the text by their Roman numerals (I - V). Four out of five have been published in relevant peer-reviewed journals (Articles I, IV), at an international conference (Article II), and a magazine (Article III), all within the fields of Human Computer Interaction (HCI) and Software Engineering. Article V has been submitted to an international conference for review.

Article I investigates the willingness of people to donate potentially sensitive personal data in public settings using public display probes. Article II provides an investigation into the perspective of people regarding the ethicality and concerns felt about various data management scenarios. Article III explores crowdsourced open health data as an enabler of future software solutions. Article IV investigates the perceptions of people with chronic pain of the use and/or non-use of mobile health (mHealth) solutions with respect to data ownership and privacy. Finally, Article V investigates the perceived monetary value people attach to their personal health data.

All the articles used in this thesis were largely informed by the findings from earlier studies. Article I consisted of two studies (A and B) of which Study B was a follow-up study informed by emerging issues identified from Study A. Similarly, Articles II and V were informed by Article I's discovery that individuals experienced a loss of ownership of their data and had no real credible insights about monetization of their data. Additionally, Article IV's call for greater collaboration among healthcare stakeholders and the need to gain knowledge about transparent and ethical data management practices stems from Articles I, III, and V's highlight of users' dissatisfaction with private companies' practices. Users' reactions to the data management scenarios in Article II

inspired Article III's investigation into their perception of reuse of personal data donated for public use.

1.2.1 *Research aim, objectives, and questions*

The aim of this thesis is to investigate people's perception of personal data particularly within the health domain. More specifically, this thesis sheds light on the valuation, reuse, and ethics of personal data in the building of future health services and tools. With this aim, this thesis pursues the following two key objectives:

– O1: To develop research probes for collecting personal data in the health domain and
– O2: To conduct a series of case studies on the perception of personal data using the developed research probes

In pursuit of the research aim and objectives, this thesis seeks to answer the following research questions (RQs).

RQ1: How do human perceptions of ownership and the ethics of reuse differ concerning personal health data of varying sensitivity levels?

The first part of the thesis investigates the perceptions of users concerning the ownership of their personal data, as well as the privacy and ethical considerations surrounding the donation of personal data (Articles I-V). We also examine the privacy-related issues surrounding the use of mobile health solutions (Article IV) and its implications for personal data ownership and management (Articles I and IV).

RQ2: What values (ethical and monetary) do people attach to their personal data?

We investigate people's perceptions of the ethicality of data management practices of differing types. We conduct this study under the framework of MyData, an emerging data management principle and set of guidelines that seeks to place the user at the centre of the data management process (Article II). Further, we investigate the ethical (Article I) and monetary (Article V) values people attach to their personal data of different types.

RQ3: What ethical issues of personal health data reuse do people identify in future software-driven healthcare solutions?

Table 1. The contributions of each original article to the research questions.

Research questions	Articles
RQ1: How do human perceptions of ownership and ethics of reuse differ concerning personal health data of varying sensitivity levels?	I-V
RQ2: What values (ethical and monetary) do people attach to their personal data?	I, II, V
RQ3: What ethical issues of personal health data reuse do people identify in future software-driven healthcare solutions?	III, IV

We investigate the ethical issues that users identify in the use of their personal data for building software solutions. We investigate RQ3 within the context of user reactions to privacy, ethical, and data-related issues (Article IV) in connection with the reuse of previously donated data for future software solutions and initiatives such as open data (Article III).

1.2.2   Articles, contributions, and author's role

The author developed two mobile applications (Articles I and V) to elicit experiences from users. Table 2 provides a summary of the two technical solutions resulting from this thesis. Article II contributes a survey-based method and instrument for assessing data management perceptions of people. Articles III and IV were survey-based studies and are therefore not listed in Table 2.

In Table 1, a summary of the contributions of each article to the research questions is presented. The following outlines the research contributions of each original article as presented in this thesis:

Article I explores user perceptions of issues surrounding the donation of personal data in public using interactive public display technologies and the ethical issues surrounding the collection of such data. The article contributes a dynamic, easy-to-install public display tool (referred to as *Videosourcing*) for the collection of selfie video files tagged with metadata from passers-by. The Videosourcing application was deployed for 61 days, supported by one post questionnaire and one narrative transportation study. This article describes the analysis of the studies and highlights the perceptions of people donating different types of personal data in public spaces (RQ1). The article further provides an understanding of aspects of the data collection concept and ethical concerns to be considered in future research ecosystems that exploit situated technologies as user-facing personal data sources (RQ2). Author contribution: The author of this thesis

designed the app used in the study, designed the questionnaires, and led the writing of the article. With the co-authors, the author also analysed the collected data.

Article II presents a survey-based method and instrument for assessing data management scenarios. We collect people's assessment of ethicality and the concern they feel about topical data management scenarios (RQ2). We identified stark differences across parameters that affected people's evaluation of ethicality and the concern they felt about the presented scenarios. Participants were excited at the prospect of being in control of their personal data (RQ3) – especially their health data. Our results also highlight how people respond to the management of their personal data. Author contribution: The author of this thesis designed the survey instrument used in the study, designed the questionnaire, and led the writing of the article. The author also took the main responsibility for the qualitative analysis with two other co-authors.

Article III focuses on crowdsourced open data as an enabler of future software solutions. The article elicits data donor perceptions on issues surrounding the ethics (RQ1) and reuse of their previously donated health data in software (RQ3). The results of the qualitative analysis contributes an in-depth investigation into a range of perceived opportunities and threats in using crowdsourced open data to enable future software solutions. Author contribution: The author of this thesis designed the questionnaire used in the study, planned the study, and led the analysis and writing of the article.

Article IV contrasts the experiences and expectations of people with chronic pain concerning the use of mHealth solutions for managing chronic pain. The article highlights the benefits of mHealth solutions and the perceived obstacles to their adoption. It also examines the privacy concerns users have about mHealth solutions collecting their data and further explores data ownership-related issues (RQ1). Author contribution: The author of this thesis supported the design of the questionnaires and led efforts in the qualitative analysis of the survey data. The author also contributed to the writing of the manuscript.

Article V presents a mobile application that investigates the monetary value that people attach to their personal health data (RQ2). The app was deployed for 14 days, with 56 participants. The article outlines the perceptions of people of the value of their data, the concerns they have, and their opinions about benefiting financially from their personal data. The article contributes a mobile app (referred to as *LBP*) for tracking sleep and low back pain data daily. It also includes an auction mechanism for valuing personal data. Author contribution: The author of this thesis designed the mobile app used in the study, designed the online questionnaire, planned and conducted the study, and led the analysis and writing of the article.

Table 2. Technical contributions of the thesis.

Article	Description
I	Tablet-based Android application for eliciting video donation on public displays
II	Web-based instrument for assessing data management perceptions
V	Android mobile application for collecting daily health data and running auctions using the data

1.2.3 Structure of the thesis

The remainder of the thesis is organised as follows. Chapter 2 describes the background of personal data, crowdsourcing marketplaces as a human subject pool, data re(use) ethics, and user perceptions. In Chapter 3, an overview of the five articles included in this thesis is presented alongside a summary of the research methods used. In Chapter 4, we explore the contributions from the five articles included in this thesis. In Chapter 5, we revisit the research questions and discuss the findings and limitations as well as implications for the future. Chapter 6 concludes the thesis.

2 Related work

2.1 Personal data

Any information related to an identified or identifiable individual can be considered personal data (Edwards, 2018). Personal data today have an increasingly significant social, economic, and practical value; a valuable resource pioneering a new economic asset class (Zuboff, 2019) that will affect all aspects of society (Kuikkaniemi et al., 2015). Many attempts have been made in recent years to assign legal claims to personal data. Of particular interest is the call for individual property rights to be applied to personal data (Lanier, 2014). A counter-argument is that personal data are not creative outputs but simple facts which cannot be owned (Cohen, 2019). How personal data are intertwined in our everyday lives makes it difficult to separate which data belongs to whom. For example, does one's credit card information belong to the person or the credit card company? Who can claim ownership of such data? Cohen (2019) further argues that if personal data were treated as property, it would be both conceptually and methodologically difficult to determine which personal data belongs to whom. Regardless of the position taken in this argument, the fact remains that the collection and use of personal information remains critical globally (Birch, Cochrane, & Ward, 2021). The advance in technology has empowered the collection, storage, and use of data on a global scale. Personal data are one such resource that is being collected through various digital processes and tools (Birch et al., 2021; Cohen, 2019; Poikola et al., 2020).

Research has shown that the type of personal information affects its valuation. For example, a study by Huberman, Adar, and Fine (2005) found that personal information was valued according to its usefulness in a particular context. The subjective privacy value of personal information has been empirically measured across a variety of contexts. The disclosure of personal information online (Hann, Hui, Lee, & Png, 2007) and access to location data (Cvrcek, Kumpost, Matyas, & Danezis, 2006) are examples. There is a growing interest within the Human-Computer Interaction and Ubiquitous Computing fields in understanding people's perceptions of their personal data: how they are collected, stored, processed, who can access them, and for what (Barkhuus et al., 2008; Consolvo et al., 2005; Toch & Levi, 2013).

2.1.1   *MyData - a new data model*

MyData is a model for the use of personal data based on the presupposition that data producers (people) themselves can use, manage, and give permissions for the use of data collected about them (Poikola et al., 2020). MyData is thus a new data management approach that seeks to transform the current organisation-centric system to a human-centric system of data ownership and control, transferring the ownership and control of one's data to the individual (Kuikkaniemi et al., 2015; Rissanen, 2016). The vision of MyData has emerged to provide technical and ethical guidelines for a balanced data management between individuals who produce data and organisations that harness the data for various use cases. This has opened a new frontier for both researchers and practitioners (Poikola et al., 2020). In the health sciences, MyData has the potential to revolutionise how research is conducted, as patients bring with them vast amounts of data to analyse. To effectively use collected health-related data, data producers should be willing to measure, store, and manage their own data (Kim, Park, et al., 2012), as having control of one's personal data leads to better motivation to take care of one's own health (Baudendistel et al., 2015).

Indeed, the benefits of MyData have been found to surpass simply having access and control of one's data to becoming an enabler of better self-care (Baudendistel et al., 2015). The GDPR regulation passed by the European Union aims to ensure the rights of data producers (European Commission, 2016). However, it still follows the organisation-centric model of data management practices. MyData aims to build on top of GDPR by altering the model from an organisation-centric view to a human-centric one, in which the data producer is at the centre of the data management process. MyData's human-centric approach pushes for the release of personal data from the confines of monopolistic data holders for the full potential of personal data to be realised (Lehtiniemi & Haapoja, 2020; Poikola et al., 2020). The MyData approach envisions an infrastructure in which data producers can access and control their personal data, grant permission for the use of such data, and revoke such use.

2.2   Health-themed research using crowdsourcing marketplaces

Crowdsourcing (CS) is the process of distributing work to the many, as opposed to performing it by the specialised few (Howe et al., 2006). Crowd workers are people who complete tasks in CS marketplaces for monetary compensation. CS has been employed for a variety of tasks, including ideation, problem solving, and opinion gathering (Kittur et al., 2013) - including in healthcare (Brabham, Ribisl, Kirchner, & Bernhardt, 2014;

Hosio et al., 2018; Swan, 2012). In addition to online crowdsourcing, the ubiquitous nature of smartphones has made both mobile crowdsourcing and public display setups gain prominence in the collection of data in specialised social and cultural contexts (Alt, Shirazi, Schmidt, Kramer, & Nawaz, 2010).

An analysis of crowdsourced health research studies showed that participatory health, the involvement of people in their own healthcare, is fast becoming part of the public health ecosystem. Facilitated partly by the growing use of the internet and social networking tools, participatory health transforms healthcare from a traditional focus on curing disease to a more personalised preventive approach, making crowdsourced health research a promising complement to traditional clinical trials as a model for conducting health research (Swan, 2012).

In this regard, a study was undertaken by Aghdasi et al. (2015) to evaluate how well a sizable group of laypeople could rate a surgical procedure (cricothyrotomy) carried out on a simulator. Amazon Mechanical Turk (MTurk) was employed as a human subject pool. In addition, three experienced surgeons (not crowd workers) watched the footage and provided their evaluations. Compared to expert surgeons, the crowd workers took significantly less time to complete the analyses (10 h vs. 60 days). The crowd workers' evaluations of the sophisticated surgical performance also matched those of the professionals (correlation coefficient 0.833). Crowd workers in this study were therefore considered to offer an effective, precise, and inexpensive way to assess surgical performance, even when applied to complex procedures. Although the conclusion is specific to the mentioned study, it must be noted that there are cases where expert knowledge is critical and needed. Crowdsourcing does not fit all situations as specific domain knowledge is critical for the accomplishment of certain tasks. As such, checks should be made to ensure crowd worker's performance is acceptable.

It can be difficult, time-consuming, and expensive to successfully recruit cancer survivors for psycho-oncology research, especially for subgroups like young adult cancer survivors. An option for locating cancer populations is to use online crowdsourcing marketplaces. Arch and Carr (2017) conducted a study to evaluate the presence of cancer survivors on MTurk and the viability of MTurk as an effective, affordable, and trustworthy platform for recruiting and conducting psycho-oncology research. Compared to many other available psycho-oncology recruitment sources, they discovered that the recruited individuals were a more geographically, medically, and socio-demographically diverse group of cancer survivors. MTurk was time- and cost-effective, and the data produced were reliable.

Crowdsourcing marketplaces can and increasingly are being used as subject pools to conduct research. These marketplaces have been valuable to research and have

been used extensively in research in fields such as food science (Simmonds, Woods, & Spence, 2018), economics (Marreiros, Tonin, Vlassopoulos, & Schraefel, 2017), psychology (Callan, Kim, Gheorghiu, & Matthews, 2017), and mental health (Jones, Du, Panattoni, & Henrikson, 2019; Meeussen & Van Laar, 2018).

We relied on CS marketplaces as a human subject pool to recruit participants for all the studies in this thesis. We used two major CS marketplaces in the studies, Prolific and Amazon Mechanical Turk (MTurk). These marketplaces offer the opportunity to access a large pool of users and produces results with high validity, even in empirical research (Brown et al., 2018; Pedersen et al., 2013; Strickland & Stoops, 2018).

2.3　　Data reuse ethics

Reusing one's data in a way that deviates from the original purpose for which the data were collected raises ethical concerns regarding reciprocity and fair benefit sharing (Ballantyne, 2019). What further complicates issues is that health data today transcend clinical records and have expanded to include lifestyle, wellness, and health-related behaviour data captured by people's personal digital devices and applications such as wearables, smartphones, social media and even loyalty cards. Such observed data are under the control of device manufacturers and application developers, often giving them the legal right to use, share, or sell such data at their discretion (Ballantyne, 2019; Kotut, Stelter, Horning, & McCrickard, 2020).

To give a truly informed consent today, when software algorithms can mine data for even presently unknown anomalies, a shift in the perspective of data management of a more human-centric perspective is needed (Swan, 2012). Involving data producers in debating ethical issues is warranted, and this approach is employed throughout this research, especially concerning the reuse of their data in the future. Krutzinna, Taddeo, and Floridi (2019) argue that it is an ethical failure not to reuse important data that matter for improving public health. They further argue that past data management practices are to blame, and that people's personal data should be made available for reuse particularly in scientific settings beyond people's lifetime. This can be done by encouraging people to donate their data posthumously, similarly to how human organs are currently donated.

Calls have been made for data producers to take "ownership" of their data so they can understand the value the data present and the need to take action against threats such as identity theft (Zou & Schaub, 2018). Today, data producers are offered a take it or leave it opportunity to use mobile applications through consent forms and user permission. One is often compelled to grant permission that may not be required for the

use of the application. Refusing to grant such permission may lead to being unable to use the application, forcing users to compromise on the value of the data being collected by these applications and exchange them for the use of the application (Bahrini, Wenig, Meissner, Sohr, & Malaka, 2019). Martani, Shaw, and Elger (2019) investigated how health insurers' mobile applications encouraged customers to share their personal behavioural and health data in exchange for monetary rewards. Even within the research domain, the sensitive and personal nature of health data presents many ethical and legal challenges. This echoes the need for more attention to addressing the regulatory and ethical issues of personal data (Andanda, 2020; Kostkova, 2018).

Throughout this thesis, we focus on the reuse of people's personal data for various use cases and investigate the perceptions of data producers of the use and reuse of their data.

2.4 Investigating user perceptions with research probes

As technology becomes increasingly ubiquitous, its usage tends to be task- and work-related. Understanding the fundamental driving forces that influence people's opinions is crucial to determining the values (social, hedonistic, emotional, financial, etc.) they attach to different things. Karat, Karat, and Vergo (2004) contend that understanding people and their values in life might help to better understand how to meet their requirements in relation to the increasingly diverse ways in which they use technology.

Closely related to this is the study by Crawford and Renaud (2014), which investigated people's perceptions of transparent authentication, and the sensitivity of tasks and data on smart devices. Using a custom built mobile application, they also explored user perceptions of the levels of protection provided to their data and applications on smart devices. They identified varying opinions from participants about the experiment tasks, justifying the need for a more granular smartphone security method. They also found that one commonly cited goal in authentication research, the complete removal of security barriers, was not aligned with user perceptions, as users preferred to have some barriers in place. With mobile computing having expanded beyond the confines of laptops, tablets, and smartphones, Häkkilä, Vahabpour, Colley, Väyrynen, and Koskela (2015) conducted a study to explore design directions for smart glasses based on user research grounded use cases. They developed three non-functional smart glasses as research probes to allow study participants to roleplay imaginary situations and then share their opinions. User perceptions of the usefulness of smart glasses and privacy and social concerns were then elicited to inform future design.

Similar studies have been conducted within the health domain. Opoku Asare, Visuri, Vega, and Ferreira (2022) conducted an exploratory investigation into using smartphones to detect the onset of depression. They investigated the possibility of using unsupervised anomaly detection methods to monitor the fluctuations of mental health and its severity using the *Me* mobile application. The Me app was used to collect both subjective and objective data, which were then combined with anomaly detection methods to identify digital biomarkers for depression. Post-study feedback on user behaviour through a semi-structured interview highlighted an expectation of instant actionable feedback within the app, which was personalised, interpretable, and meaningful to the user. Schoen et al. (2017) also conducted a study to gain an understanding of the benefits of and possible barriers to the use of mHealth tools for community health workers. They used the *Geohealth* software tool to collect health and demographic data of community members, assessing their experiences of using the software tool, and conducted a semi-structured interviews with 57 community health workers. They posited that while the tool saved time with paperwork, organising data that needed to be collected, and replaced sheaves of paper, poor hardware, faulty software, and an overall negative perception of community members of mHealth solutions were key technical and social barriers to the successful adoption of the m-Health tool. They demanded that community health workers' opinions be sought when designing the tool to maximise its usefulness.

In summary, research probes have been deployed in different settings and contexts to collect user perceptions of different use cases. In this thesis, we develop software probes (Articles I and V) to aid the soliciting of user perceptions of various personal data management use cases.

3 Research methods

The research work embodied in this thesis was conducted using a mixed-methods approach. Mixed-methods is a mix of quantitative and qualitative research methods that is common within the HCI research community. This section provides an overview of the research methods used in the different articles in this thesis.

3.1 Summary of research methods used

In the studies mentioned in this thesis, the mixed methods approach was largely used. Quantitative data were collected from the study participants with supplementary qualitative data to gather more nuanced insights. Table 3 summarises the research methods used in the studies presented in this thesis.

3.2 Research approach

Understanding user perceptions of and attitudes towards personal data is an exploratory and interdisciplinary process. Drawing on research methods from the fields of psychology, sociology, and computer science, this thesis is conducted in the interdisciplinary research field of Human-Computer Interaction (HCI), a field that cuts across the previously mentioned disciplines Lazar, Feng, and Hochheiser (2017). The research in this thesis is primarily exploratory in nature and investigates the perceptions of people of their personal data. This thesis yields both artefact and empirical contributions and draws on mixed methods to explore user perceptions of the value, ownership, ethics, and reuse of personal data through the lens of MyData. The following research methods are used in the conduct of this research.

Two methods (field studies and online surveys) were used to collect data from the study participants in the various studies included in this thesis. Quotations from participants are highlighted in italics. Articles I-V elicited information from participants using online surveys. In some studies, the online surveys provided supplementary data to data collected through software artefacts. The online surveys included both open-ended and quantitative items.

Table 3. Summary of research methods used in each article.

Method	Study design	Data collection	Data analysis	Software used
Article I	Field study, online experiment	Public display, online questionnaire	Video analysis with OpenCV, thematic analysis, statistical tests	Custom Public display app, Google Forms
Article II	Online experiment	Online questionnaire	Thematic analysis, statistical tests	Google Forms
Article III	Online experiment	Online questionnaire	Thematic analysis	Custom decision support tool, Google Forms
Article IV	Online experiment	Online questionnaire	Directed content analysis, statistical tests	Google Forms
Article V	Field study, online experiment	Custom mobile app, online questionnaire, Vickrey auction	Statistical tests, Borda count, thematic analysis	Custom android mobile app, Google Forms

3.2.1   Online experiments

In all the studies, participants completed tasks in their own environment without researcher supervision. In Article II, we developed and deployed a web-based instrument in a crowdsourcing marketplace to present different data management scenarios to crowd workers in an online within-subject experiment. In Article III, we employed a remote study setup, a web-based software tool, with which participants interacted with on their own computers. Participants subsequently completed a post-task questionnaire, hosted on Google Forms to share their thoughts on various items.

3.2.2 *Field studies*

Articles I and V employed field studies. Field studies are a way of obtaining empirical data in the user's own environment. In the case of Article I, we deployed an interactive public display artefact at our local university. Participants used the public display tool to voluntarily donate personal data, including media data. In Article V, we developed and deployed an Android mobile software artefact which we shared with participants sourced from Prolific, a crowdsourcing marketplace.

3.3   Data analysis

We employed the use of a mix of qualitative and quantitative methods. We used quantitative methods that are reminiscent in the field of HCI. Statistical tests such as parametric tests (e.g. t-test, CHI-square) and non-parametric tests (e.g. Wilcoxon rank sum) were used. Thematic analysis (Braun & Clarke, 2006) was used to develop insights from the qualitative data. Video data were analysed using the SSD-framework (Liu et al., 2016) and the ResNet (He, Zhang, Ren, & Sun, 2016) deep learning model.

3.4   Research ethics

The ethical procedures required by the thesis author's university were duly followed during the conduct of this thesis. The author, alongside his co-authors, was effectively engaged in following good scientific practices in accordance with the Declaration of Helsinki, the national ethical guidelines of Finland, and the European Charter for Researchers. The research conducted was committed to the obligations of research such as openness, quality, and accountability (Mikkelsen et al., 2016). Software code and other study-related materials where applicable were made available in online repositories. Research study participants were engaged in all the studies embodied in this thesis, and the thesis author's university's ethical processes for the engagement of such persons for the purposes of research were followed. Participants were required to provide consent prior to participating in each of the studies. Study participants therefore consented to their data being used for the purposes of academic studies. We also recruited participants from the online crowdsourcing marketplaces Prolific and Amazon Mechanical Turk. The crowd workers from these marketplaces who participated in our studies were compensated according to the principles of fair crowd work (Silberman et al., 2018).

3.5 Experiments

3.5.1 *Study 1: Videosourcing personal health data (Article I)*

In this study, we deployed a *Videosourcing* application to be used by passers-by (students, researchers, and other administrative staff) on the university campus. Study 1 was conducted in two streams, studies A and B. In Study A, we collected data via the Videosourcing application interface and some supplementary qualitative data via a follow-up online questionnaire hosted on Google Forms. In Study B, we recruited study participants from Prolific to provide their thoughts on the issues identified in the earlier study (Study A). Here, we employed the use of narrative transportation (Green, 2008), in which we immersed participants in a story about the tool and its use. Narrative transportation is a theory that proposes that when participants are engaged in a story, their attitudes and intentions change to reflect that story (Green, 2008). The collected video data were analysed using OpenCV (Bradski, 2000). The quantitative data were analysed using the Kruskal-Wallis and Wilcoxon Signed Ranked statistical tests.

3.5.2 *Study 2: Data management scenarios (Article II)*

In this study, we deployed an online data collection tool on Amazon Mechanical Turk (MTurk). The tool was built using AWS Crowd-HTML markup and was directly deployed in the MTurk interface as a survey task (see Figure 4). The list of parameters used to generate the scenarios in the tool was uploaded via a separate CSV file. We analysed the qualitative data submitted by participants using thematic analysis (Braun & Clarke, 2006). Least Square Means (LSM) with Tukey HSD adjustment was used to investigate the pairwise relations of the levels of each parameter. Shapiro-Wilk and one-way ANOVA tests were used to test statistically significant differences between the parameters.

3.5.3 *Study 3: Reuse of open personal data (Article III)*

Prior to study 3, we had conducted a study in which we invited participants to contribute and assess mental health self-care techniques using a data collection and decision support tool. During the data collection, the participants were informed that all the data they provided would be used freely for research and as an open dataset accessible by anyone online, thus open health data (OHD). We used Prolific as our human subject pool. The questionnaire was hosted on Google Forms. Participants were given a link that directed users to explore the decision support system (software) that had been built

32

using their previously donated data. After interacting with the developed system, users answered the questionnaire. Thematic analysis was employed for the data analysis (Braun & Clarke, 2006) following a deductive approach.

3.5.4   Study 4: Evaluation of mHealth use (Article IV)

In Study 4, we investigated how mHealth solutions were perceived by two groups: users and non-users. Our goal was to understand the expectations and assumptions underlying mHealth non-users' concerns about the technology, as well as to evaluate how far these assumptions were from users' actual experiences. We conducted a questionnaire-based study hosted on Google Forms and deployed on Prolific, targeting people with varying experiences of chronic pain. We applied the directed content analysis method (Hsieh & Shannon, 2005) to analyse the qualitative data. The Mann-Witney-Wilcoxon statistical test was also used for the Likert-type questions.

3.5.5   Study 5: Assessing monetary value of personal health data (Article V)

In this study, we built a custom mobile application to collect health data including sleep duration, sleep quality, pain intensity, and wake-up times from people living with Low Back Pain (LBP). These data were then later valued to be sold to a purchasing entity (a for-profit company, the government, or an academic institution) using a reverse second-price auction mechanism implemented in the app. We applied the reverse second-price closed-bid Vickrey auction (Vickrey, 1962). An end-of-study questionnaire hosted on Google Forms was sent to the participants at the end of the 14 day study to gather qualitative insights. The quantitative data were analysed using statistical tests (e.g. Shapiro-Wilk, Kruskal-Wallis, one-way ANOVA, Bonferroni correction, etc). *Borda count* (Saari, 2000) was also used to weight user preferences. Qualitative data were analysed using the deductive thematic analysis approach (Braun & Clarke, 2006).

4   Research contributions

In this chapter, we present the research contributions from Articles I-V in which we explore and investigate the views of people concerning various issues that concern their personal data. We explore this from the angles of data ownership and privacy, personal data management and reuse, and valuation.

4.1   Ownership and privacy of personal health data

Article I introduces Study 1 with two parts (Study A and Study B) in which the participants provided insights into using public display technologies to collect personal data in public settings in the form of tagged media. In Study A, participants used the designed public display tool (see Fig. 1) to donate selfie videos tagged with metadata to

Fig. 1. The _Videosourcing_ desk deployed in the corridor of the campus. Reprinted with permission from Article I under the terms of the Creative Commons Attribution 4.0 licence, 2020.

be leveraged in any non-profit research. In Study B, an online narrative transportation study was used to elicit rich qualitative insights into key emerging aspects from Study A on issues of trust in the data collection entity, the intended use of the collected data, and the reasons for the collection of the data. Since in Study A we obtained input only from people who already used the deployment, Study B targeted people online who did not use the deployment. As such the use of a narrative transportation was warranted. Narrative transportation enables participants to engage in a story, which primes their attitudes and intentions to reflect the circumstances of the given story narrative (van Laer, de Ruyter, Visconti, & Wetzels, 2013). The narrative presents the setup in its entirety to the study participant along with a narrative of the scenario that participants in the first study experienced. The narrative described how the user first encounters the display, interacts with, etc. The field study (Study A) showed people's willingness to donate even highly sensitive personal data about themselves in public settings. A total of 199 15-second video selfies was donated over a 61-day period with 78 of them having voluntarily labelled metadata attached to them by the donors. The online narrative transportation study (Study B) provides deeper understanding of various issues arising from Study A that can be leveraged in the future design of systems such as 'Videosourcing' for collecting data in public. Study B revealed interesting perceptions of people regarding the sharing of their sensitive data in public settings, reservations about data security, and ownership rights to their personal data.

Participants in Study A reported a mean truthfulness score of 5.4 (SD = 0.8) on a scale of 1 to 7 for their answers to the questions used as metadata to tag the videos with no significant differences between genders or age groups. Using the Kruskal-Wallis test in Study B, we identified a significant difference between the willingness to donate different types of data ($p < 0.01$). Further analysis using the Wilcoxon Signed Rank tests showed that data types such as religion and racial identity were more likely to be donated than email, location, serious health conditions, or one's surname ($p < 0.01$). The figure 2 shows fairly stark differences in the data types. Data types such as surname and serious health conditions were types participants were unwilling to reveal. Non-serious health conditions and email addresses were more likely to be revealed than surnames ($p < 0.01$).

The results from Article I reveal the participants in general were willing to share their personal data, even in public settings. However, concerning the entity collecting the data, the participants highlighted some differentiation. To them, not all entities were equal and that the reputation of an entity played an important role in their decision to donate data. Even in academia, participants viewed private and public research institutions differently, with participants more easy-going with public research institutions.

Type of Data Participants Would Donate

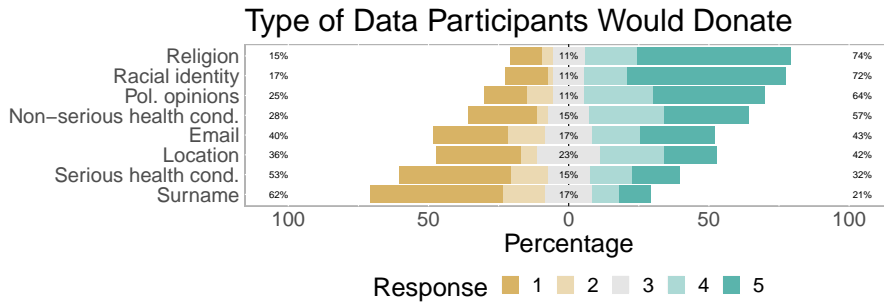| Category	1 (%)	3 (%)	5 (%)
Religion	15%	11%	74%
Racial identity	17%	11%	72%
Pol. opinions	25%	11%	64%
Non–serious health cond.	28%	15%	57%
Email	40%	17%	43%
Location	36%	23%	42%
Serious health cond.	53%	15%	32%
Surname	62%	17%	21%

Response ■ 1 ■ 2 ■ 3 ■ 4 ■ 5

Fig. 2. Responses to a Likert Scale on the types of data the participants would be willing to donate (1 = not at all willing, 5 = extremely willing). Reprinted with permission from Article I under the terms of the Creative Commons Attribution 4.0 licence, 2020.

In Articles I and V, the participants expressed the need to be in charge of making decisions about their data. To them, owning and being in control of one's data was a right, allowing them to choose who should have access to their data, what to share, and the opportunity to benefit financially from their data. Some participants also highlighted the concern that they could lose ownership of their data once they donated it. This was particularly important to the participants, as they wanted to be in control of how their data were used and managed. Most participants agreed that their data should only be used for the defined purpose.

In Article I, although the participants were willing to donate different types of data, they still expressed some reservations regarding privacy, trust, and the future use of their data. The participants also clearly demonstrated that internally, they valued different types of data differently, as can be seen in Figure 2. They assigned sensitivity levels to their data, willing to donate less sensitive data and unwilling to donate more sensitive data. Considering the sensitivity of health data, the participants expressed their willingness to donate more only if the setup was anonymous. Since the setup in Article I involved video collection, it might have had affected the kinds of data participants were willing to donate. However, in general, the majority of participants were willing to donate some data. Despite not probing participants about the misuse of their data, this issue was raised repeatedly by participants, indicating a clear pain point for people.

In both Studies A and B of Article I, people were willing to donate personal data in public settings using public display technologies, even when rewards are not guaranteed. The use of PDs could act as a complementary form of data collection to the existing tried and tested traditional methods such as online data collection systems. Further, should concerns such as trust in the organisation behind the data collection, the intended use of the collected data, and the security of the data collection platform be alleviated,

people were willing to provide much more detailed and sensitive information about themselves. In explaining the intricacies of data collection, "why" is more important than "what" (Fiesler & Hallinan, 2018). People should be aware of "why" their data is being collected and this should be addressed proactively at the outset of the user interaction flow. This would increase participation in data collection initiatives that adopt PDs, reduce concerns about data use, and improve the ethicality of the data collection procedure. For more details on Studies A and B, see Article I.

4.1.1 *Privacy and data management in using m-Health solutions*

Mobile health (mHealth) is a medical or public practice supported by mobile devices (Rowland, Fitzgerald, Holme, Powell, & McGregor, 2020). It contains a variety of contexts such as the use of mobile phones for care delivery, patient communication, medication monitoring, and even adherence support (Becker et al., 2014). In Article IV, we focus on the perceptions of people with chronic pain of mHealth. Specifically, in this thesis, we focus on the privacy and data management aspects of the study. The study was conducted with two groups: those who use mHealth solutions and those who did not. We investigated how mHealth solutions were used and experienced from the perspectives of ease of use, reliability, functionality, usefulness, satisfaction, and privacy using questionnaire responses from 62 participants: 31 m-Health users and 31 non-users.

The fear of data abuse and/or misuse was a strong concern for participants. Real world events such as the increasing number of data breaches, particularly of health-related data, influenced on people's willingness to donate their data. In Article IV, we observed that mHealth non-users were more concerned about the privacy of their data than m-Health users shown in Figure 3 ($p < 0.01$). Such privacy concerns acted as inhibitors to the adoption of technologies that could support health and well-being. Participants expressed very serious reservations about sharing sensitive data such as credit card numbers and biometric data citing identity theft as a major fear. Similarly, concerning mHealth solutions, manufacturers having access to users' personal data collected using mHealth solutions, we observed statistically significant differences between mHealth users and non-users ($p < 0.01$).
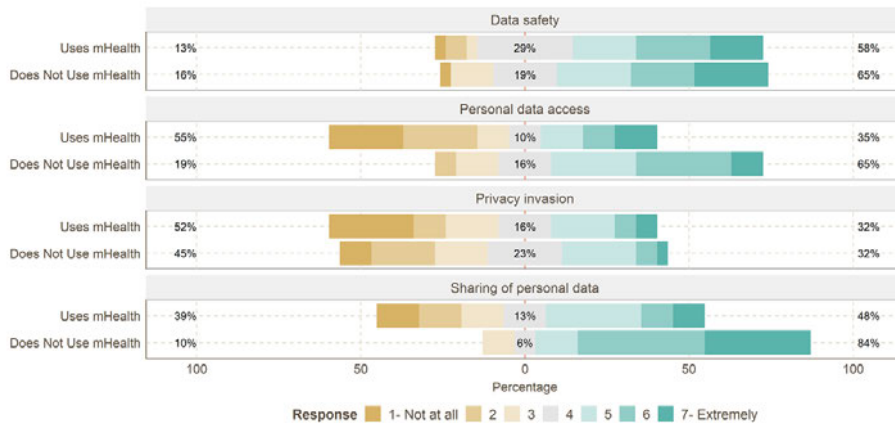
Fig. 3. Likert item answers to privacy. mHealth: mobile health. Reprinted with permission from Article I under the terms of the Creative Commons Attribution 4.0 licence, 2022.

While expressing concern about privacy, the participants also demonstrated diverse views concerning the future management of their personal health data. In Article II, the participants claimed they would not trust platforms with their data, but also preferred to have their health data linked to their health insurance provider. Similarly, in Article IV, the participants were unconcerned about the m-Health solutions provider having access to their data. However, in Articles II and III, they were unreservedly concerned. This contrast in opinions may very well be due to a lack of clarity on what privacy actually means. People have a natural expectation of the flow of information. Privacy issues may surface if this fails to happen. Participants were concerned about the sharing of their personal data with third parties, and mHealth solutions frequently fell short of giving any guarantees despite gathering and sending personal data (Blenner et al., 2016; Grindrod et al., 2017). Personal data are sensitive by nature, making access to them a threat to data security (Tangari, Ikram, Ijaz, Kaafar, & Berkovsky, 2021).

In summary, concern for the security of one's personal health data is one of the impediments to the adoption of digital tools that could help improve people's overall health. We suggest that privacy concerns among non-users of mHealth may have hindered their adoption of mHealth solutions. Conversely, it could also be argued that their prior experience or lack thereof with mHealth solutions has made them more concerned about data privacy as sometimes, only after people use a technology do they realize that their initial concerns may have been exaggerated. Whatever the case is, we believe that identifying and removing impediments that may be crucial to mHealth adoption could be critical to healthy lifestyle promotion and even self-care. A "neutral"

entity that commands user trust could be helpful in this regard by taking responsibility for the managing of personal data to curtail potential misuse and abuse.

4.2   Reactions to personal data management issues

Building on key findings of Article I, we designed a further study to investigate people's reactions to various data management scenarios. The aim was to better understand what people considered acceptable when it comes to managing their personal data. Article II focuses on people's perceptions of various data management practices, modelled after the related literature (Fiesler & Hallinan, 2018; Stevens, 2001) and various privacy and potentially ethical violations of people's data in the real world (Harwell, 2019; Helm, 2019; Lapaire, 2018; Marcia, 2019). The research was conducted in two stages (Stage 1 and Stage 2). In Stage 1, participants assessed the ethicality of various data management scenarios using a novel quantitative scenario-based approach, as shown in Figure 4 and parameters shown in Table 4. In Stage 2, there was an online questionnaire to collect demographic data, as well as open-ended questions to collect additional qualitative insights to support the quantitative results from Stage 1. The research used in Article II was conducted with a special focus on MyData (Poikola et al., 2020), a human-centric data management model and set of guidelines that aims to empower individuals to access, use, manage, and give permission to their personal data. Using a variety of different data management scenarios as experimental stimuli, we collect data on ethics and concern on 96 different scenarios that emulate common types of personal data use and misuse. A total of 1920 individual scenario assessments was obtained in Stage 1.

A Cumulative Link Mixed Model fitted with the Laplace approximation regression model was built using the scenario assessments from Stage 1. We observed statistically significant effects of the different parameter levels on the perceived ethicality of a given scenario. Further, a least square means analysis with Tukey HSD adjustment was used to investigate the pairwise relations of the levels of each parameter. Similarly, significant statistical differences were discovered between all the parameters (see Table 4). For more details on how different parameters affect the perceived ethicality of the scenarios, see Article II.

Further, the study also shows how different parameters affected the extent to which the presented scenarios were perceived as concerning. The results however, differ substantially from those of the ethicality of the scenario, indicating that the two variables (ethicality and felt-concern) are decoupled. A least square means analysis comparing pairwise the different parameter levels in the context of how concerning the conduct was

Evaluate how ethical or unethical this scenario is

Certain terms that you're likely to see are defined below:

1. **MyData** is data you have generated about yourself and to which you have access, control, and ownership. This includes data collected by your mobile devices (mobile phones and wearables), wellness applications, medical records, employment records, etc.

2. **MyData operator** is a company that stores your MyData. MyData is collected from multiple sources and saved on a server through which you can access and control your data.

In this task, a scenario depicting how your **MyData** is treated and acted upon by your MyData operator and a third party entity is presented. The scenario includes 1) action taken by the MyData operator, 2) type of data used, 3) the purpose, and 4) a reason for the event to occur. Based on these, you will be asked to rate how ethical or unethical the scenario is.

Scenario:
${HIT_operator_action}. This data includes your ${HIT_datatype}. ${HIT_purpose}. ${HIT_whyhappening}.

Details:
MyData operator action: ${HIT_operator_action}

Data type: ${HIT_datatype}

Purpose: ${HIT_purpose}

Why did this happen: ${HIT_whyhappening}

Your response:
How ethical was the conduct in the scenario above Use a scale from not at all ethical (value 1) to extremely ethical (value 5):

1: not at all **ethical**

5: extremely **ethical**

Which of the following **affected most your choice of ethicality?**

○ MyData operator action

○ Data type

○ Purpose

○ Why did this happen

How concerning was the conduct in the scenario above Use a scale from not at all concerning (value 1) to extremely concerning (value 5):

1: not at all **concerning**

5: extremely **concerning**

Submit

Fig. 4. The Human Intelligence Task (HIT) deployed to MTurk. The variables denoted inside curly brackets were populated in runtime by MTurk with the parameter levels presented in Table 4, i.e. there were 96 unique scenarios deployed through this setup. Participants used the two sliders and the set of radiobuttons to indicate the perceived levels of ethicality and concern, and the decisive parameter for ethicality. Reprinted with permission from Article II ©️ ACM, 2021.

41

Table 4. Parameters (column headers) and levels (rows) in the study. Unique combinations of levels (N=96) constitute the assessed data management scenarios. Adapted with permission from Article II © ACM, 2021.

MyData operator action	Data type	Purpose	Consent
The MyData operator has shared your MyData with a third party without getting paid for it	Diagnostic and medical data (diseases, medical history, medicine prescription)	Your MyData is used by academic researchers to conduct publicly available research	The action was taken against your consent, i.e. you did not agree to this happening
The MyData operator has sold your MyData to a third party and profited from the transaction	Personal health tracker data (sleep, heart rate, daily activities)	Your MyData is used by for-profit companies to build proprietary commercial solutions.	The action taken was within your consent, i.e. you agreed to this happening
	Location data	Your MyData is used by the government to implement policy changes related to healthcare	The action taken was within your signed consent but the terms and conditions were confusing, i.e. you were unaware of having granted such permission
	Personal media files (pictures and video)		The action was taken against your signed consent, but legal agreements in your country allowed for the actions to take place

perceived as also yielded statistically significant differences between the levels. Details of the pairwise comparison can be found in Article II.

A Shapiro-Wilk test on normality on the ethicality and concern responses was not normally distributed ($p < .0005$). The response variance for ethicality was 1.88 (SD = 1.37), while the variance for concern was 1.15 (SD = 1.37). When considering the variance across all 96 scenarios, one-way analysis of variance (ANOVA) showed significant differences on both the ethicality response (F = 1.539) and the concern response (F = 1.931). ANOVA was used due to its tolerance of non-normal data with a small effect on the Type I error rate. This verifies that all the different scenario types had an influence on the participant responses.

Stage 2 was conducted with the participation of 75% of the invited participants of the Stage 1 study. Here, we collected qualitative insights to complement the first stage study. The MyData vision was an exciting concept for our participants, as it enabled them to be the owners and controllers of their data. We observed that the prospect of being in control of one's data was not only exciting but also offered an opportunity to monetise such data. However, concerns about privacy remained, as participants expressed doubt about whether it was even possible to be in control of one's data in a world in which security and privacy were potentially in the hands of malicious actors and hackers. Some were concerned about how much control they could realistically have. However, the attitude of participants was generally positive, especially concerning healthcare data ownership and management and its vast potential.

Our participants were very vocal regarding the abuse of personal data, expressing in uncertain terms how they perceived their data being abused on a daily basis by corporations for monetary benefits. Some of the issues highlighted issues include the sale of their data to third parties for marketing purposes and data breaches at a global level. Our participants not only voiced concern but positively embraced the idea posited by the MyData vision and guidelines, not only in offering the opportunity to be in control of their data but also to provide access to tackling healthcare problems by enabling access to their health-related data to healthcare providers.

Most ethical issues about data are related to data collected about people - personal data (Hand, 2018). Considering how parameters in Stage 1 affected the perceived ethicality and concern, it is clear that data management perceptions should be studied at a more granular level. People do not view ethicality and concern in the same way. What one may consider unethical may not necessarily therefore be concerning and vice versa. Table 5 gives a clue about what aspects of the parameters are important to people. As such, prioritising these aspects which people consider most decisive in their internal accounting could be useful in improving the legibility of documents such as

disclaimers and consent forms. The results from Stage 2 also highlight a level of distrust of companies involved in the collection and processing of personal data. To this end, trust is an essential component that must be developed in all the stakeholders in the data management chain.

Table 5. Counts of how many times each parameter was the decisive factor in the participant's choice of ethicality, and the counts of levels that were shown if the parameter in question was the decisive factor. Adapted with permission from Article II © ACM, 2021.

Parameter	Level counts and proportions as the decisive factor for ethicality
MyData operator action (499 / 26.0%)	Sold (282 / 56.5%) Shared (217 / 43.5%)
Data type (691 / 36.0%)	Media (229 / 33.1%) Diagnostic (184 / 26.6%) Health tracker (139 / 20.1%) Location data (139 / 20.1%)
Purpose (435 / 22.7%)	Government (152 / 34.9%) For-profit (143 / 32.9%) Academic research (140 / 32.2%)
Consent (295 / 15.3%)	Against consent (84 / 28.5%) Within consent (77 / 26.1%) Within but confusing (72 / 24.4%) Against but legal (62 / 21.0%)

In summary, we present end-user perceptions on personal data management through the examination of various data management scenarios. The perspectives elicited were realistic given participants' knowledge and/or experience with various data mismanagement occurrences in the real world. We posit that end-user perceptions of personal data are relevant to scientific, societal, and industrial research on data ethics. To support research in this area, we contribute a survey-based method and instrument for the assessment of data management perceptions.

4.3   Reusing open personal health data

Our study in Article III focused on the perceptions of data contributors concerning the use and reuse of their health data contributed to Open Data initiatives. Article II highlights users' opposition to the malicious use of their data without their consent. This study aimed at gaining an understanding of people's view of their personal data beyond its initial purpose in a situation where they have renounced ownership of the donated data. Open Health Data (OHD) as used within the context of this study refers to any type of publicly accessible health-related data (Kostkova et al., 2016; Martin, Helbig, & Birkhead, 2015). Researchers have explored crowdsourcing as a means of collecting bespoke OHD as an input for digital health software solutions. Understanding the decision making process of potential data donors is crucial, as such understanding can aid the software community in taking ample steps to alleviate any concern.

We conducted a study (not included in this thesis) in which crowd workers from Prolific were invited to contribute data to a decision support system on mental health self-care. These participants were informed that all the data they provided would be used openly in research and made available online to anyone as an OHD dataset. Participants were requested to assess an initial set of self-care techniques based on a set of criteria (see Figure 5[B]). Participants could build on these initial set of techniques by contributing their own techniques which then became available for assessment by other participants. A total of 1071 people participated in the study with 188 unique self-care technique contributions. In Article III, a subset of these participants (N = 80) were invited to a follow-up study to explore the decision support system which could analyse their previously contributed self-care techniques and turn it into an interactive exploration interface helpful in finding suitable self-care techniques to try (see Figure 5). Participants then answered an online questionnaire about their perceptions of OHD reuse.

Anonymity (87.5%, N = 70) and *how the data would eventually be used* (88.75%, N = 71) were the most important considerations of study participants when deciding to donate data for public use. Of the 80 participants, only 37.5% (N = 30) considered the perceived societal benefit as an important criterion during the decision-making process. Thus, despite donating the data for public "benefit", participants put themselves first, not in terms of benefit per se, but the security and privacy of their identity and data.

We evaluated how much trust participants have in public, private, or societal stakeholders in building software using OHD. As expected, public (50%, N = 40) and societal stakeholders (48.75%, N = 39) were considered more trustworthy by participants than private stakeholders (17.5%, N = 14). See Figure 6. Similarly, in

Find self-care techniques for mental health in higher education.

Use one or more of the sliders to indicate your preferences. **A**

Use the sliders to indicate what type of self-care techniques are you looking for? Please try at least 3-4 different searches. Click 'discover best matches' to see a list of best-matching techniques, which are all based on the data you also helped donate earlier.

The list is in ranked order; The higher up a technique shows up, the better match it is to the search condition you set up with the sliders

B

Familiarity: 87
How familiar are you personally with this method?

Not at all familiar — Extremely familiar

Effectiveness: 70
Is the technique effective?

Not at all effective — Extremely effective

Affordability: 28
In general, how affordable is this method for higher education students?

Not at all affordable — Extremely affordable

Required level of sociality: 42
How much social interaction or cooperation does this method require?

Not at all much — Extremely much

Time required to get started : 69
Does it take a long time to get started with this technique?

Not at all long — Extremely long

Ease of getting started: 44
How easy is it to just get started with trying out this method?

Not at all easy — Extremely easy

DISCOVER BEST MATCHES

RESET SLIDERS

C

Pets (animals)
You can snuggle or go out with a pet, depending on the situation.

Get a pet
Get a pet. Dog or cat or what do you like the most

Travel_Explore
Someone has to choose a few destinations the he/she would like to visit, but not for sun and beach tourist, and get to know that culture, history, some facts about the local language, explore places out of the beaten track without risking potential injury. Be an amateur explorer and giving time to enjoy what his doing, and not focusing on partying or buying souvenirs.

Seek professional help
Find your way to a mental health professional - there are specialists for pretty much everything from not being able to sort out adult life to suicidal ideation, and while waitlists are long the process can start from student healthcare

Taking care of a pet
Caring for and spending time with a pet helps to de-stress and can force you to get out of a rut and do something. Petting a cat is known to lower bloodpressure.

Working out in gym. | ☑
I felt that working out helps me ease the stress and pressure but I did not start it as a self care method. But it helped me quite a lot.

Lifting heavy at gym
Go to your gym and push yourself to your limits

Spending time with animals

Playing an instrument
One can combine writing about a stressful topic with the art of creating music. You can evaluate and work through your situation, with the help of your own music therapy.

Better diet
Eating more healthy

Fig. 5. Interface of the Decision Support System developed to suggest mental health self-care techniques to participants [A) General instructions B) Criteria for finding self-care techniques C) Suggested self-care techniques based on the selected criteria]. Reprinted with permission from Article III © IEEE, 2021.

Article V, participants considered public (government) stakeholders more trustworthy than private stakeholders (banks, and insurance companies). Participant perspective about challenges that OHD presents include preserving user privacy and anonymity, preventing abuse and misuse of health data, and ensuring quality of contributed data. Despite these challenges, the potential application areas of OHD are enormous. OHD could transform the development of digital health software, opening opportunities to
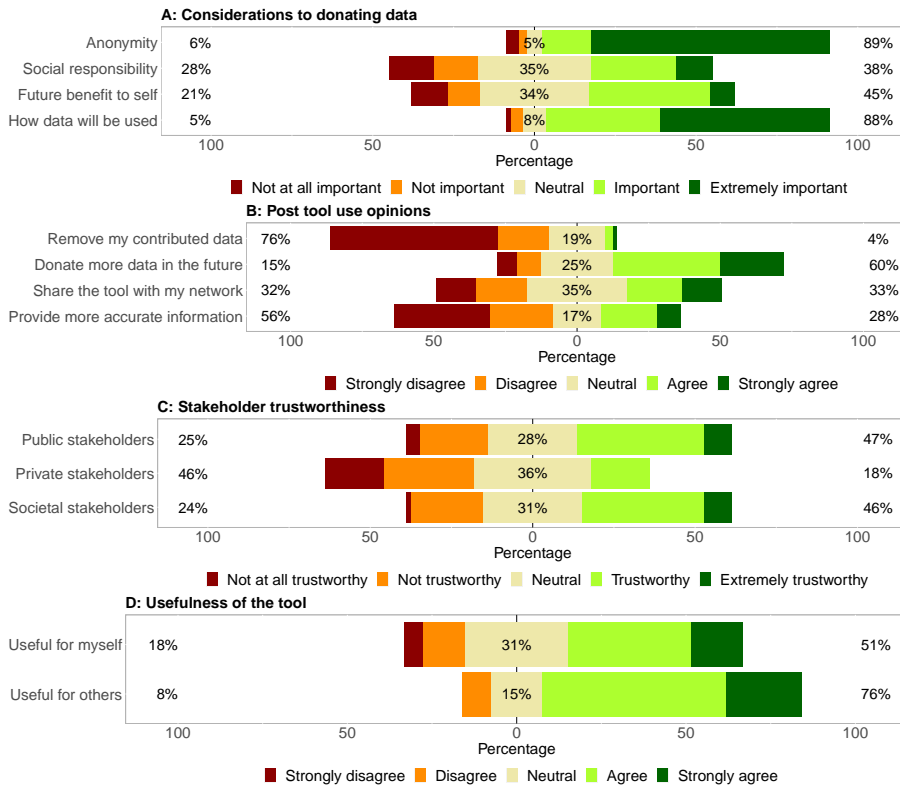
Fig. 6. The Likert-scale responses to questions concerning (a) aspects to consider prior to donating data, (b) user opinions about their data after using the designed tool, (c) stakeholder trustworthiness, and (d) the perceived usefulness of the tool. Adapted with permission from Article III © IEEE, 2021.

build tools and services for the improvement of the health and well-being of the general public, as well as pushing the boundaries of research.

Overall, participants perceive OHD as having very broad potential. It was considered suitable for the creation of health- and wellness-related software tools, the creation of new knowledge, fuelling of scientific research, and fostering the detection and prevention of previously unknown diseases (Bietz et al., 2016; Dolley, 2018). One promising development in this area is the interest of crowd workers in donating their health data to open data initiatives. However, there is a lingering concern about the prospects of private stakeholders such as pharmaceutical and insurance companies using people's personal data donated to OHD initiatives. Thus, despite relinquishing their right to the data, participants still wanted to have a say in how the data were eventually used, by whom, for what, and where. With trust in private stakeholders low, there is a fear about possible

data abuse by private entities, particularly insurance companies. Similarly, participants' strong position on anonymity presents a roadblock for healthcare solutions that may require identifiable user data (Martin et al., 2015). The capacity to examine the data and target-specific demographic groups to discover what conditions may be prevalent in these groups may be hampered by the deidentification of user data. An opportunity presented here is for public and societal stakeholders to join efforts in the development of software solutions based on OHD as they command more trust from data donors.

4.4 Monetary value of personal health data

In Article V, we focus on the perceived monetary value of personal health data and developed an Android application for collecting user data daily through an in-app questionnaire (see Figure 8). Article V was informed by previous findings that highlight users' inability to put a realistic price on their own data (Article I) as well as optimism surrounding the MyData concept enabling users to benefit financially from their personal data (Article II). Participants were recruited from Prolific, an online crowdsourcing platform after having met the eligibility criteria. Participants engaged in an auction once a day which lasted from 6am to 6pm. In each auction, participants were able to place bids on four units of data (sleep duration, sleep quality, pain intensity, and number of wake-up times) to be sold to one of three entities (a for-profit company, the government, or an academic institution). Each participant therefore had the option to place four bids a day. Participants were offered the opportunity to skip placing a bid on a unit of data if they wished. A total of 2656 bids was placed over the 14 day study period across all four categories, with an average of 165.3 bids per day (SD = 11.56). Fifteen participants won in at least one category (X = 17.93, SD = 17.18). We present density plots for the various data categories throughout the study in Figure 7 to demonstrate the variation in participants' assessments of the monetary value of their health data.

The auction data were not normally distributed according to a Shapiro-Wilk test on normality. A non-parametric Kruskal-Wallis test indicated a statistically significant difference in sleep duration ($\chi^2(2)$ = 10.994, p = 0.004), sleep quality ($\chi^2(2)$ = 8.455, p = 0.015), pain intensity ($\chi^2(2)$ = 9.469, p = 0.009) and number of wake-up time ($\chi^2(2)$ = 9.410, p = 0.094) scores between the different buyers. Using a one-way ANOVA, we investigated if bid distributions within a given data category showed any significant differences across the various entities (a for-profit company, the government, and an academic institution). We observed statistically significant differences in how the data categories were valued between the different buyers (p < 0.05), as shown in Table 6.
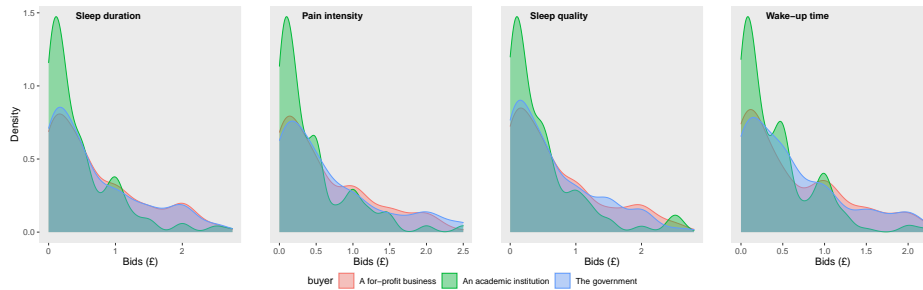
Fig. 7. Density plot of bids per buyer per data category. Reprinted from Article V (under review).

Bonferroni correction t-tests were conducted to determine the pair-wise differences between the buyers. More details on the pair-wise differences can be found in Article V.

Article V focused solely on individual data units (e.g. sleep duration for the past night). Compared to other studies (e.g. Carrascal, Riederer, Erramilli, Cherubini, and de Oliveira (2013); Staiano et al. (2014)), we recorded a lower monetary valuation of data units, an indication of a somewhat limited understanding of the value of such data to a purchasing entity. There is an opportunity for data purchasing firms to simply buy individual data units at a minimal cost for subsequent aggregation. On the other hand, by pricing individual units of data at a low fee, data producers may be setting themselves up for low financial benefits from their data. In a similar vein, the downward trend of personal data valuation favours purchasing entities to acquire more data at a lower cost. The low financial benefits to data producers could be discouraging in and of itself, prompting users to opt out of selling such data altogether.

The studies in this thesis indicate clearly that people are willing to share their data, albeit at a fair value (Li, Liu, & Motiwalla, 2021). In Article V, we learned that when it comes to selling one's data to entities such as the government and for-profit companies, people were very interested in getting monetary rewards in return. However,

Table 6. Summary of the one-way ANOVA test for differences in participant bids across the various purchasing entities. Reprinted from Article V (under review).

Groups	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Signif. code
Sleep duration - Buyer	2	6.130	3.065	7.460	0.001	**
Sleep quality - Buyer	2	4.110	2.055	5.033	0.007	**
Pain intensity - Buyer	2	4.980	2.493	6.794	0.001	**
Wake-up times - Buyer	2	4.450	2.223	7.354	0.001	**

Signif. codes: 0   '***'   0.001   '**'   0.01   '*'   0.05

for non-profits such as academic institutions, people were willing to even offer their data for "free", as long as it was used for the intended purpose. In this regard, users placed lower bids for the same units of data when the purchasing entity was "an academic institution". Contributing to the "public good" was one cited motivation for this attitude by our participants. Participants viewed non-profit organisations to be much more caring for the interests of the public. Regarding trust, our study in Articles III and V show very similar results; that people primarily trust themselves to be custodians of their own data, followed by societal stakeholders.

In summary, Articles I and V confirm an indication of a clear desire of people to benefit financially from their personal data, and that they are willing to sell such data, irrespective of who the purchasing entity is. With private companies already finding smart ways of accessing such data through consents and other means, the opportunity for data producers to benefit financially from their data is very welcome despite privacy and data protection concerns. The economic value placed on units of data is thus shaped by a mental image of the resources available to the purchasing entity. Our participants had self-selected as human subjects in academic studies, and our participants originate only from the UK. Therefore, although our results are in line with existing research and also indicate broader trends, they do not generalize over the general population. Also, as the data entered by our participants was questionnaire data, and despite the study relying on a commonly used mechansism, the Vickrey auction, the data may suffer from subpar ecological validity.

4.5   Technological tools

Throughout the conduct of this thesis, bespoke software tools and a research instrument were designed to conduct the studies highlighted in Table 2. An Android based mobile application was built and deployed on a tablet device to serve as a Public Display unit for the conduct of the studies in Article I. This tool allowed for custom-remote configuration of the app through push notifications, making it easy to modify the study parameters as needed, even while the deployment was ongoing. This tool also facilitated data collection which is crucial for analyzing the results of the study (see Figure 1). In Article V, a cross-platform mobile application was designed and built to run on both Android and iOS devices for the daily logging of health in-app questionnaire data and also for the conduct of in-app auctions of user data (see Figure 8). The tool did not only enable the easy and convenient daily logging of health-related in-app questionnaire data but also provided a platform for the fair and ethical auctioning of user data. Article II contributes a web-based instrument for the assessment of data management scenarios
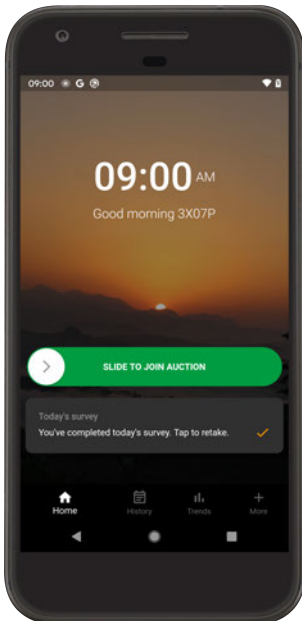
50

(see Figure 4). This tool provides a customisable framework for evaluating different data management strategies, thereby helping researchers make informed decisions.
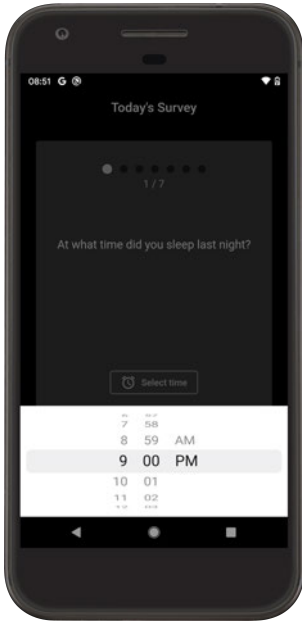
(a) Daily survey
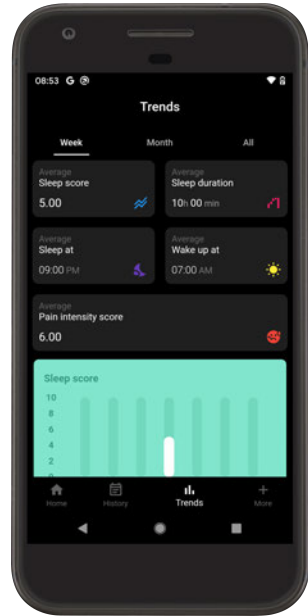
(b) Weekly trends

(c) Join auction

(d) Winnings

Fig. 8. Different screenshots of the 'LBP' app. Reprinted from Article V (under review).

5   Discussion

In this chapter, we revisit the research questions outlined at the beginning of the thesis and highlight how they were addressed. We discuss personal data management and its implications for digital services that rely on personal data. We conclude this chapter with some of the limitations reported in the articles presented in this thesis.

5.1   Attaining the objectives

This thesis has two primary objectives:
– O1: To develop new software and research probes for crowdsourcing personal data in the health domain.
– O2: To conduct case-studies on perceptions of privacy, ownership, and valuation of such personal data as well as the ethics of data reuse.

For objective O1, we created two software applications, *Videosourcing* (a public display app) and *LBP* (a mobile application) to collect personal data from participants in the wild. Both applications run on the Android platform. The Videosourcing application enabled the collection of short videos from passers-by with accompanying metadata. The second software tool, LBP, enabled the daily logging of pain and sleep data through an in-app questionnaire and facilitated the running of a daily auction in which users bid to sell their personal data to a purchasing entity.

The second objective, O2, was to conduct case studies on people's perceptions of various ethical and data management concerns related to their personal data. This was achieved in several areas, including privacy, data ownership, data reuse, and valuation. The case-studies focused on open health data (OHD), data donation in public settings using public display probes, an assessment of various data management practices, and the monetary valuation of health data. The LBP and Videousourcing applications were instrumental in the conduct of some of these studies.

Next, we discuss the research questions.

5.2   Revisiting the research questions

5.2.1   *RQ1 - How do human perceptions of ownership and ethics of use differ concerning personal health data of varying sensitivity levels?*

In Article V, we observe that users primarily trust themselves to be custodians of their own health data. As data producers, they want to remain in control of their data. Studies on personal data have shown that giving users more control of their personal data increases their trust and willingness to share it (Smith, Dinev, & Xu, 2011). Other studies such as (Li et al., 2021) share the opinion that users are willing to share their data if they are fairly financially rewarded.

Across Articles I-V, our participants perceive control of their personal data as a right. Having gained an understanding in Article II of the potential MyData offered in terms of ownership and control, they unreservedly supported the idea. People are not unwilling to share their data for the public good as evidenced in Articles I and III. However, they simply want to be in control of whom they share the data with, for how long, and for what purpose. For example, in Article IV, the participants agreed to share the ownership of their data with the platforms that hosted such data. However, when it comes to whom such data can be shared with and for what, the participants believe the prerogative is theirs alone. Efforts should be made to ensure concerns raised by people about issues such as privacy, data misuse, and abuse are duly addressed to facilitate future donation of their data.

In Article II, the participants expressed how important various parameters were to their decision-making about ethicality. We believe this inter-parameter importance provides a clue to data operators about which aspects (parameters and levels) are important to people. Although clear concrete data use policies are therefore already pervasively required, prioritisation of aspects people find most critical in their internal accounting can help increase user trust in consent forms and disclaimers, for example. One interesting observation from Article I is user concerns about "why" data about them is collected. While this may seem obvious, the ability to answer this simple question means participation in data donation initiatives can be actively increased, as participants begin to view the collection as more ethical and less suspicious. However, answering the "why" is less simple than it sounds, as it inherently involves identifying and addressing user concerns. More granular data management mechanisms must therefore exist in the future to guarantee people can dictate who will benefit from their data.

Our findings in this thesis indicate that participants value the sensitivity levels of their data. Throughout the various studies, the participants were unwilling to disclose

data they considered very sensitive. This included certain health issues, credit card information, and biometric data. Of particular concern to the participants was the unknown implications for donating such sensitive data. For example, in Article III, it was clear that the participants feared knowledge of certain diseases could negatively affect their access to insurance premiums. Similarly in Article I, the participants were unwilling to disclose data such as their surnames, serious health conditions and location data.

5.2.2   RQ2: What values (ethical and monetary) do people attach to their personal data?

We answer RQ2 in Articles I, II, and V. In Article II, we investigated people's perceptions of various data management scenarios. The participants expressed their felt-concerns about "malicious" practices being exercised by the corporations that currently managed their data. We observed that the seriousness people attached to the privacy and security of their personal data was such that simply informing them of the use of high-end technology to secure their data meant nothing to them. There was a need to build and maintain a trusting relationship with them.

To ensure a transparent relationship between data producers and data operators, trustworthiness must therefore be a core pillar demonstrated at every level of the relationship. Indeed, our studies reveal that what people considered unethical and what they felt concerned about were often not the same. This may be partly attributed to the degree of "online apathy" people experience, when although they know something is being done incorrectly, they simply do not care any longer as they consider it inevitable (Hargittai & Marwick, 2016).

Our findings from Article I indicate an underlying willingness of people to donate detailed personal data about themselves if they could 1) trust the entity behind the data collection effort, 2) trust the security of the platform where the data will be stored, and 3) know the intended use of the data. Indeed, on this front, it is evident that the reputation of the data collection entity is of the utmost importance. Considering entities that collect data - for example, academic research institutions - there seems to be a dichotomy between private and public institutions, something on which the participants focused. Similarly, in Article V, we observe an obvious dislike of insurance companies accessing user data. Indeed, insurance companies were singled out as being the least trustworthy by our participants.

Our findings indicate that the monetary value people attach to different units of their personal data has been decreasing over the years. For this, we hypothesise two

possible scenarios: 1) either people have grown tired and simply do not care any longer and would request a smaller amount of money for their data just so they could gain some financial benefit from it; or 2) people feel detached from the value of their data. We believe the latter could be due to how data are spread across multiple devices and technologies and managed by different institutions.

In Article V, some participants considered their data to be of no value simply because they could not perceive how such data would be useful to a purchasing entity. This assumption is probably due to the individual nature of the data. Should these individual units of data be aggregated, it's usefulness may become clearer, at which point people would consider it much more valuable. One participant notes in Article V, *"I cannot see how these organisations can meaningfully take advantage of these data"*. The lower valuations for the various data units may therefore be due to participants' assumption of the "little information" about their health and lifestyle conveyed by the individual data units. This could be a result of participants' lack of comprehension of how valuable such data could be to the purchasing entity.

In Article I, it was clear that users did not have a clear idea of what their data were worth, as responses ranged from zero to hundreds of thousands of dollars with no convergence points. However, the participants expressed the need for payment for the collection and use of their data. There was an emphasis on the intended use and its subsequent contribution to society. While participants considered their data to be priceless, they were willing to give them up for free for research purposes that would benefit the wider society.

5.2.3 RQ3 - What ethical issues of personal health data reuse do people identify in future software-driven healthcare solutions?

Open health data have broad potential for use in the creation of health and wellness-related software solutions, scientific research, and for the identification and prevention of previously unknown diseases. Our participants (Article III), despite having donated their data as open data, still wanted to have a say in how the data were to be used, by whom, and where. Our findings indicate that the attachment people have to their data makes it difficult for them to understand that by donating it, they have essentially revoked control and ownership of the data. This unconscious perception of ownership could be an issue to contend with when using such data for the development of software solutions, particularly by private stakeholders. The participants' highlighting of past experiences of data abuse by private stakeholders such as insurance companies, creates a concern about who might have access to their donated data in the future. With the privacy and

anonymisation of personal data not enough to prevent deidentification (Sweeney, Abu, & Winn, 2013), people are concerned about their identity being revealed, despite donating data anonymously. On the contrary, deidentification of personal data also limits the ability to target specific demographic groups and presents severe roadblocks for health solutions that depend on identifiable data.

The potential benefits of using open health data to develop software solutions for understanding and diagnosing previously unknown diseases are extensive. A major issue of concern here is third-party data sharing (Article IV). In Article IV, we note that some software solutions routinely fail to provide transparency in collecting and transmitting personal data to third-parties (Blenner et al., 2016; Grindrod et al., 2017). While the sharing of the data with third-parties is often within legal barriers due to users having agreed to terms and conditions by simply ticking a box, the sharing of data is often warranted to enhance user experience or even monetise the applications (Binns et al., 2018). As such, data collected by one party may be shared with another for the "benefit" of the user. Unfortunately, the third party may use such data in a way that disproportionately targets the user through targeted advertising, for example. Similarly, individuals also passively trade their personal data in exchange for free access to various software applications. They also willingly share sensitive information about themselves and others on social media (which are also software solutions).

5.3   Implications for digital health services

Digital services that rely on personal health data can be assured of people's willingness not only to sell their data but also to donate it if necessary. Thus, depending on the use case, digital services can either get such data for free or purchase them from users for a fee. The decreasing cost of personal data as evidenced in Article V indicates that large amounts of data could be bought at a relatively cheap price. However, this is not so simple, because the purchasing entity has a part to play in this. For example, in Article III, people are willing to donate their data to open data initiatives that focus on services for the greater societal good. Similarly, in Article V, we observe that the value people placed on their data when the purchasing entity was a research institution was cheaper than other purchasing entities. In essence, the role of the entity that is collecting the data in building a service influences how available such data could be and how much they will cost. MyData could play a pivotal role here as people will be able to easily grant or revoke access to the use of their data or even sell it for a fee, depending on the intended use or the purchasing entity.

People view societal stakeholders (e.g. non-governmental organisations and academic research institutions) more favourably than public and private stakeholders. Interestingly, in Article V, private and public stakeholders were considered almost as equals, as data sold to these entities were priced almost equally. Our findings indicate that people trust societal stakeholders with their data far more than they do with public (government) and private stakeholders. One avenue created by this level of trust is for societal stakeholders to join in the development of digital services and software solutions that rely on people's personal data. This is important, as people trust societal stakeholders to use their data for the broader societal benefit.

Crowd workers' willingness to donate data both for free and at a price is also interesting. Crowdsourcing is known to offer a state-of-the-art means for obtaining ecologically valid data at scale (Kittur et al., 2013) and has been hypothesised as key to finding new solutions to wicked healthcare problems (Brabham, 2008). By harnessing the power of the crowd, large amounts of data could be accessed to enable the building of such future digital health services. Another important opportunity lies with OHD. Efforts can be made to remove the barriers that are prohibiting users from contributing to OHD initiatives, as this can open the space for the building of bespoke services that rely heavily on personal health data.

Most ethical issues about data are related to data collected about people - thus, personal data. To this end, our research highlights a distrust of companies involved in the collection and processing of personal data. Although not surprising in and of itself, from both the academic literature (Bahrini et al., 2019; Zou & Schaub, 2018) and numerous real life cases of data misuse and abuse such as the infamous Facebook-Cambridge Analytica scandal (Lapaire, 2018), it outlines an Achilles heel of initiatives such as MyData. Active steps should be taken not only to fulfill legal and ethical considerations but to address the concerns users have about their personal data. For solid advances to be made in the healthcare domain, trust must be developed among all the stakeholders in the data management chain.

5.4   Towards a new paradigm

The concerns of people about retaining control of their personal data calls for new data management models. MyData, an emerging human-centric data management model and set of guidelines that aims to empower people to access, use, manage, and grant permission to their personal data, is closely related (Poikola et al., 2020). In Article II, our participants expressed support for the MyData initiative, as it serves as an

independent intermediary between them and the private stakeholders that currently hold most of their data.

MyData could play an important role in facilitating access to quality data to the software industry, particularly to software solutions that rely on massive amounts of data (e.g. artificial intelligence software solutions). It could also affect the creation of future digital health solutions that rely on personal data as core building blocks. Even if the MyData initiative does not see the light of day, we believe that data management should be studied at a more granular level.

5.5 Future work

First, it will be valuable to explore the potential of the concept of *"healthy carefulness"* in relation to data ethics as well as consider the risks and moral obligations of data reuse for public health purposes. Additionally, it will be important to investigate how various stakeholders (researchers, private and public institutions, governments, software designers, etc) may best navigate the often conflicting perspectives of users on the sharing and reuse of their data.

Secondly, the implementation of the MyData framework could offer a promising avenue for achieving an ethical and serendipitous approach to data collection. Should the MyData framework take root, it could empower individuals through simplified practices of consent, access, and control, and provide opportunities for them to benefit monetarily from their personal data. It could also be instrumental in facilitating the creation of future digital health software that use people's health data as core building blocks.

5.6 Limitations

In Article I, for example, our results were biased towards people who were willing to interact with the public display setup. Most of the video submissions were made after campus hours between 5pm and midnight. It is possible that some people may have been discouraged from participating because of embarrassment or awkwardness associated with engaging with the technology probe in public.

The majority of the limitations recorded in the articles presented in this thesis are inherent to crowd workers. As crowd workers are online primarily to make money, there is a chance that we missed more qualified candidates due to the competitive nature of jobs available to these workers. For example, task timeout was a major issue that crowd workers faced when they were unable to deliver tasks on time. While we verified tasks

to ensure that tasks marked as timed out were paid for as long as they were submitted, we are sure that some workers may have abandoned the tasks because of these time restraints.

There is also a potential for skewed results due to the recruitment of participants from crowdsourcing marketplaces for the studies used in this thesis. Although these marketplaces provide a convenient and cost-effective way to recruit a large sample size quickly, they may attract participants with different characteristics than those of the general population. For example, individuals who frequent these platforms may be more tech-savvy, more motivated to participate in online studies, or have more flexible schedules than individuals who do not use these platforms. This could potentially bias the results of the study and limit its generalizability to the broader population.

Further, the results from the studies are limited to the context and culture of the locations in which the data were collected. Indeed, by relying on crowdworkers, our results (Articles I - V) are biased toward those who know about and are willing to participate on these platforms. In Articles I-V, the studies were conducted online, and in Article V, the participants were expected to have an Android device, live in the UK, and have low back pain, which means only those with the condition, equipment, and the ability to use these technologies were eligible. Moreover, the bulk of our participants on these crowdsourcing marketplaces come from the US, India, UK, and Finland. In Article V for example, it can be posited that the openness of crowd workers to sell their personal data may be in part due to the fact that by working on a crowdsourcing platform, they are already predisposed to make money that way. This means our results cannot be generalised for the general population. Despite these limitations, our results are in line with existing research and are indicative of much broader trends, as results from online marketplaces have been valuable to research with high external validity even in empirical research (Brown et al., 2018; Pedersen et al., 2013; Strickland & Stoops, 2018; Swan, 2012).

Additionally, while we ensured that our data collection was of high quality, it remains possible that some users performed the tasks casually (e.g. using the public display setup in Article I and the mobile app in Article V). Although these casual attitudes to research are consistent with previous findings, we are confident that they represent only a small fraction of the data collected.

To conclude, due to both the aforementioned limitations, we emphasise that our approaches are not a silver bullet for gathering contributions from the crowd but provide an additional support to existing approaches.

6   Conclusion

This research investigated user perceptions of the ownership, valuation, and reuse ethics of personal data. The broad aim was to gain an understanding of people's perception of personal data particularly in the health domain, and to shed light on the valuation, reuse, and ethical factors involved in the building of innovative future health services based on personal data.

Throughout the thesis, we present insights into personal data management based on empirical results from situated public display tasks, online experiments, and mobile data collection and auction tools. We employed in-the-wild field studies with purpose-built applications acting as technology probes, complemented by online experiments. We argue that gaining an understanding of user perceptions of their personal data can benefit services that rely on such data. We see an opportunity for crowdsourcing marketplaces to play an important role in serving as as a resourceful human-subject pool for accessing personal health data.

However, we believe the work presented in this thesis only serves as a first step towards an extensive investigation into user perceptions in relation to personal data management. Various research avenues remain unexplored in this domain. For example, it is important to explore the differences in perspectives of people across different geographical areas or even from different countries within the same geographical area.

In summary, we are quite optimistic about the future of personal data management. If efforts are made to resolve the concerns of people concerning the management of their data, we believe more and more people will be open to the idea of sharing, selling, or donating their data to various initiatives, including research efforts. This work contributes empirical insights and implications based on data ownership, valuation, and reuse.

References

Aghdasi, N., Bly, R., White, L. W., Hannaford, B., Moe, K., & Lendvay, T. S. (2015). Crowd-sourced assessment of surgical skills in cricothyrotomy procedure. *Journal of surgical research*, *196*(2), 302–306.

Alt, F., Shirazi, A. S., Schmidt, A., Kramer, U., & Nawaz, Z. (2010). Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proceedings of the 6th nordic conference on human-computer interaction: Extending boundaries* (pp. 13–22).

Andanda, P. (2020). Ethical and legal governance of health-related research that use digital data from user-generated online health content. *Information, Communication & Society*, *23*(8), 1154–1169.

Arch, J. J., & Carr, A. L. (2017). Using mechanical turk for research on cancer survivors. *Psycho-oncology*, *26*(10), 1593–1603.

Bahrini, M., Wenig, N., Meissner, M., Sohr, K., & Malaka, R. (2019). Happypermi: Presenting critical data flows in mobile application to raise user security awareness. In *Extended abstracts of the 2019 chi conference on human factors in computing systems* (pp. 1–6). Glasgow, Scotland UK: Association for Computing Machinery.

Ballantyne, A. (2019). Adjusting the focus: a public health ethics approach to data research. *Bioethics*, *33*(3), 357–366.

Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M., & Chalmers, M. (2008). From awareness to repartee: sharing location within social groups. In *proceedings of the sigchi conference on human factors in computing systems* (pp. 497–506).

Baudendistel, I., Winkler, E., Kamradt, M., Brophy, S., Längst, G., Eckrich, F., . . . Ose, D. (2015). The patients' active role in managing a personal electronic health record: a qualitative analysis. *Supportive Care in Cancer*, *23*(9), 2613–2621.

Becker, S., Miron-Shatz, T., Schumacher, N., Krocza, J., Diamantidis, C., Albrecht, U.-V., et al. (2014). mhealth 2.0: experiences, possibilities, and perspectives. *JMIR mHealth and uHealth*, *2*(2), e3328.

Beigi, G., Shu, K., Zhang, Y., & Liu, H. (2018). Securing social media user data: An adversarial approach. In *Proceedings of the 29th on hypertext and social media* (pp. 165–173).

Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., . . . Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association*, *23*(e1), e42–e48.

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third party tracking in the mobile ecosystem. In *Proceedings of the 10th acm conference on web science* (pp. 23–31).

Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? the measurement, governance, and valuation of digital personal data by big tech. *Big Data & Society*, *8*(1), 20539517211017308.

Blenner, S. R., Köllmer, M., Rouse, A. J., Daneshvar, N., Williams, C., & Andrews, L. B. (2016). Privacy policies of android diabetes apps and sharing of health information. *Jama*, *315*(10), 1051–1052.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75–90.

Brabham, D. C., Ribisl, K. M., Kirchner, T. R., & Bernhardt, J. M. (2014). Crowd-sourcing applications for public health. *American journal of preventive medicine*, *46*(2), 179–187.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Brown, G., McAlpine, C., Rhodes, J., Lunney, D., Goldingay, R., Fielding, K., ... others (2018). Assessing the validity of crowdsourced wildlife observations for conservation using public participatory mapping methods. *Biological Conservation*, *227*, 141–151.

Callan, M. J., Kim, H., Gheorghiu, A. I., & Matthews, W. J. (2017). The interrelations between social class, personal relative deprivation, and prosociality. *Social Psychological and Personality Science*, *8*(6), 660–669.

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2013). Your browsing behavior for a big mac: Economics of personal information online. In *Proceedings of the 22nd international conference on world wide web* (pp. 189–200).

Cohen, J. E. (2019). *Between truth and power*. Oxford University Press.

Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., & Powledge, P. (2005). Location disclosure to social relations: why, when, & what people want to share. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 81–90).

Crawford, H., & Renaud, K. (2014). Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, *1*(1), 1–28.

Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006). A study on the value of location privacy. In *Proceedings of the 5th acm workshop on privacy in electronic*

society (pp. 109–118).

Dolley, S. (2018). Big data's role in precision public health. *Frontiers in public health*, 68.

Drexl, J., Hilty, R., Desaunettes-Barbero, L., Greiner, F., Kim, D., Richter, H., ... Wiedemann, K. (2016). Data ownership and access to data-position statement of the max planck institute for innovation and competition of 16 august 2016 on the current european debate. *Max Planck Institute for Innovation & Competition Research Paper*(16-10).

Edwards, L. (2018). Data protection: Enter the general data protection regulation. *Law, Policy and the Internet (Hart Publishing, 2018)*.

Eke, P. I. (2011). Using social media for research and public health surveillance. *Journal of dental research*, *90*(9), 1045.

European Commission. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*. European Commission. Retrieved from `https:// eur-lex.europa.eu/eli/reg/2016/679/oj`

Farkas, T. J. (2017). Data created by the internet of things: the new gold without ownership. *Rev. Prop. Inmaterial*, *23*, 5.

Fiesler, C., & Hallinan, B. (2018). We are the product: Public reactions to online data sharing and privacy controversies in the media. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 53).

Goddard, M. (2017). The eu general data protection regulation (gdpr): European regulation that has a global impact. *International Journal of Market Research*, *59*(6), 703–705.

Green, M. C. (2008). Transportation theory. *The international encyclopedia of communication*.

Grindrod, K., Boersema, J., Waked, K., Smith, V., Yang, J., & Gebotys, C. (2017). Locking it down: The privacy and security of mobile medication apps. *Canadian Pharmacists Journal/Revue Des Pharmaciens Du Canada*, *150*(1), 60–66.

Häkkilä, J., Vahabpour, F., Colley, A., Väyrynen, J., & Koskela, T. (2015). Design probes study on user perceptions of a smart glasses concept. In *Proceedings of the 14th international conference on mobile and ubiquitous multimedia* (pp. 223–233).

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big data*, *6*(3), 176–190.

Hann, I.-H., Hui, K.-L., Lee, S.-Y. T., & Png, I. P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of management information systems*, *24*(2), 13–42.

Hargittai, E., & Marwick, A. (2016). "what can i really do?" explaining the privacy paradox with online apathy. *International journal of communication*, *10*, 21.

Harwell, D. (2019, Apr). *Is your pregnancy app sharing your intimate data with your boss?* WP Company. Retrieved from https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/?arc404=true

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Helm, T. (2019, Dec). *Patient data from gp surgeries sold to us companies*. Guardian News and Media. Retrieved from https://www.theguardian.com/politics/2019/dec/07/nhs-medical-data-sales-american-pharma-lack-transparency

Hosio, S. J., Karppinen, J., Takala, E.-P., Takatalo, J., Goncalves, J., Van Berkel, N., ... Kostakos, V. (2018). Crowdsourcing treatments for low back pain. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 276).

Howe, J., et al. (2006). The rise of crowdsourcing. *Wired magazine*, *14*(6), 1–4.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277–1288.

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *IEEE security & privacy*, *3*(5), 22–25.

Janeček, V. (2018). Ownership of personal data in the internet of things. *Computer law & security review*, *34*(5), 1039–1052.

Jones, S. M., Du, Y., Panattoni, L., & Henrikson, N. B. (2019). Assessing worry about affording healthcare in a general population sample. *Frontiers in psychology*, *10*, 2622.

Karat, J., Karat, C.-M., & Vergo, J. (2004). Experiences people value: The new frontier of task analysis. *The handbook of task analysis for human-computer interaction*, 585–603.

Kim, J., Park, H.-A., et al. (2012). Development of a health information technology acceptance model using consumers' health behavior intention. *Journal of medical Internet research*, *14*(5), e2143.

Kish, L. J., & Topol, E. J. (2015). Unpatients—why patients should own their medical data. *Nature biotechnology*, *33*(9), 921–924.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., . . . Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 1301–1318).

Kostkova, P. (2018). Disease surveillance data sharing for public health: the next ethical frontiers. *Life sciences, society and policy*, *14*(1), 1–5.

Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., . . . others (2016). Who owns the data? open data for healthcare. *Frontiers in public health*, *4*, 7.

Kotut, L., Stelter, T. L., Horning, M., & McCrickard, D. S. (2020). Willing buyer, willing seller: Personal data trade as a service. In *Companion of the 2020 acm international conference on supporting group work* (pp. 59–68).

Krutzinna, J., Taddeo, M., & Floridi, L. (2019). Enabling posthumous medical data donation: A plea for the ethical utilisation of personal health data. *The Ethics of Medical Data Donation*, 163–180.

Kuikkaniemi, K., Poikola, A., & Honko, H. (2015). Mydata a nordic model for human-centered personal data management and processing.

Lanier, J. (2014). *Who owns the future?* Simon and Schuster.

Lapaire, J.-R. (2018). Why content matters. zuckerberg, vox media and the cambridge analytica data leak. *ANTARES: Letras e Humanidades*, *10*(20), 88–110.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Lehtiniemi, T., & Haapoja, J. (2020). Data agency at stake: Mydata activism and alternative frames of equal participation. *new media & society*, *22*(1), 87–104.

Li, X.-B., Liu, X., & Motiwalla, L. (2021). Valuing personal data with privacy consideration. *Decision Sciences*, *52*(2), 393–426.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37).

Marcia, E. (2019). *Case studies: High-profile cases of privacy violation.* SGR Law. Retrieved from `https://www.sgrlaw.com/ttl-articles/case-studies -high-profile-cases-of-privacy-violation/`

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. (2017). "now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization*, *140*, 1–17.

Martani, A., Shaw, D., & Elger, B. S. (2019). Stay fit or get bit-ethical issues in sharing health data with insurers' apps. *Swiss medical weekly*, *149*(2526).

Martin, E. G., Helbig, N., & Birkhead, G. S. (2015). Opening health data: what do

researchers want? early experiences with new york's open health data platform. *Journal of Public Health Management and Practice*, *21*(5), E1–E7.

Meeussen, L., & Van Laar, C. (2018). Feeling pressure to be a perfect mother relates to parental burnout and career ambitions. *Frontiers in psychology*, *9*, 2113.

Mikkelsen, Ø., Hartvigsen, S. H., Hauge, K. H., Nordenson, S., Nordtveit, E., Nydal, R., . . . others (2016). Guidelines for research ethics in science and technology.

Opoku Asare, K., Visuri, A., Vega, J., & Ferreira, D. (2022). Me in the wild: An exploratory study using smartphones to detect the onset of depression. In *International conference on wireless mobile communication and healthcare* (pp. 121–145).

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., . . . De Vreede, G.-J. (2013). Conceptual foundations of crowdsourcing: A review of is research. In *2013 46th hawaii international conference on system sciences* (pp. 579–588). Wailea, Maui, HI, USA: IEEE.

Poikola, A., Kuikkaniemi, K., Kuittinen, O., Honko, H., Knuutila, A., & Lähteenoja, V. (2020). Mydata—an introduction to human-centric use of personal data. *Finnish Ministry of Transport and Communications, 55p*, *8*.

Rissanen, T. (2016). Public online services at the age of mydata: a new approach to personal data management in finland.

Rowland, S. P., Fitzgerald, J. E., Holme, T., Powell, J., & McGregor, A. (2020). What is the clinical value of mhealth for patients? *NPJ digital medicine*, *3*(1), 1–6.

Saari, D. G. (2000). The mathematics of voting: Democratic symmetry. *Economist*, *83*.

Schoen, J., Mallett, J. W., Grossman-Kahn, R., Brentani, A., Kaselitz, E., & Heisler, M. (2017). Perspectives and experiences of community health workers in brazilian primary care centers using m-health tools in home visits with community members. *Human resources for health*, *15*(1), 1–10.

Silberman, M. S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., & Zaldivar, A. (2018). Responsible research with crowds: pay crowdworkers at least minimum wage. *Communications of the ACM*, *61*(3), 39–41.

Simmonds, G., Woods, A. T., & Spence, C. (2018). 'show me the goods': Assessing the effectiveness of transparent packaging vs. product imagery on product evaluation. *Food quality and preference*, *63*, 18–27.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS quarterly*, 989–1015.

Staiano, J., Oliver, N., Lepri, B., de Oliveira, R., Caraviello, M., & Sebe, N. (2014). Money walks: a human-centric study on the economics of personal mobile data. In *Proceedings of the 2014 acm international joint conference on pervasive and*

ubiquitous computing (pp. 583–594).

Stevens, B. (2001). Hospitality ethics: Responses from human resource directors and students to seven ethical scenarios. *Journal of Business Ethics*, *30*(3), 233–242.

Strickland, J. C., & Stoops, W. W. (2018). Feasibility, acceptability, and validity of crowdsourcing for collecting longitudinal alcohol use data. *Journal of the Experimental Analysis of Behavior*, *110*(1), 136–153.

Swan, M. (2012). Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *Journal of medical Internet research*, *14*(2), e46.

Sweeney, L., Abu, A., & Winn, J. (2013). Identifying participants in the personal genome project by name (a re-identification experiment). *arXiv preprint arXiv:1304.7605*.

Tangari, G., Ikram, M., Ijaz, K., Kaafar, M. A., & Berkovsky, S. (2021). Mobile health and privacy: cross sectional study. *bmj*, *373*.

Toch, E., & Levi, I. (2013). Locality and privacy in people-nearby applications. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (pp. 539–548).

Van den Hoven, J. (2013). *Internet of things factsheet ethics.*

van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2013, 08). The Extended Transportation-Imagery Model: A Meta-Analysis of the Antecedents and Consequences of Consumers' Narrative Transportation. *Journal of Consumer Research*, *40*(5), 797-817. doi: 10.1086/673383

Vickrey, W. (1962). Auctions and bidding games. *Recent advances in game theory*, *29*(1962), 15–27.

Zou, Y., & Schaub, F. (2018). Concern but no action: Consumers' reactions to the equifax data breach. In *Extended abstracts of the 2018 chi conference on human factors in computing systems* (pp. 1–6).

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack obama's books of 2019*. Profile books.

Original publications

I   Alorwu A, van Berkel N, Goncalves J, Oppenlaender J, López M B, Seetharaman M, & Hosio S (2020). Crowdsourcing sensitive data using public displays—opportunities, challenges, and considerations. Personal and Ubiquitous Computing, 1-16. doi: 10.1007/s00779-020-01375-6

II   Alorwu A, Kheirinejad S, van Berkel N, Kinnula M, Ferreira D, Visuri A, & Hosio S (2021). Assessing MyData Scenarios: Ethics, Concerns, and the Promise. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-11). doi: 10.1145/3411764.3445213

III   Alorwu A, Visuri A, van Berkel N, & Hosio S J (2021). (Re) using Crowdsourced Health Data: Perceptions of Data Contributors. IEEE Software, 39(1). doi: 10.1109/MS.2021.3117684

IV   Kheirinejad, S, Alorwu, A, Visuri, A, & Hosio, S (2022). Contrasting the Expectations and Experiences Related to Mobile Health Use for Chronic Pain: Questionnaire Study. JMIR Human Factors, 9(3), e38265. doi: 10.2196/38265

V   Alorwu, A, van Berkel, N, Visuri, A, Suryanarayana, S, Yoshihiro, T, & Hosio, S (under review, 202X). Monetary Valuation of Personal Health Data in the Wild. *IJHCS*

871. Alzaza, Ahmad (2023) Cementitious binders for construction in northern cold regions

872. Kantanen, Pekka (2023) The role of finely divided retained austenite on the mechanical properties of QP and ART processed novel 0.3C ultrahigh strength steels

873. Markkula, Juho (2023) Communicating smart grid demand response and control data through LTE network related solutions

874. Laukkanen, Johanna (2023) Alkali-activation of industrial aluminosilicate sidestreams : application for mine water treatment and sediment remediation

875. Ismail, Mostafa (2023) Advanced imaging of lignocellulosic and cellulose materials

876. Abdelrahim, Ahmed (2023) Towards lower CO_2 emissions in iron and steelmaking : hydrogen-assisted reduction and cement-free briquettes

877. Zhou, Jin (2023) Conductive composites of engineered nanomaterials with excellent gas sensing properties

878. Ali, Farooq (2023) A framework for analyzing, developing, and managing stakeholder network relationships in collaborative hospital construction projects

879. Rönkkö, Pasi (2023) Circular economy as an enabler of improved resilience and material availability in supply chains

880. Isteri, Visa (2023) Alternative ye'elimite (CSA) cement clinkers from industrial byproducts

881. Aprialdi, Dwinata (2023) Forest hydrological studies to improve water management in marine lowlands in Palembang, Indonesia

882. Koskela, Aki (2023) Utilisation of lignin-based biocarbon in pyrometallurgical applications

883. Jääskä, Elina (2023) Game-based learning methods in project management higher education

884. Adediran, Adeolu Idowu (2023) Alkali activation of iron-rich fayalite slag : fresh, hardened and durability properties

885. Mansoori, Solmaz (2023) Improving data utilization in construction : on the path towards industrialization

886. Hannila, Esa (2023) Quality assessment and reliability of additively manufactured hybrid structural electronics

ACTA UNIVERSITATIS OULUENSIS

SERIES EDITORS

A
SCIENTIAE RERUM NATURALIUM
University Lecturer Mahmoud Filali

B
HUMANIORA
University Lecturer Santeri Palviainen

C
TECHNICA
Senior Research Fellow Antti Kaijalainen

D
MEDICA
University Lecturer Pirjo Kaakinen

E
SCIENTIAE RERUM SOCIALIUM
University Lecturer Henri Pettersson

F
SCRIPTA ACADEMICA
Strategy Officer Mari Katvala

G
OECONOMICA
University Researcher Marko Korhonen

H
ARCHITECTONICA
Associate Professor Anu Soikkeli

EDITOR IN CHIEF
University Lecturer Santeri Palviainen

PUBLICATIONS EDITOR
Publications Editor Kirsti Nurkkala

UNIVERSITY OF OULU