Assessing MyData Scenarios: Ethics, Concerns, and the Promise

Andy Alorwu
University of Oulu
Oulu, Finland

Saba Kheirinejad
University of Oulu
Oulu, Finland

Niels van Berkel
Aalborg University
Aalborg, Denmark

Marianne Kinnula
University of Oulu
Oulu, Finland

Denzil Ferreira
University of Oulu
Oulu, Finland

Aku Visuri
University of Oulu
Oulu, Finland

Simo Hosio
University of Oulu
Oulu, Finland

ABSTRACT

Public controversies around the unethical use of personal data are increasing, spotlighting data ethics as an increasingly important field of study. MyData is a related emerging vision that emphasizes individuals' control of their personal data. In this paper, we investigate people's perceptions of various data management scenarios by measuring the perceived ethicality and level of felt concern concerning the scenarios. We deployed a set of 96 unique scenarios to an online crowdsourcing platform for assessment and invited a representative sample of the participants to a second-stage questionnaire about the MyData vision and its potential in the field of healthcare. Our results provide a timely investigation into how topical data-related practices affect the perceived ethicality and the felt concern. The questionnaire analysis reveals great potential in the MyData vision. Through the combined quantitative and qualitative results, we contribute to the field of data ethics.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Security and privacy → Human and societal aspects of security and privacy**; *Social aspects of security and privacy*.

KEYWORDS

MyData, Ethics, Healthcare, Crowdsourcing, Privacy

ACM Reference Format:
Andy Alorwu, Saba Kheirinejad, Niels van Berkel, Marianne Kinnula, Denzil Ferreira, Aku Visuri, and Simo Hosio. 2021. Assessing MyData Scenarios: Ethics, Concerns, and the Promise. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3411764.3445213

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445213

1 INTRODUCTION

There is a global concern around the misuse of personal data. Such data are increasingly being generated and collected through the myriad hardware devices and software applications that we use daily, including smartphones, fitness trackers, and different types of applications. In addition to this, companies and governments store various types of data about us, such as medical records or location-related activity logs. The concerns often deal with issues of data sharing with third party organizations, such as governments or businesses, and sometimes without clear user consent or even illegally [21, 25, 59]. Notable public scandals in this regard include, for instance, tech vendors pre-installing software on their hardware to collect consumer data without the knowledge or consent of consumers and selling this data to third parties [49], or the infamous Cambridge Analytica scandal where millions of Facebook users' personal data were used to support political campaigns [44]. In the health domain, examples of unethical conduct include *e.g.* surgery data of millions of patients being sold to pharmaceutical companies for research purposes [29], or data about female employees' fertility, pregnancy, and childbirth being shared with their employers [28].

Given the role of personal data across all use of contemporary technology, data management and data ethics are timely and lively research topics. In our focus is *MyData*, a data management vision and a set of principles that posit the individual as the owner and controller of the generated personal data [38, 60]. MyData aims to empower individuals to better control their data and promote its best use [51]. As a result, it aims to simplify the challenges centered around consent, access, and control of one's personal data, and creates opportunities for service innovation [24]. Many ethical issues about data are related to data collected about people – personal data [26]. Specifically in the health domain, MyData has been seen as an ideal – even if currently somewhat optimistic – model that would allow the individual (data producer) to be the controller of the data that are currently largely managed and owned by large corporations [24]. MyData has been prominently discussed in both public healthcare and governmental contexts [24, 29, 35].

There is a clear identified need to understand user perceptions towards personal data management and/or use of personal data by third-parties, thus organizations that store and control the data [3, 5, 21, 41, 48]. A pivotal topic here is the ethicality of data management, which has been increasingly investigated in the field of

Human-Computer Interaction (HCI) as well, *e.g.* concerning ethical and appropriate data use [48], ethical use of qualitative data [53], use of family civic data [12], apps using location data and other sensitive data [3], or indeed ethical issues surrounding data use in healthcare [5]. Another dimension is the felt subjective concern of data management; individual's expression of concern for privacy can often be even contrary to their actions on e.g. social networks due to the inconvenience related to e.g. switching a platform [1, 64]. In a similar vein, after hearing about the Cambridge Analytica scandal users hardly reacted or even expressed much concern [30].

In this paper, we investigate end-user perceptions of data management practices using a novel quantitative scenario-based approach supplemented with qualitative data. At the center of this investigation is MyData, a data management vision that is yet to catch the attention of the CHI community. Using a variety of different data management scenarios as the experimental stimuli, we collected quantitative data using Amazon's Mechanical Turk [18] on ethics and concern on 96 different scenarios that emulate common types of personal data use and misuse, as modelled after related literature and various historical real-world events [19, 21, 28, 29, 44, 49, 68]. We additionally invited the contributors back for a second-stage study to gain descriptive statistics about our sample and their qualitative insights on broader issues around MyData. The key contributions of our work are:

(1) We present an understanding of end-user perceptions on personal data management.
(2) We introduce the MyData vision to the CHI community and its relevance to research on data ethics.
(3) We contribute a survey-based method and instrument to assess people's data management perceptions.

We collected 1920 submissions from 172 Amazon Mechanical Turk (MTurk) workers, and invited them to a second-stage questionnaire. Our results identify stark differences across parameters that affect people's evaluation of ethicality and the overall felt concern about the presented scenarios. We also quantitatively show that parameters that directly affect ethicality do not necessarily at all affect the felt concern, i.e. the perceived ethicality and concern are not tightly coupled. Our qualitative analysis highlights participants' excitement towards the prospect of having control over their personal data – especially their health data – and deciding who has access to it. At the same time, participants demonstrate a degree of negligence to their data. Finally, our discussion is helpful for paving the way for the MyData vision and contributes to research on data ethics.

2 RELATED WORK

Our work is situated in the intersection of MyData, data ethics, and data privacy. We focus in here on recent works especially in the field of HCI.

2.1 MyData

MyData is an emerging personal data management and processing vision that focuses on the consumer's perspective and considers personal data as a resource that the consumer can access and control [60]. MyData's human-centric approach pushes for the release of personal data from the confines of monopolistic data holders in order for the full potential of personal data to be realized [46]. According to Poikola et. al [60], MyData can be looked at in two folds: 1) a new approach in the management and processing of personal data, putting the individual at the center of the process, and 2) personal data about an individual that is under the control of that individual. MyData is therefore both an infrastructure-level approach for ensuring data interoperability and portability, making it possible for individuals to change service providers without proprietary data lock-ins, and a consent-based data management approach that puts the control of an individual's data in the hands of the individual [60]. Born out of the open data initiative, MyData includes personal information such as health and fitness data, social media data, data from internet usage, and financial data, among others. The term *MyData* therefore refers to both a data management vision and data as a resource [60].

Related to this vision, a study by Kim & Park aimed at describing consumers' behavioural intention of using health information technology found out that the effective use of collected health-related data is dependent on the behavioural intention of consumers to measure, store, and manage their own data [36]. With MyData aiming to enforce consumers have their personal data under their control, it is important for consumers to be incentivised to take active actions with their data as new solutions and services will be built based on this data [38]. The benefits of MyData have been found to surpass simply having access and control over one's data to becoming an enabler of better self-care. To this end, a study by Baudendistel et al. suggests that having control over one's personal data leads to better motivation to take care of one's own health [10].

Globally, there are similar projects aimed at moving the control of personal data from companies to the data creators. For instance, Sir Tim Berners-Lee, the pioneer of the world wide web, has embarked on a mission to make personal data a resource for consumers [6, 14]. The father of virtual reality, Jaron Lanier has also made the call for the establishment of a balance between data holders (companies) and consumers by rewarding consumers for the use of their personal data [43]. In a study comparing data access trends, Iemma proposed the use of smart disclosure programs to release machine-readable personal data from firms to consumers [34]. In a similar vein, Lehtiniemi introduced software that has been developed to provide users the means to have control over the collection and use of their personal data [45].

2.2 Data ethics

HCI researchers have acknowledged ethical concerns regarding data use. In the context of family civic data, i.e. any kind of data concerning families stored by a public authority, Bowyer et al. [12] found out that there is a need for the development of interfaces that support Human-Data Interaction [16], to make it possible for families to co-operatively manage their data. The worries about family data use were related to potentially misleading data in the form of numbers and standardized labelling, such as e.g. 'domestic violence', not capturing the nuanced reality. Regarding credit data, Zou & Schaub identified a need for empowering consumers to 'take ownership' of their data by managing their credit data, so that they understand the need to take protective actions against e.g. identity theft [74]. Researchers have also been concerned about

apps unnecessarily accessing certain types of data in mobile devices and teaching users to be aware of the kind of permissions they give different apps [7, 50]. In the context of higher education, the use of student data in learning analytic systems raised students' concerns of how their data was being used as well as expressed the need to be involved in designing the learning analytics process [69]. The need to teach students ethical issues related to data collection and use has also been identified [65]. Generally, it can be said that HCI research aims for user empowerment as regards their data, finding ways for users to be educated and in control of how their data is used and by whom [41], in line with MyData aims.

Health data today is being used outside the parameters of the original consent agreed to by the data producer as many of such uses were not predictable at the time of data collection. The deviation of future uses of data from the expectations of the data producer and the purposes for which the data collection was disclosed raises a number of ethical concerns [8]. Complicating issues further is that health data is no longer about clinical records only but has expanded to include fitness and health-related behaviour 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 [8, 40, 41]. Martani et al. investigated how health insurers' mobile applications encourage customers to share their personal behavioural and health data with them in exchange for monetary rewards [50]. The sensitive and personal nature of health data poses many ethical challenges even for research purposes. Therefore, more attention is needed to address regulatory and ethical issues of data [39]. Andanda investigated legal and ethical concerns that appear from utilizing health-related data generated by users of some online platforms for the purposes of health-based research [4].

With no existing ethical framework or process for the donation of one's medical data to a public institution for research purposes, Krutzinna et al. argued that it is an ethical failure not to exploit important data that matter to improve public health. This they believe is a failure of past data management practices and as such argue that people's personal data should be made available for scientific research by encouraging the people to donate their data posthumously similar to how human organs are currently donated [42].

2.3 User Concerns about Data Management

The concern people feel around data management results from a variety of issues, including for example privacy-related risks and practices. As no single source for concern exists, end-user centric exploration of concern is warranted, e.g. by considering the needs and viewpoints of different stakeholders [2]. User concerns are also exacerbated in the context of sensitive data, such as health related data, by the fact that most health data is electronic and thus easily shareable [61].

Acquisti and Gross conducted a study on members of a social network to investigate the impact of data management concerns and compared users' stated attitudes to actual behaviour on a social network platform [1]. Their results showed that individual's expression of concern for privacy was contrary to their actions on the network as they revealed a great amount of personal information. This 'paradox' is an attitudinal issue that reveals underlying behavioural processes such as lack of self-control or lack of impulse control. Indeed, although people feel concerned about data management practices of certain companies, they are often unwilling to switch services due to the associated inconvenience [61, 64].

In the wake of the Facebook-Cambridge Analytica scandal, Hinds et al. found out that respondents, after hearing about the scandal did not change their privacy settings, delete their accounts, or even express much concern [30]. Similarly, a study by Hargittai & Marwick revealed that young people believe various types of data mismanagement issues are inevitable and thus just do not care anymore [27], while others express the sentiment that they do not have control anyway over how their data is used [66].

While these works have covered ground in terms of understanding types of consumer concerns regarding their data, the HCI research community has recently identified a further need to understand these concerns in a wider and more general level [21]. Our goal is to add to this understanding of ethicality of usage of personal data.

3 THE STUDY

Our investigation is primarily focused on the ethicality and the felt concern of the conduct of various common data management scenarios, and secondarily on a qualitative investigation on the issues and potential of the MyData vision with a special emphasis on the health domain.

3.1 Method

We employed crowdsourcing for the data collection. Crowdsourcing, originally coined as *"the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call"* [31], has matured into an adaptive and valid approach to empirical research [58, 70]. We recruited workers from Amazon Mechanical Turk (MTurk), a platform used commonly by the HCI research community for a variety of tasks, including quantitative studies as well as creative work [15, 18, 52, 56]. The participants were required to be fluent in English and to have successfully completed at least 10000 tasks with an accept rate of 95% or higher.

3.2 Stage 1

In our study, a data management *scenario* is a construct of four different *parameters* – MyData operator action, data type, purpose, and consent – that all have one or more *levels*, as clarified in Table 1. The first stage required participants to read and assess the ethicality of a given data related scenario, indicate which specific parameter most affected this choice, and rate the overall felt concern about the scenario. The task was implemented using Amazon Web Service's (AWS) Crowd-HTML markup and deployed directly into the MTurk interface as a survey task, as depicted in Figure 1.

We crafted the scenarios based on prior works [19, 21, 68], and drew further inspiration from recent privacy and potentially ethical violations of people's data in the real world [28, 29, 44, 49]. The authors met three times to ideate a range of scenarios (diverging), followed by the selection of our final set of scenarios based on their

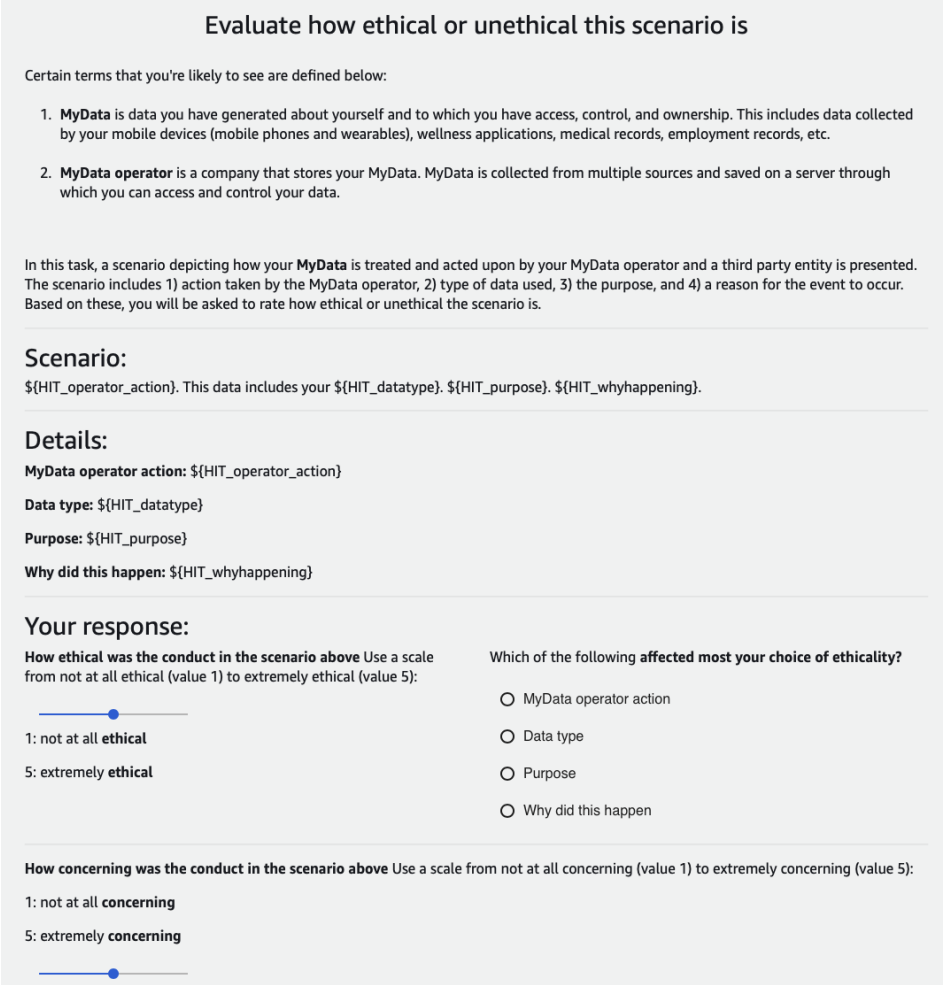

Figure 1: The Human Intelligence Task (HIT) deployed to MTurk. The variables denoted inside curly brackets were populated in runtime by Mturk with the parameter levels presented in Table 1, i.e. there were 96 unique scenarios deployed through this setup. Participants used the two sliders and the set of radiobuttons to indicate the perceived levels of ethicality and concern, and the decisive parameter for ethicality.

relevance and distinctiveness (converging). The first parameter, **MyData operator action**, consists of two levels – highlighting the difference between a MyData operator which financially profited from the transaction [28, 49] versus a MyData operator that did not have a financial incentive [21]. **Data type** is split across four levels, covering diagnostic and medical data [29], data from health trackers [32], location data (found to have an influence on health as it provides information about the social and environmental context within which patients live [11, 57, 73]), and personal media files (facial expressions found in pictures and videos have been used to detect health issues such as depression and having pain [17, 54, 62, 67]). The **purpose** of the data management covers the use by academic researchers [71], commercial companies [29], and the government [35]. Finally, the parameter of **consent** describes how the data is managed – without the user's consent, within a signed consent, within a signed consent but without the user being aware, and against the user's consent but within the existing legal framework [33, 44, 47, 63].

In Stage 1, participants were shown a MyData-related scenario. Given the four parameters and their unique levels, we had a total of 96 unique permutations available as scenarios to rate (2x4x3x4). We asked the participants to provide a response to three different items after observing one scenario at a time:

(1) The perceived ethicality of the scenario
(2) Which of the parameters most affected their choice of ethicality
(3) The degree of personal concern, given the overall conduct in the scenario

Ethicality and degree of concern were collected using a scale from 1-5, and the choice of parameter that most affected ethicality was done with a radio button group. We used Amazon's crowd-HTML elements to implement the task so that it was presented

Table 1: Parameters (column headers) and levels (rows) in the study. Unique combinations of levels (N=96) constitute the assessed data management scenarios.

MyData operator action	Data type
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)
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)
	Location data
	Personal media files (pictures and video)
Purpose	**Consent**
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
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
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 not aware of having granted such permissions
	The action was taken against your signed consent, but legal agreements in your country allowed for the actions to take place

directly inside the MTurk worker interface, as this yields a seamless working experience for the workers. We set up the task so that each participant would always see a set of different permutations by using MTurk's batch functionality via a .CSV file input. As a result, a participant may assess several scenarios, but on each occasion, the participant is presented with a different permutation of the total 96 available ones. The source code for the designed instrument is available on *GitHub* [1] and can be assessed publicly.

Participants were rewarded with $0.15 for each of the tasks (*i.e.*, assessing one scenario). We invited 20 unique contributions per each of the available scenarios.

3.3 Stage 2

We invited all the participants from the first stage study to the second stage study. Not all invitees responded to the call to the second stage study with about 75% of them responding. The second stage study is complimentary to the first stage study and was aimed at supporting the robust quantitative results from the first study. It comprised of a questionnaire to 1) collect additional qualitative insights concerning MyData, and 2) obtain demographic information on our sample. The questionnaire was hosted in Google Forms and contained items as follows.

- Demographic details: age, gender, nationality, and highest academic qualification.
- Familiarity with MyData: Likert-style item from 1 (not at all familiar) - 5 (extremely familiar) of participants' prior knowledge of MyData.
- Checklist of devices and applications participants use: activity trackers, smartphones, and smartphone applications for health tracking.
- How much is your personal data being abused: Likert-style item 1-5 identifying perceived degree of one's personal data being abused unethically by a third party.
- Elaborate on the previous question. Why?
- Open-ended fields to discuss participants' excitement about having control of their own data, what concerns they may have about MyData now and in the future, as well as provide insights into who they will grant access and usage rights to their data to.
- Standardized ethics position questionnaire, the EPQ [22].

[1] https://github.com/alorwu/mydata-ethicality-tool

4 RESULTS

4.1 Participant Overview

Our initial study – Stage 1 – was completed with contributions from 172 workers who together contributed all the available 1920 scenario assessment tasks (20 submissions per each of the 96 available permutations). Based on the sample (N=129) who completed Stage 2, the mean age of our sample was 32.71 years (SD = 7.17 years). Of the Stage 2 participants, 97 were men, 31 women, and 1 non-binary. The participants had diverse academic qualifications with 65 holding a bachelor's degree, 36 with a master's degree, 17 holding high school diploma, 9 professional degree, 1 associate degree, and 1 music conservatory degree. Participants' nationalities were equally diverse with highest being India (66), USA (38), Brazil (5), Italy (4), and some identifying as Asian (3). Others include Britain (2), Spain (2), Ukraine (2) and one participant each from Bulgaria, Canada, Germany, Ireland, Kenya, Nigeria, and Pakistan.

We inquired about participants' familiarity with the term *'MyData'* using a single 1-5 scale (1 - not at all familiar to 5 - extremely familiar). The mean value was 3.37 (SD: 1.42). We also probed people's position on how much they believed their personal data was being currently unethically used by any third party, using a similar scale from 1-5 (1 - not at all much to 5 - extremely much). Here, the mean was 3.53 (SD: 1.10).

Finally, we used the Ethics Position Questionnaire (EPQ) to describe the ethical ideology of the sample [22]. The EPQ has been applied in various research notably in social psychology [23] and business and marketing [9, 37], and it assesses the degree of idealism and rejection of universal moral rules in favor of relativism. The EPQ questionnaire consists of 20 items to measure the extent to which respondents agree or disagree to each of the statements (1 - completely disagree to 9 - completely agree). Thus, people can score between 10 and 90 on the ideology and relativism scales. Our participants scored an average of 67.80 (SD=14.87) on the ideology scale, and 61.22 (SD =15.49) on the relativism scale.

4.2 Quantitative Analysis

The ethics position of the participants had practically no effect on the perceived ethicality or felt concern of the scenarios, with the exception of relativism and perceived ethicality being weakly correlated, $r(1701) = .28$, $p < .001$. The remainder of this section unpacks data on the perceived ethicality and the felt concern of the presented MyData scenarios.

Table 2: Summary of the fitted ordinal logistic regression model for participants' responses on the ethicality of the presented scenario.

Coefficients – Ethicality:

	Estimate	Std. Error	z value	Pr(>|z|)	
MyData operator action (sell)	-0.306	0.096	-3.182	0.002	**
data type (health tracker)	0.177	0.137	1.297	0.195	
data type (location data)	0.202	0.136	1.488	0.137	
data type (media)	-0.391	0.137	-2.853	0.004	**
purpose (for-profit)	-0.294	0.118	-2.490	0.013	*
purpose (government)	0.150	0.118	1.277	0.202	
consent (against consent)	-0.267	0.136	-1.963	0.050	*
consent (within but confusing)	0.221	0.133	1.668	0.095	
consent (within consent)	1.028	0.139	7.391	<0.001	***

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

4.2.1 Effect of Parameters to the Perceived Ethicality. We built a regression model (Cumulative Link Mixed Model fitted with the Laplace approximation) using the 1920 individual assessments obtained in Stage 1 using the R package *ordinal* [2], modeling the effects of the four parameters on the indicated ethicality score. We also included the participant ID as the mixed effect. Table 2 depicts the model for how the different variables affect the perceived ethicality of the scenarios.

We observed statistically significant effects of the different parameter levels on the perceived ethicality of the scenario, most notably a strong positive effect of *within consent* as well as the negative effects of *media* as the data type and *sell* as the MyData operator action. We then also conducted a least square means analysis with Tukey HSD adjustment (using the R package lsmeans [3]) to investigate the pairwise relations of the levels of each parameter and discovered significant statistical differences as follows. The two MyData operator actions, *shared* and *sold*, differed significantly (p<0.01); concerning data types, *health trackers* and *media data* differed (p<0.001), *location data* and *media data* differed (p<0.001), and *diagnostic data* and *media data* differed (p<0.05); concerning purpose of the data, *for-profit* and *government use* differed significantly (p<0.001) as well as *academic research* and *for-profit* (p<0.05). Finally, concerning consent, we observed significant differences between *against but legal* and *within consent* (p<0.001), *against consent* and *within but confusing* (p<0.001), *against consent* and *within consent* (p<0.001), and also with *within but confusing* and *within consent* (p<0.001).

4.2.2 Effect of Parameters to the Felt Concern. Table 3 depicts the model for how the different variables affected the degree of how concerning overall the scenarios were perceived. Interestingly, the results differ substantially from the results concerning ethicality, indicating that the two variables are decoupled: What is unethical may not necessarily be concerning and vice versa. Here, we observe that managing consent the correct way, i.e. *within consent*, as well as *health tracker* and *location data* being in question significantly decrease user concern. Media data, for instance, had no observable effect. We will talk about this decoupling more later in the discussion.

We conducted a least square means analysis comparing pairwise the different parameter levels in the context of how concerning

[2]https://cran.r-project.org/web/packages/ordinal/index.html
[3]https://www.rdocumentation.org/packages/lsmeans/versions/2.27-2/topics/lsmeans

Table 3: Summary of the fitted ordinal logistic regression model for participants' responses on the perceived concern about the presented scenario.

Coefficients – Concern:

	Estimate	Std. Error	z value	Pr(>|z|)	
MyData operator action (sold)	0.090	0.096	0.939	0.348	
data type (health tracker)	-0.407	0.136	-2.999	0.003	**
data type (location data)	-0.514	0.136	-3.777	<0.001	***
data type (media)	0.202	0.138	1.467	0.142	
purpose (for-profit)	0.402	0.119	3.377	0.001	***
purpose (government)	-0.024	0.117	-0.209	0.835	
consent (against consent)	0.073	0.136	0.532	0.595	
consent (within but confusing)	0.036	0.134	0.264	0.792	
consent (within consent)	-0.754	0.137	-5.513	<0.001	***

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

Table 4: 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.

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

the conduct was perceived. Our results again reveal statistically significant differences between the levels, as follows. The two levels of MyData operator actions did not differ significantly, but concerning data types, we observed that *diagnostic* and *health tracker* differed ($p<0.05$), *diagnostic* and *location data* differed ($p<0.001$), *health tracker* and *media* differed ($p<0.001$), and *location data* and *media* differed significantly ($p<0.001$). Concerning purpose, we discovered a significant difference between *academic research* and *for-profit* ($p<0.01$), and between *for-profit* and *government* ($p<0.001$). Further, concerning consent, we observed a significant difference between how *against but legal* and *within consent* ($p<0.001$) affected the overall concern, as well as between *against consent* and *within consent* ($p<0.001$), and finally also between *within but confusing* and *within consent* ($p<0.001$). Again, the pairwise directions of these differences are best observed in the overall model, depicted in Table 3.

4.2.3 Variance Between Scenarios. Both the ethicality response and the concern response data are not normally distributed according to the Shapiro-Wilk test on normality ($p < .0005$). The response variance for ethicality was 1.88 with standard deviation of 1.37 and variance for concern was 1.15 with standard deviation of 1.37. One-way analysis of variance (ANOVA) shows significant ($p < .005$ with 95 degrees of freedom) differences on both the ethicality response ($F = 1.539$) and the concern response ($F = 1.931$) when considering the variance across all 96 parameter combinations (scenarios). Although ANOVA is not recommended for non-normal data it can tolerate non-normal data with only a small effect on the Type I error rate. This verifies that the different scenario types indeed had an influence on the participant responses, and thus this also validates our study design as successful.

Finally, we asked the users to choose the one parameter per scenario that most affected their choice of ethicality. Table 4 lists the counts as well as the proportions of the levels of the parameters that were the most decisive in the participants' choice of ethicality. We note that while the comparison between parameters is not directly applicable, as they had different amount of levels, the inter-parameter comparisons are interesting here as they reveal which levels were mostly used in people's decision-making processes.

4.3 Qualitative Results

We analysed the 129 responses in stage 2 following a deductive thematic analysis process [13, 20] based on each of the individual questionnaire items. In this analysis method you derive predetermined aspects from questionnaire responses. From each of the questionnaire items, the main author first identified the most informative responses. Then three of the paper's authors got together virtually to identify aspects from these responses that would be the most interesting to present and discuss in each question's context. In the following sections, we will present the findings from four of these questionnaire items.

4.3.1 Potential of MyData. Participants expressed enthusiasm and hope about the MyData vision where they would be the owners and controllers of their data. One participant duly noted, "*I am excited about not giving my data up to third-party advertisers, where I have no control over who gets to see my data or how it is used.*" (34, Nonbinary) while others expressed optimism of being in total control of their data, "*I like the prospect of being in total control of data which I never would have had access to before*" (Male, 25), "*I love the idea of being in control of my own data, it feels like freedom*" (26, Male), and "*Having control over the access of MyData and keeping updated about my health status*" (29, Female). Another participant agrees with this assessment, refusing to get too excited about the prospect of this happening "*Not much excitement but the changes in the right direction gives some excitement*" (45, Male). Looking beyond having control of one's data, one participant also identified a unique opportunity for monetizing one's MyData, "*...most exciting would be an ongoing paid opportunity to disclose more if I wanted to*" (35, Female). One participant however expressed a great deal of hope for the future of MyData; that MyData could become a means of capturing our entire livelihood and a visualization of this vast amount of data (especially media) could enable us relive the past. The participant noted, "*Data that could make the time stop for nostalgic purposes, like having lenses that capture and maintain all of our days in a google photos alike website. A data to keep all of our audios too, where we could remember that voice of a deceased mother*" (45, Male).

4.3.2 Concerns and Doubts. With regards to concerns that participants have about the MyData vision, it was clear that the optimistic ethos did not really alter their view of existing privacy issues and concerns – and naturally we did not even expect it to. Participants were doubtful of whether it is even possible to have control of their data in a world where security and privacy are potentially at the hands of malicious actors and hackers, "*I am afraid of hackers, they*

know their way to steal data and I'd never want my record to be public even if I am not a celebrity. I care about my privacy and I know I am not totally in control when hackers are around" (40, Female). One participant was much more concerned with how to secure the data from third parties, stating, "*Security and privacy of my data and saving this from the third party*" (45, Male). Others were concerned about how much control they can ever actually realistically have "*The main concern is that to what extend our data is actually in our control*" (29, Male) and "*It could fall into the wrong hands*" (34, Male). One participant expressed concern about why such an initiative to hand over control of one's data to the individual is not being done by the government but a company (referring to the MyData operator in the scenario) "*I don't like how having control of my data has to be done by a company instead of the government*" (25, Male).

4.3.3 Perceptions on Current Data Abuse. When asked about the current state of data abuse, participants expressed rather clear overarching frustration. One participant notes, "*I think privacy is bs today, they use my data as much as they want, they make profits and I am the one who is constantly bummed by unwanted phone calls trying to sell me trading online services and all sorts of scams. That happens because companies sell my data and the people calling me know everything about me, it freaks me out on a daily basis*" (40, Female). Others based their assessment on news items, noting, "*There are some breach of data being said in news all over the world. Cases where our email is leaked in relationship sites*" (26, Male) and "*I track data breaches every day and the amount of them happening every day, all the time, is ALARMING. There's just no way some of my data from one app or another isn't for sale. Even without that, I think it is unethical for my data to be used in advertising (even if I have to allow it to use an app or service) to a third party*" (34, Nonbinary). Some participants were doubtful of their data being abused by a third party as one stated, *"I don't think any 3rd party can use my personal data since my device has been safe"* (31, Male). Another one echoed similar sentiments, *"Because I think most of my data are kept secret"* (23, Male). One participant was however unwavering in his believe that his data was being abused unethically by a third party due to past experiences, *"I am absolutely certain my data is being used unethically because I have been in several major hacks"* (31, Male).

4.3.4 MyData Vision in Health Care. We asked participants what their thoughts were on the notion of MyData particularly within the health domain. The majority of the participants expressed an overwhelmingly positive response to this. As some participants noted, *"I think it's a great idea. I want to be able to control my data"* (35, Male), and *"Health-related data is one of the most important types of data and I'd love to be fully in control when it comes to it"* (30, Male). Another participant expressed similar ideas, *"I believe it is perfectly in order for every individual to have absolute right to monitor and as well approve of who uses their data"* (46, Female).

Issues of access control were central in this category. For instance, the accessibility of health information to healthcare providers can be a crucial element to the delivery of healthcare. This understanding was shared by participants, *"If my health related data is easily accessible to my healthcare provider and my vital parameters from a health tracker can be monitored by them in real time, it could potentially help them to preempt an adverse health condition that might cause irreversible damage to my body"* (38, Male) and *"The benefit is that this information can be used by health professionals to access medical needs quickly. Hopefully to identify and prevent medical emergencies quickly"* (49, Male). Or, *"I think that it is a move in the right direction. For medical professionals and hospitals to have access to my medical data is a good thing and would enable better medical care"* (49, Male). One participant also pitched the idea of giving the government access to the personal health data, *"This can be used by government and health organisations to improve the overall quality"* (59, Male). Contributing one's data towards research was also highlighted, *"I would want my data to be anonymised and used for medical research but my consent must be taken first"* (29, Female).

Several participants saw potential in how MyData could be used to promote health and wellbeing considering the current health crisis caused by COVID-19. One participant notes, *"Data being used to help understand viruses like COVID-19..."* (21, Male). Another participant sees a broader picture of how continuous access of his data to medical staff can help them in early identification of potential health issues stating, *"The possibility that MyData can be used to help prevent and identify health issues"* (49, Male). Medical diagnosis and treatment could be better personalized based on MyData as suggested by one participant, *"I am sure that this data will help me in the future if my health deteriorates. Health specialists will be able to find solutions based on the data"* (31, Male).

Even if data is controlled by the individual, technically, some organization will manage the data – somebody has to maintain the infrastructure. To this end, participants voiced certain concerns about who would be a worthy custodian of their data: *"I do not believe that a private company should keep my medical data, because a change in that company or that in the future it was acquired by another one could lose control of my own data. Medical data must be hosted by a public entity, such as a universal public health system that has no financial interest in my data."*

5 DISCUSSION

In this paper we investigated the effect of different parameters in data related scenarios concerning the perceived ethicality and the felt concern about the overall conduct. We also inquired about the potential of MyData in this context as a potential way forward in data management. This is important because such end-user-centric investigations are needed to shed more light on user perceptions on data practices [21]. Further, given the prevalence of scandals in this domain, user attitudes toward any improvements – be they however optimistic and complex at this point – will inevitably affect their realisation and ultimate acceptance.

Most ethical issues about data are related to data collected about people – personal data [26]. MyData is an emerging vision for ethical data management that has not yet received the attention in the field of HCI as it has for instance in the medical field [38]. Yet, it is not a far-fetched utopia, but rather firmly rooted in a commonsensical notion of one having more control over one's data. As such, it would also offer an additional layer of support to the General Data Protection Regulation (GDPR) in terms of data protection and control. In the case of GDPR, data subjects are given *"greater control"* and will have easier access to *"their own"* data while MyData aims

to take this a step further, by providing them with direct access, control, or even ownership of their data.

5.1 Data Management Perceptions

Looking at the results of our study especially on how the common parameters rooted firmly in related work or modeled after real-world incidents affect perceived ethicality and concern, it is clear that data management perceptions can and should be studied on a granular level. In our study, we included a handful of common levels per parameter (see Table 1) but could still detect a total of nine levels that were either negatively or positively associated with a change in the perceived ethicality or concern.

Related to this, studies have shown that one of the Internet's biggest modern misconceptions has been that clicking "Agree" for consent protects one's data from abuse [55]. Trustworthiness must be shown at all levels of the user-operator relationship. Merely informing data producers of using cryptographic and high-end security infrastructures to secure their data means nothing to them. To this end, we also clearly noted in the qualitative analysis the frustration of participants regarding the felt current dubious or malicious practices by the current companies managing their data.

Curiously, our results also show that ethicality and the concern indeed are different things to begin with. This can be at least partially related to a degree of 'online apathy' that people feel – while they know that something is not being done correctly, they do not care as it is seen as inevitable [27]. In our findings, for instance, the top three parameters affecting ethicality of the scenarios were acting within consent, media data, and the MyData operator action of selling data (see Table 2), while the top three parameters affecting the overall felt concern were acting within consent, location data, and for-profit as the purpose (see Table 3).

5.2 Implications

The quantitative results of our work concretely inform academic work through investigating the relative effect sizes – user-centric perceived importances – of the different factors that are all coupled in related work or real-world practices on the perceived ethicality and the felt concern of realistic data management scenarios. To the best of our knowledge our work is the first to report on these relative effect sizes of the factors that are commonly reported anecdotally in interview studies. To exemplify this, in our study *e.g. within consent* affected ethicality approximately five times more than *within consent but confusing*.

The qualitative results in our work highlight a level of distrust towards the companies involved in the collection and processing of personal data. While this is perhaps not surprising in the light of both recent academic literature [7, 74] and the numerous cases of data misuse that were brought to light, it reveals the Achilles heel of initiatives such as MyData. If MyData or any other similar initiative is to make solid advances for instance in the healthcare domain, users have to develop trust in all stakeholders in the data management chain, including data operators. To this end, MyData operators should potentially aim to not only fulfill legal and ethical considerations, but also address the concerns raised by (potential) users. This is also connected to another related issue: Crafting disclaimers and consent forms. The inter-parameter importances (see Table 4) provide clues to what aspects are important to people. Thus, while clear articulation of data use policies is already pervasively required, prioritizing the aspects that people find as most decisive in their internal accounting will be useful in further improving the legibility of such documents.

Zou & Schaub showed that even when users have concerns, they might not know how to take appropriate actions [74]. In addition to lack of knowledge, some of our participants even presented a defeatist perspective and highlighted that they do not feel in control of their data. Therefore, data operators should provide users the ability to directly rescind any ongoing data sharing agreements. Further, our findings are in line with previous research, showing that design of systems that help users to understand how their data is used and by whom (see [7, 74]), supported with education related to data collection and use [7, 65, 74], is needed. In addition to that, we need to develop systems that help users to gain back the declining trust on their data use. A question is, however, how to develop those? Initiatives such as MyData are not optimally efficient if people do not trust the data operators, but instead ask, *quis custodiet ipsos custodes* – who watches the watchmen?

Finally, the inherent potential of regression models like this is what they enable when developed further into production and beyond the scope of an individual research paper. We need not take many steps to use the same underlying data to build ethicality prediction machines for data management. These could be used for speculating the acceptability of data management practices, and, on the other hand, for investigating the tradeoffs on how different data management decisions that might be otherwise beneficial for a company's operations ultimately degrade the perceived user experience. Such an approach can provide useful insights to co-operations, but may also enable commercial organisations to push the envelope until just before an ethical border is crossed.

5.3 Future Work

To scope the work, we left out several other interesting scenario ideas (consisting of parameters and levels) that are worth investigating in future studies. Some of the ideas we brainstormed during the study design stage include non-profits[72], user profiling (metadata derived from raw data), temporal dimensions of data collection, and data access by family members.

While this line of investigation deals with people's perceptions of scenarios, we are also interested in examining similar issues of ethics and data management in a more realistic setup so that participants would actually make the calls on real data that they have produced on their own and are given full governance over. In other words, we are building a data collection setup that allows users to collect data and then make choices about its sharing.

5.4 Limitations

We acknowledge limitations in our work. First, our results originate from MTurk and thus while the results are certainly indicative of broader trends, they do not naturally generalize over the general population. Similar to samples from most MTurk studies, the bulk of our participants come from India and the US. However, results from such online marketplaces have been valuable to research and the external validity is high, even in empirical research [58, 70].

Further, we did not include idealism and relativism in the models since the data for analyzing the idealism and relativism come from only a subset of the participants (75%). Therefore, including such data in the models would deteriorate their quality and predictive power.

6 CONCLUSION

In this paper, we investigate people's reactions to various data management scenarios. Our study was conducted in two parts: first, we crowdsourced assessments from Amazon's Mechanical Turk. Second, we conducted a complementary qualitative study to collect further insights. Our quantitative results revealed differences in how the selected parameters affect people's perceptions on ethicality as well as felt concern. Our qualitative analysis revealed concerns and doubts about MyData, perceptions on current data abuse, and the potential MyData has especially in health care. Put together, we hope our findings can act as a solid conversation starter within the CHI community as well as the field of data ethics as an exhaustive end-user-centric exploration to data management ethics and concerns.

ACKNOWLEDGMENTS

This research is connected to the GenZ strategic profiling project at the University of Oulu, supported by the Academy of Finland (project number 318930), and CRITICAL (Academy of Finland Strategic Research, 335729). Part of the work was also carried out with the support of Biocenter Oulu, spearhead project ICON.

REFERENCES

[1] Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies*. Springer, Berlin, Heidelberg, Cambridge, United Kingdom, 36–58.

[2] Eunice Eno Yaa Frimponmaa Agyei and Harri Oinas-Kukkonen. 2020. GDPR and Systems for Health Behavior Change: A Systematic Review. In *International Conference on Persuasive Technology*. Springer, Aalborg, Denmark, 234–246.

[3] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your location has been shared 5,398 times! A field study on mobile app privacy nudging. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Association for Computing Machinery, Seoul, Republic of Korea, 787–796.

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

[5] Ross Anderson. 2015. The collection, linking and use of data in biomedical research and health care: ethical issues.

[6] Mark Andrejevic. 2014. Big data, big questions| the big data divide. *International Journal of Communication* 8 (2014), 17.

[7] Mehrdad Bahrini, Nina Wenig, Marcel Meissner, Karsten Sohr, and Rainer Malaka. 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*. Association for Computing Machinery, Glasgow, Scotland UK, 1–6.

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

[9] Tim Barnett, Ken Bass, and Gene Brown. 1994. Ethical ideology and ethical judgment regarding ethical issues in business. *Journal of Business Ethics* 13, 6 (1994), 469–480.

[10] Ines Baudendistel, Eva Winkler, Martina Kamradt, Sarah Brophy, Gerda Längst, Felicitas Eckrich, Oliver Heinze, Bjoern Bergh, Joachim Szecsenyi, and Dominik Ose. 2015. The patients' active role in managing a personal electronic health record: a qualitative analysis. *Supportive Care in Cancer* 23, 9 (2015), 2613–2621.

[11] Maged N Kamel Boulos. 2003. Location-based health information services: a new paradigm in personalised information delivery. *International journal of health geographics* 2, 1 (2003), 2.

[12] Alex Bowyer, Kyle Montague, Stuart Wheater, Ruth McGovern, Raghu Lingam, and Madeline Balaam. 2018. Understanding the family perspective on the storage, sharing and handling of family civic data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Montreal QC, Canada, 1–13.

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[14] Katrina Brooker. 2018. I was devastated": Tim Berners-Lee, the man who created the World Wide Web, has some regrets. *Vanity Fair* 1 (2018).

[15] Michael Chmielewski and Sarah C Kucker. 2020. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science* 11, 4 (2020), 464–473.

[16] Andy Crabtree and Richard Mortier. 2015. Human data interaction: historical lessons from social studies and CSCW. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*. Springer, Oslo, Norway, 3–21.

[17] Wheidima Carneiro de Melo, Eric Granger, and Abdenour Hadid. 2019. Combining global and local convolutional 3d networks for detecting depression from facial expressions. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, Lille, France, 1–8.

[18] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 238–247.

[19] T Selwyn Ellis and David Griffith. 2000. The evaluation of IT ethical scenarios using a multidimensional scale. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 32, 1 (2000), 75–85.

[20] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[21] Casey Fiesler and Blake Hallinan. 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*. Association for Computing Machinery, Montreal QC, Canada, 1–13.

[22] Donelson R Forsyth. 1980. A taxonomy of ethical ideologies. *Journal of Personality and Social psychology* 39, 1 (1980), 175.

[23] Donelson R Forsyth. 1985. Individual differences in information integration during moral judgment. *Journal of Personality and Social Psychology* 49, 1 (1985), 264.

[24] MyData Global. 2020. MyData: Applying human-centric principles to health data. *Medical Writing* 29 (2020), 64–69.

[25] Timothy R Graeff and Susan Harmon. 2002. Collecting and using personal data: consumers' awareness and concerns. *Journal of consumer marketing* 19, 4 (2002), 302–318.

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

[27] Eszter Hargittai and Alice Marwick. 2016. "What can I really do?" Explaining the privacy paradox with online apathy. *International journal of communication* 10 (2016), 21.

[28] Drew Harwell. 2019. Is your pregnancy app sharing your intimate data with your boss? https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/?arc404=true

[29] Toby Helm. 2019. Patient data from GP surgeries sold to US companies. https://www.theguardian.com/politics/2019/dec/07/nhs-medical-data-sales-american-pharma-lack-transparency

[30] Joanne Hinds, Emma J Williams, and Adam N Joinson. 2020. "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies* 143 (2020), 102498.

[31] Jeff Howe. 2006. The rise of crowdsourcing. *Wired magazine* 14, 6 (2006), 1–4.

[32] Matthew B Hoy. 2016. Personal activity trackers and the quantified self. *Medical reference services quarterly* 35, 1 (2016), 94–100.

[33] Luke Hutton and Tristan Henderson. 2017. Beyond the EULA: Improving consent for data mining. In *Transparent Data Mining for Big and Small Data*. Springer, Cham, 147–167.

[34] Raimondo Iemma. 2016. Towards personal data services: a view on some enabling factors. *International Journal of Electronic Governance* 8, 1 (2016), 58–73.

[35] Gang-Hoon Kim, Silvana Trimi, and Ji-Hyong Chung. 2014. Big-data applications in the government sector. *Commun. ACM* 57, 3 (2014), 78–85.

[36] Jeongeun Kim and Hyeoun-Ae Park. 2012. Development of a health information technology acceptance model using consumers' health behavior intention. *Journal of medical Internet research* 14, 5 (2012), e133.

[37] Susan Bardi Kleiser, Eugene Sivadas, James J Kellaris, and Robert F Dahlstrom. 2003. Ethical ideologies: Efficient assessment and influence on ethical judgments of marketing practices. *Psychology & Marketing* 20, 1 (2003), 1–21.

[38] Timo Koivumäki, Saara Pekkarinen, Minna Lappi, Jere Väisänen, Jouni Juntunen, and Minna Pikkarainen. 2017. consumer adoption of future MyData-based

preventive eHealth services: an acceptance model and survey study. *Journal of medical Internet research* 19, 12 (2017), e429.

[39] Patty Kostkova. 2018. Disease surveillance data sharing for public health: the next ethical frontiers. *Life sciences, society and policy* 14, 1 (2018), 16.

[40] Patty Kostkova, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, Peter Knight, Corinne Marsolier, Rachel A McKendry, et al. 2016. Who owns the data? Open data for healthcare. *Frontiers in public health* 4 (2016), 7.

[41] Lindah Kotut, Timothy L Stelter, Michael Horning, and D Scott McCrickard. 2020. Willing Buyer, Willing Seller: Personal Data Trade as a Service. In *Companion of the 2020 ACM International Conference on Supporting Group Work*. Association for Computing Machinery, Sanibel Island, Florida, USA, 59–68.

[42] Jenny Krutzinna, Mariarosaria Taddeo, and Luciano Floridi. 2019. Enabling posthumous medical data donation: an appeal for the ethical utilisation of personal health data. *Science and Engineering Ethics* 25, 5 (2019), 1357–1387.

[43] Jaron Lanier. 2014. *Who owns the future?* Simon and Schuster, New York, NY.

[44] Jean-Rémi Lapaire. 2018. Why content matters. Zuckerberg, Vox Media and the Cambridge Analytica data leak. *ANTARES: Letras e Humanidades* 10, 20 (2018), 88–110.

[45] Tuukka Lehtiniemi. 2017. Personal data spaces: An intervention in surveillance capitalism? *Surveillance & Society* 15, 5 (2017), 626–639.

[46] Tuukka Lehtiniemi and Jesse Haapoja. 2020. Data agency at stake: MyData activism and alternative frames of equal participation. *new media & society* 22, 1 (2020), 87–104.

[47] Jens-Erik Mai. 2016. Big data privacy: The datafication of personal information. *The Information Society* 32, 3 (2016), 192–199.

[48] Ellen B Mandinach, Brennan M Parton, Edith S Gummer, and Rachel Anderson. 2015. Ethical and appropriate data use requires data literacy. *Phi Delta Kappan* 96, 5 (2015), 25–28.

[49] Ernst Marcia. 2019. Case Studies: High-Profile Cases of Privacy Violation. https://www.sgrlaw.com/ttl-articles/case-studies-high-profile-cases-of-privacy-violation/

[50] Andrea Martani, David Shaw, and Bernice Simone Elger. 2019. Stay fit or get bit-ethical issues in sharing health data with insurers' apps. *Swiss medical weekly* 149, 2526 (2019), 1–8.

[51] Matthew S McCoy, Steven Joffe, and Ezekiel J Emanuel. 2020. Sharing Patient Data Without Exploiting Patients. *Jama* 323, 6 (2020), 505–506.

[52] Morgan N McCredie and Leslie C Morey. 2019. Who are the Turkers? A characterization of MTurk workers using the personality assessment inventory. *Assessment* 26, 5 (2019), 759–766.

[53] Donna M Mertens. 2014. Ethical use of qualitative data and findings. In *The SAGE handbook of qualitative data analysis*. Sage London, England, London, England, 510–523.

[54] Ghulam Muhammad, Mansour Alsulaiman, Syed Umar Amin, Ahmed Ghoneim, and Mohammed F Alhamid. 2017. A facial-expression monitoring system for improved healthcare in smart cities. *IEEE Access* 5 (2017), 10871–10881.

[55] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.

[56] Jonas Oppenlaender, Kristy Milland, Aku Visuri, Panos Ipeirotis, and Simo Hosio. 2020. Creativity on Paid Crowdsourcing Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu, HI, USA, 1–14.

[57] Niclas Palmius, Althanasios Tsanas, KEA Saunders, Amy C Bilderbeck, John R Geddes, Guy M Goodwin, and Maarten De Vos. 2016. Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering* 64, 8 (2016), 1761–1771.

[58] Jay Pedersen, David Kocsis, Abhishek Tripathi, Alvin Tarrell, Aruna Weerakoon, Nargess Tahmasbi, Jie Xiong, Wei Deng, Onook Oh, and Gert-Jan De Vreede. 2013. Conceptual foundations of crowdsourcing: A review of IS research. In *2013 46th Hawaii International Conference on System Sciences*. IEEE, Wailea, Maui, HI, USA, 579–588.

[59] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. 2000. Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing* 19, 1 (2000), 27–41.

[60] Antti Poikola, Kai Kuikkaniemi, and Harri Honko. 2015. Mydata a nordic model for human-centered personal data management and processing.

[61] Karpurika Raychaudhuri and Pradeep Ray. 2012. Privacy challenges in the use of eHealth systems for public health management. In *Emerging Communication Technologies for E-Health and Medicine*. IGI Global, 155–166.

[62] Sourav Dey Roy, Mrinal Kanti Bhowmik, Priya Saha, and Anjan Kumar Ghosh. 2016. An approach for automatic pain detection through facial expression. *Procedia Computer Science* 84 (2016), 99–106.

[63] Bart W Schermer, Bart Custers, and Simone van der Hof. 2014. The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. *Ethics and Information Technology* 16, 2 (2014), 171–182.

[64] Michel Schreiner and Thomas Hess. 2015. Examining the role of privacy in virtual migration: the case of WhatsApp and Threema. In *Twenty-first Americas Conference on Information Systems*. AIS eLibrary, Puerto Rico, 1–11.

[65] Ben Rydal Shapiro, Amanda Meng, Cody O'Donnell, Charlotte Lou, Edwin Zhao, Bianca Dankwa, and Andrew Hostetler. 2020. Re-Shape: A method to teach data ethics for data science education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu, HI, USA, 1–13.

[66] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Toronto, Ontario, Canada, 2347–2356.

[67] Jeffrey Soar, Ghazal Bargshady, Xujuan Zhou, and Frank Whittaker. 2018. Deep learning model for detection of pain intensity from facial expression. In *International Conference on Smart Homes and Health Telematics*. Springer, Cham, Singapore, Singapore, 249–254.

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

[69] Kaiwen Sun, Abraham H Mhaidli, Sonakshi Watel, Christopher A Brooks, and Florian Schaub. 2019. It's My Data! Tensions Among Stakeholders of a Learning Analytics Dashboard. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Glasgow, Scotland UK, 1–14.

[70] Melanie Swan. 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 (2012), e46.

[71] Mark Walport and Paul Brest. 2011. Sharing research data to improve public health. *The Lancet* 377, 9765 (2011), 537–539.

[72] Shannon K. Yavorsky and Kelly DeMarchis Bastide. 2018. Europe's New Data Law: What Nonprofits Need To Know. https://www.thenonprofittimes.com/npt_articles/data/

[73] Chaoqun Yue, Shweta Ware, Reynaldo Morillo, Jin Lu, Chao Shang, Jingbo Bi, Jayesh Kamath, Alexander Russell, Athanasios Bamis, and Bing Wang. 2018. Fusing location data for depression prediction. *IEEE Transactions on Big Data* (2018), 1–1.

[74] Yixin Zou and Florian Schaub. 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*. Association for Computing Machinery, Montreal QC, Canada, 1–6.