Futures for Health Research Data Platforms From the Participants' Perspectives

Giovanna Nunes Vilaza
Technical University of Denmark
Kgs. Lyngby, Denmark
gnvi@dtu.dk

David Coyle
University College Dublin
Dublin, Ireland
d.coyle@ucd.ie

Raju Maharjan
Technical University of Denmark
Kgs. Lyngby, Denmark
rajm@dtu.dk

Jakob Bardram
Technical University of Denmark
Kgs. Lyngby, Denmark
jakba@dtu.dk

ABSTRACT

In clinical cohort studies, researchers analyse the life history of population groups to understand the evolution of diseases. Health research data platforms came to facilitate such studies as they allow multiple projects to share access to cohorts' non-identifiable health information. Some latest initiatives are also starting to include mobile-generated data in their research programmes. Although seemingly beneficial, it is not yet clear how potential participants feel about contributing to the new platforms: there is a need to investigate potential factors related to the acceptance in this specific context. In this paper, previous works from related contexts were brought together and, along with a qualitative study, composed a participant-centred perspective of enablers and barriers for contribution. We found that there is an apparent misalignment between current implementations and participants' preferences, leading us to propose design guidelines for future developments which can make participation more ethical and engaging.

CCS CONCEPTS

• **Human-centered computing** → Empirical studies in HCI; • **Applied computing** → Health informatics.

KEYWORDS

health research; data platforms; personal sensing; acceptance; privacy; ethics.

ACM Reference Format:

Giovanna Nunes Vilaza, Raju Maharjan, David Coyle, and Jakob Bardram. 2020. Futures for Health Research Data Platforms From the Participants' Perspectives. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*, October 25–29, 2020, Tallinn, Estonia. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3419249.3420110>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

NordiCHI '20, October 25–29, 2020, Tallinn, Estonia

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7579-5/20/10.

<https://doi.org/10.1145/3419249.3420110>

1 INTRODUCTION

Clinical cohort studies are used to investigate the natural history of diseases. By analysing demographic, behavioural and health data in chronological order, clinical researchers look for correlations, causes and effects [88]. Some of the challenges of these studies are the recruitment of large samples and the costly collection of longitudinal data [135].

Recently, health research data repositories have emerged to facilitate this process. They consist of large-scale shared-access databases containing different types of information on patients and healthy individuals, allowing multiple researchers to conduct analytical studies without having to recruit new participants [36, 112, 141]. Examples of data stored in such repositories may range from Electronic Health Records (e.g. Danish Civil Registration System [128]) to daily habits and behaviours (e.g. Evidation [44]).

Even though such platforms aim at benefiting society and health-care, previous research in other contexts found that when it comes to sharing behavioural and sensitive data, people have several concerns. In the context of information sharing for health treatments, a benevolent end goal is not enough for patients to agree with disclosing every data type to any receiver [106]. Similarly, in the context of genetic and bio-repositories, individuals feel concerned about their DNA data being misused in the future even though they understand the need for research [96].

Given such evidence from other contexts, individual concerns may represent a risk for the future of the new platforms: lack of acceptance can deter potential participation and compromise cohort diversity [125]. While a large body of research has focused on willingness to share data with bio-repositories [39], genetic testers [70] and clinicians [69], there is a gap in research that focuses on the specific context of the emerging large-scale platforms and research programmes. The latest initiatives differ from bio-repositories because they contemplate the use of continuous behavioural monitoring, which means it is not only DNA being stored but a complete picture of one's daily life.

Therefore, there is a need to investigate how potential participants feel about contributing to shared-access repositories, especially those which include behavioural monitoring. Ethical, legal, and social issues related to the topic have been previously discussed [125], but not much has been considered in terms of the

participants' preferences, concerns and priorities. Data repositories for health research have seen a rise in the last years [33], and since the COVID-19 pandemic, the expectation is that they will become even more predominant worldwide [48]. However, the context in question is underexplored in Human-Computer Interaction (HCI) [70]. It is time to consider how participants want to be engaged in such initiatives [59], as a better understanding of their viewpoint will enable the design of acceptable features aligned with people's concerns and expectations.

The goal of this paper is to investigate the perspectives of those who are not yet participants in these large-scale platforms and inspect their willingness to share different types of data. In order to achieve this goal, an empirical study with young adults was conducted. The motivation to focus on young adults came from the observation that those under 40 years old are often less favourable towards contributing to research repositories [50] and expect more control over how data is used [133], which makes them a more challenging group in terms of acceptance. Through the empirical study, our goal was to tease out their reasons for a potential resistance so that future developments can take these factors into account.

Moreover, this paper also provided a comprehensive compilation of the literature related to acceptance, privacy and motivation, as well as a summary of representative examples of existing platforms. The goal of this extensive narrative review was to put together findings and theories from scattered contexts, and show how they provide the foundation for the understanding of the participation in the new initiatives. Such aggregation is a useful resource to ground future research as well. Finally, we also contribute to design guidelines which are put in contrast with the current platforms to demonstrate what can be done to better align with participants' preferences.

2 LITERATURE REVIEW

The review was conducted in two parts. First, to characterise the spectrum of existing platforms, a set of examples was identified. Some of them had their origin in traditional research, but others came from a commercial or non-profit context, which means they were not listed in academic search engines. Therefore, searches were conducted online, on Google Scholar, via Twitter. Even though not systematic, this was considered as the most appropriate and feasible approach, given the heterogeneous field of search. Platforms were selected and categorised according to their distinctive characteristics to compose a diverse and representative set. Different from other reviews [9], particular emphasis was given to repositories which offered more agency to its participants. Only platforms with details available in English were included, which could mean that some examples, operating locally in their own language, were not detected.

Second, previous works about behavioural monitoring and research participation were identified and reviewed. Google Scholar, ScienceDirect, PubMed and the ACM library were searched for a combination of the keywords: participation in research, open-access repositories, bio-banks, genetic data-banks, motivation, willingness to share, acceptance, trust, privacy, informed consent, behavioural monitoring, self-tracking, research engagement. Papers' reference lists were also consulted (snowballing). We aimed at including a

broad set of reference, in order to illustrate that many how different factors can influence participation, but we were also mindful of the quality of the publication venues. The choice of topics to highlight in the narrative emerged from several rounds of analysis and discussions among authors until the core elements relevant to the context were traced. Particular emphasis was placed on identifying key challenges for acceptance, willingness to share and consent. A summary of these can be seen in Table 1.

2.1 Health research data platforms

A health research data platform is a repository which stores health-related data from population groups, to be used for analytical studies, often by multiple projects [92]. The development of such repositories represents a significant advance on how research can be conducted. Studies using data from existing platforms can provide evidence about the relationship between behaviours and health outcomes in an efficient and less costly way [54]. These repositories have been implemented differently, and the involvement of participants also differs among them.

2.1.1 National initiatives. A well-established example of a national research repository is the Danish Civil Registration System, which holds all medical registers for the entire Danish population [128]. Large-scale cohort studies have been conducted using this data, without much involvement of the patients. Researchers who want to get access to the data need to submit their projects for internal approval. Similar centralised national-level approaches are under development in India [124] and China [80].

Countries that do not have centralised health records have invested in the recruitment of volunteers for bio-banks. The *UK Biobank*, by the Wellcome Trust, collected clinical data from 500,000 patients until 2010 [136]. Their web page shows studies approved and published [89], as well as summaries of the available data [16]. Researchers can request access to this data, but the involvement of patients ended after data collection.

A more recent initiative is the *All of Us Research Programme*, which aims at recruiting one million volunteers and contemplates the collection of mobile-generated data [125]. According to their protocol [107], they want to allow volunteers to use the data to keep track of their health. Access is meant to be provided only to a group of researchers selected by the Programme.

2.1.2 Data portals for researchers. Another type of repository gathers data from previously conducted clinical trials. For instance, the non-profit organisation *Vivli* [25] works as a 'neutral broker' by connecting institutions and companies who are willing to share results from past projects [15]. Each organisation has its own data-sharing policies and access conditions, but *Vivli* run de-identification algorithms to the personal data beforehand. Participants receive the benefits offered by the individual clinical trials but are not involved in the processes at *Vivli*. Other data portals have aggregated information about genes, proteins and published papers, like the European Union's COVID-19 Data Portal [117]. In these platforms, it is up to the researchers to share their data, not the participants.

2.1.3 Opt-in sharing. Individuals may start to collect data for themselves but agree to share it for research afterwards. For instance, *Achievement* [2] is a commercial mobile app for personal health

Acceptance and Trust	People trust some institutions more than others	[34, 39, 118, 132]
	Trust can be hindered by previous experiences	[3, 5, 34]
	Acceptance decreases if data collection is hard	[32, 71]
	Acceptance decreases if the purpose is not useful	[75, 78, 137, 152]
Willingness to share	People worry about data misuse	[78, 95, 96]
	People want to preserve their reputation	[65, 77, 111, 119, 121]
	Some types of data are more sensitive	[14, 47, 61, 86, 123, 146]
	Some people are more concerned than others	[8, 63, 86]
Consent and Ethics	Consent forms are lengthy and complex	[84, 110]
	People cannot understand the risks	[31, 94, 115, 131]
	Lack of flexible sharing options and control	[5, 27, 57, 64, 73, 91, 106]
	Lack of transparency about data use	[39, 99, 142]

Table 1: Challenges identified in the literature, which can represent barriers to participation in health research platforms.

tracking in which users earn points by logging daily behaviours, answering surveys and being passively tracked. They can opt-in to share this data with *Evidation* [44], a platform which uses the information for clinical studies. Before the surge of COVID-19, not many details could be found about the on-going projects of Evidation, besides a publication list, but currently, they post weekly updates for the public. A similar opt-in approach is also used by the DNA genetic testing company *23andMe* [70] and the social network *PatientsLikeMe* [149]. In none of these examples are participants in control of how the data is used once they give their consent to share it for research.

2.1.4 Participant-driven. In some platforms, the initiative to join an initiative and collect data for that comes from the participants themselves. An example is the community-based *Open Humans* [53], in which projects are posted, and members are invited to contribute with their Fitbit, Twitter or symptoms data. Members can also propose projects of their own, and share tools for self-experimentation. With the surge of COVID-19, organisers have been hosting online meetings as a way to open the discussion about pandemic-related projects. A similar example is the recent *Project Baseline*, from Alphabet (parent company of Google) [90]. Their projects range from virtual mood tracking to in-person bio-sample collection. Since the rise of COVID-19, they partnered with local governments to send tests and share results with Departments of Public Health - but so far, this data does not seem to be openly available. Besides such platforms which aggregate calls for several projects, there are also crowd-sourcing projects focused on health, also called 'participatory disease surveillance', in which individuals are invited to self-report symptoms and answer surveys for epidemiological control [103]. The COVID-19 emergency increased the number of such initiatives all over the world [48].

2.2 Behavioural monitoring and research participation

As it can be observed from the above overview, and from previous studies [29, 38, 47, 52, 85], smartphones and wearable data are starting to be considered for health research. Also called 'digital phenotyping', such data sources can provide information about users' environments [38], mental states [52], physical status [29] and daily habits [85]. This data can be collected passively (personal sensing) or manually through self-reports.

When it comes to using this data for health research, individuals are asked to contribute with their personal (and often sensitive) data for a purpose which at first may not directly benefit them - it is meant to help society. In this complex context, there might be a desire to contribute, but privacy concerns might hinder the willingness to share data. Besides, the goal of many of the platforms is to allow data to be reused across multiple projects. Within genetic data repositories, there are several discussions about how consent should be asked in this situation [39]. This section, therefore, reviews factors which are key in the context of data platforms storing health and behavioural information for research beyond the scope of a single project.

2.2.1 Motivation. In the context of participating in research, motivation is the force which drives individuals to contribute and may come from different sources. It is common to reward participants with coupons, lottery tickets and gift cards [76], which are external sources of motivation. Money payments is another potential source of motivation to share personal data [13]. However, individuals might feel more motivated if their behaviours are self-determined: they act because they believe in the inherent value of their action [35]. A common finding in related works about willingness to participate in research is that people are motivated by contributing to science [150], helping patients [49] and seeing the outcomes [39].

In order to enhance motivation, researchers have looked for ways to get participants to be more engaged in the process [21,

40, 46]. It has been discussed that participants should take part in agenda-setting, study design, recruitment, project evaluation [130], as well as decisions about the usage of the data [5]. By being more integrated with the project, they may feel more responsible for the results [7].

Getting test results is a primary motivation of those who send samples for genetic testing, such as 23AndMe [70] and previous works have pointed out to the use of personal data for reflection and behavioural change [45, 68, 82]. In addition to learning from the data collected, people can also feel motivated by comparing themselves to others [87], sharing their goals [30], and receiving feedback on their progress [51, 66]. However, there can also be negative consequences for using these as motivation drivers. When looking at their stats, some may experience anxiety [81, 147], feel guilty for not keeping up [42], lower their self-esteem [28, 100], and make incorrect interpretations [93, 109, 147]. Another issue is that when it comes to self-tracking, motivation may be necessary for keeping adherence to the data collection. It can be challenging for participants to stay consistent after joining a project [122], especially if they have to provide inputs frequently [26, 32].

2.2.2 Privacy. In the same way that motivation is an important enabler to participation in the research, privacy is a major barrier [74]. People worry about the potential for discrimination, undisclosed marketing interests, government surveillance and data used for generating profit [78, 95, 96]. Being constantly monitored can lead to feelings of surveillance [140], paranoia [127], and loss of freedom [24].

Privacy boundaries are often defined to protect internal values of autonomy, social acceptance and impression management [119]. In a study about students' perceptions on disclosing self-tracked data, they mentioned experiencing social pressure, stigma and embarrassment about their data [65]. The fear of a data breach is related to the potential damage to one's reputation and how others may react [111, 121]. By selecting which aspects of their lives to disclose, people can also control the image others have of themselves [77, 113].

However, willingness to share can depend considerably on the context [104], and there is a known mismatch between stated intentions and actual behaviours (privacy paradox) [105]. Therefore, even though privacy concerns may be a barrier to participation, in the context of research, individuals apply a 'usefulness heuristic': they prefer to share information that they know will be useful [83], and that they believe will not increase the risk for personal harm [121]. The Amazon Mechanical Turks were found to base their decision to share personal data for research on pay rate, requester, purpose, and perceived sensitivity of the request. However, they also engage in privacy-protective behaviours, such as abandoning tasks or providing inaccurate data [126].

Some factors can affect the intensity of privacy concerns. Previous research has found that more sensitive data types are disclosed less often [86, 123], such as audio recordings [37], browser history [14], message/phone logs and social media activity [47], camera pictures [121], financial information [146], home address [83], feelings of loneliness [61], sexually transmitted diseases [146], toilet use [12], and any health-related data in general [141]. Privacy is also perceived differently across individuals [22]. Personality traits

of agreeableness, conscientiousness, and openness to experience, can affect privacy perceptions [63]. Privacy fundamentalists are less likely to disclose information [86] as well as anxious individuals [8]. The benefits obtained from information disclosure are also a factor that can modulate the importance people attribute to privacy risks [70].

2.2.3 Acceptance. The decision to share data for research consists of weighing the motivation to contribute against the potential risks [83, 134]. It is clear that in the context of data sharing for research, motivation and privacy are critical components for acceptance. It is only when individuals 'accept' to contribute to a health research platform that their participation begins, but these acceptance should also be considered in the longer term, as it would be undesirable to have people dropping out [101].

The literature has extensively discussed factors for technology acceptance in general - usability and usefulness are known to be key factors for it [71]. In the context of participation in research, acceptance can increase if the burden of data input is low, which can be achieved by high usability [32, 57]. Regarding usefulness, people are more accepting of risks if they see their benefits, which can be for future patients [75], global causes [78, 152], researchers [137] or themselves [70]. However, usefulness can be highly contextual. Individuals may be less willing to participate if they think information is not relevant for the purpose [121, 123, 147]. On the other hand, individuals who are health-motivated [60], familiar with research [96] and who have disabilities [12] can be more accepting, because they understand how data can be useful.

The acceptance of information sharing with research repositories is also closely related to trust, especially on those who will keep and use the data [5, 118]. People may be willing to leave concerns aside when they trust those requesting their information [62, 79]. Scientific institutions are often trusted more than commercial, governmental and insurance institutions holding health information [39]. However, the development of trust is a continuous process, and trust can change over time [34]. Some enablers of trust in digital systems have been listed as fair data access, recommendation by others, customised design features, initial face-to-face contact, stakeholder engagement and improved communication [3].

2.3 Ethics and informed consent

It is well-known that before using any data from research subjects, the ethical approach is to obtain their consent, especially when data is collected through smartphones [55]. Subjects should also be allowed to withdraw from any study effortlessly [31]. The recent European General Data Protection Regulation (GDPR) established that in any data sharing procedure, people need to be at least informed about the information being shared, with whom and for what reasons [144]. However, consent forms are often overly complex [84], lengthy [110] and not efficient in informing participants [19, 58]. Many people do not understand what they are signing up for [145], how much information can be extracted from data [31], what data from sensors can reveal about them [94], how risky it is if someone identifies them [131], and how easily their identities can be revealed [115].

As a consequence, recent works have pointed to a desire for even more transparent and understandable information about data

usage [39, 99, 142]. Participants of genetic repositories have mentioned being interested in receiving more details about the goals of the projects [139], reports on privacy risks [145] and notifications about when data is used [114]. However, it is not trivial to provide an appropriate level of information without confusing users or making it a burden [43, 72]. The literature on how to improve the process of asking for consent is extensive, especially within research [41]. Suggestions have been made to go beyond textual forms and consider face-to-face encounters [4, 114], communication channels (e.g. chats and forum) [129], visual info-graphics [97] and interactive tools [64, 102].

Furthermore, it has been argued that people should be allowed more control over how their data is used [5, 27, 57, 64], such as the level of detail disclosed [91, 106], and how long data is retained [73]. It has been suggested that willingness to share may increase if users can choose what information is shared and to whom [1, 148], and customised settings can enhance trust [114]. User interfaces that allow a selective disclosure of data have been proposed as a way to support automatic sharing rules in order to reduce the burden of continually editing preferences [100].

However, from the review of the existing platforms, it could be observed that most of the large-scale repositories often rely on models of *broad consent* [120], which means that participants provide their consent to any future study using their data for an indeterminate period of time (e.g. All of Us, Evidation). This contrasts with the idea of *narrow consent*, which has been found to be a preferred option over broad consent in some cases [67]. This type of consent consists of consulting participants for each project using the data (e.g. Open Humans), but restricts the re-use across multiple researchers and the combination of data sources. A consent type in between narrow and broad is called the *tiered consent* [20], in which consent is only asked in certain situations, such as if studies vary significantly in nature. None of the platforms seems to follow this type of consent yet - such mid-term appears to be missing in this space.

3 EMPIRICAL STUDY

Our literature review identified enablers and barriers for participation in research and data sharing overall, but it could not identify qualitative works specifically on health research repositories that consider continuous behaviour monitoring. Moreover, the overview of the existing repositories identified different strategies to health data platforms, but it remains unclear if the expectations of participants match the available options. This empirical study is to explore individuals' attitudes and preferences regarding this scenario. The study consisted of semi-structured interviews and a card sorting activity with potential future participants who had not contributed to any research data repository yet.

3.1 Sample

A sample of 12 participants, aged between 19 and 26 (mean=23, std=1.8) was recruited through word-of-mouth. Recruitment stopped when data saturation was reached - no new themes emerged from the last two interviews. Participants were four females and eight males, from France, China, Spain, United States, Hungary, Netherlands, Italy and Scotland. One was a full-time developer, three were

part-time developers, three were doing internships (in wind energy, marine biology and photonics), two were exchange students, and two were finishing their Masters'. They reported being healthy in general, did not mention experiencing data breaches in the past, never been part of a clinical study, and four of them tracked physical activity occasionally. The study was exempt from ethical approval but was conducted according to accepted ethical standards (informed consent, benefit- not harm, confidentiality) and followed the national code of scientific conduct.

3.2 Study design

Semi-structured interviews and card sorting activities were conducted with each of the participants individually. All of the interviews had the same set of questions and cards (within-subjects design).

3.3 Materials

Before the interview, a set of cards was prepared. Each card contained a different 'information item' (see Table 2). The choice of items was based on the analysis of previous works about data sensitivity and willingness to share [12, 14, 37, 47, 61, 83, 86, 121, 123, 141, 146], data collected in the different platforms, and mobile sensing frameworks for behavioural monitoring [23]. Three medical doctors were consulted to confirm that the card set contained only data relevant to health research.

3.4 Procedure

In order to introduce participants about the concept of a health platform, on the day of the interviews, participants were first shown a 1 minute and 55 seconds video about precision medicine, made by the All of Us organisers¹. The short video did not provide any in-depth details about the platform; it was an introduction to the idea of having a data repository for health research. Participants were explained that the video was meant to be just an illustrative example.

Participants were asked to describe their initial opinions about health research platforms. Then, they were asked about potential reasons to contribute to and general concerns. Interview questions deliberately elicited both positive and negative perspectives to try to mitigate the bias of social acquiescence.

After these initial questions, participants were asked to do the card sorting. For each card, they were asked if they would feel comfortable with having the information depicted on the card stored in a repository. They were asked to place the cards in "yes", "no" and "maybe" piles and explain their decision. After the card sorting, they were asked what they would like to happen once their data had been shared and what they expected to hear from the platform. The interview ended with the collection of demographic information. Interviews lasted 45-70 minutes. They were conducted in-person (before the COVID-19 crisis), audio-recorded, and transcribed. The card sorting activity was documented in pictures. Data analysis was conducted using Thematic Analysis [17, 18], with themes being iteratively generated from the data collected (inductive approach). Transcripts were read several times, during which similar quotes

¹<https://www.youtube.com/watch?v=ti50nS7B5vI>

Calls/text content	Calendar	Financial status	Relationship	Address	Blood/urine
Mental illness	Grades	Genetic disorders	Emotional status	Family	Partner health
Toilet habits	Hygiene	Sleep patterns	Places visited	DNA	Contagious illness
Alcohol/ tobacco	Food	Local ambience	On-screen time	Distances	Physical exercises
Physical status	Leisure	Current weather	Physical illness	App	Heart rate

Table 2: List of information items (data types) shown as cards to participants.

were given codes, and groups of related codes were associated with themes.

4 FINDINGS

The interview study had the goal of investigating how young individuals perceive the idea of contributing to a health research repository and how they feel about sharing different data items. The result of the analysis showed that three main themes were recurrently discussed. These were: reasons to contribute to a research repository, characteristics of an ideal platform and the criteria for sharing the different types of information.

4.1 Motivations to contribute

When participants discussed the possibility of contributing to a health research platform in the future, they discussed the potential benefits of this choice. These fell under two categories: public and personal benefits.

4.1.1 Public benefits. Without exception, all of the participants praised the importance of health data repositories for society. Their initial comments often highlighted the usefulness of such idea: *“There could be a lot of good for a lot of people if healthcare providers could have more access to information about people’s lives”* (P3). Such altruistic motivation, of helping researchers and patients, was the first mentioned by the majority of the participants: *“I would share my data with the idea that it can help research and it can help other people to be cured”* (P9).

4.1.2 Personal benefits. Besides altruistic reasons, there was definite interest from the majority of the participants in receiving personal benefits as well, such as using the data for their own health improvement, which was mentioned often: *“If you have all the information about me, you can do more specific tailored care, maybe even like getting life recommendations”* (P3). There was also interest in comparing themselves with others: *“I would like to know how I compare with others. I feel like if I see others are using their phone significantly less time than I do, then I can control myself”* (P7).

On the hand, the prospect of receiving material compensations was rarely mentioned, and when mentioned, participants had opposing opinions. Some believed that there their contribution should be rewarded: *“Probably if it takes a lot of time, it should be rewarded, it should be paid”* (P1), but others did not: *“I don’t think I would ask for financial compensation, I think it is more of a volunteering type of work”* (P7).

4.2 Trust in the platform

Throughout the interviews, participants mentioned some key features a platform should offer to be trusted. Two main requirements were recurrent: transparency and choice.

4.2.1 Transparency. In order to verify if a platform presented the desired characteristics, participants mentioned several times the need for more transparency, so that they could make an informed judgement. They would like transparency about:

- *Who will use the data.* Most participants mentioned that they would like to know both who is the organisation behind the platform and who will be provided with access: *“I need to know who are the people behind the programme. Is it the university, is it an external research centre, is it for hospitals? And who is going to check my data?”* (P2).
- *For what purposes.* They also wanted to know precisely for which purposes the data would be used now and in the future: *“It is good not to share anything if we do not know how it can be used in the future”* (P11). The reason for that was because not all purposes were acceptable, as there was an evident objection to allowing data to be used for profit, mentioned by the majority: *“I’m happy to share this for health and public good reasons, but as soon as it starts to be for a private company, I feel like I am being taken advantage of”* (P3). Some of the participants would like to be notified about what was done with the data: *“I would be interested to know if I helped and what conclusions were drawn as a result”* (P8). This interest was not shared by everyone, however: *“I will give you my data, but I will never read the outcomes. For the people more interested, maybe”* (P12).
- *How data is secured and anonymised.* Evidence of cybersecurity measures was regarded as a significant factor for trusting a platform: *“If you feel that data is going to be leaked then I guess you wouldn’t share a lot”* (P5). One participant mentioned feeling reassured by being informed about the protocols used, even without fully understanding them: *“Knowing some technical stuff would make me comfortable even though I don’t understand that type of information”* (P4). The same importance was given for knowing that data is stored anonymously: *“If it is already anonymised in the database, then I would use the platform even if there was some kind of hack in the system because I know that I am just going to be one in the millions”* (P4).

4.2.2 Choice. Besides being thoroughly informed about the aspects above, a topic often mentioned by participants was the desire

to be able to actively keep control of the data during and after it was shared. Participants stated that they would like to choose who can have access and what data is available:

- *Selecting receivers.* Some of the participants would like to allow access to projects on a case-by-case basis: *"If they have new research and want to have access to my data, I want them to ask me if I want to release to the new research or not"* (P7). Not all of the participants agreed with this thought: *"I don't want to give the go-ahead for every researcher. There are probably too many requests, and I don't want to be bothered by this all the time"* (P5). As one of the participants mentioned, it would be interesting to be able to at least block the access for studies they do not agree with: *"I want to choose if I disagree with one of the studies and say: 'with this one I don't want to share'"* (P10).
- *Filtering the data.* Many participants mentioned that it would be good to stop the data collection on some occasions: *"Maybe in some situations is useful to stop recording some data. If you go to some places where you don't want to be followed, for example, you don't want people to know you are there"* (P2); and to be able to see what data is stored and reconsider the decision to give access: *"Once I have given the data, I think it would be nice to have a look at what is in the system, what data you saved and then have the option to maybe delete data points if I think it is not appropriate"* (P5). Additionally, many of them mentioned that they would feel more comfortable if researchers have access only to less detailed information: *"If it just to know what you are going to do in your day, how many hours you are walking, something like that, is fine"* (P10).

4.3 Willingness to share each data type

Participants often diverged opinions about which information they would feel comfortable collecting and sharing. Also, the same data type sometimes led to different types of concerns between participants. An overview of the different criteria for these decisions is listed next.

4.3.1 Relevance for the research. Having a clear understanding how the information is useful for health research was in many cases the reason why a participant would feel comfortable with sharing it: *"I would be happy to share that consumption information. For me, that has a very tangible connection to health"* (P3). On the other hand, not fully understanding the relevance of an item was mentioned by many as a reason to not share: *"I don't think my home address is relevant for the purposes of the research"* (P8). Participants did not always agree on the relevance of the different data items, with some being fine with sharing items that others would not see a reason for it.

4.3.2 Potential for material damage. When participants felt comfortable with sharing a particular data item, it often was because they could not see the potential for any adverse consequences: *"I don't think there is any harm in sharing your sleep patterns"* (P8). On the other hand, when participants said feeling uncomfortable, they mentioned fears related to physical attacks: *"If people want to locate me, it would be really easy. If I am going back home at this time, then someone is just waiting for me at the door and just breaking*

into my room" (P7); cyberattacks: *"Financial situation, no, because people maybe know how much money I have in my bank account and they would see if it is worthy hacking it or not"* (P7); and financial loss: *"If hackers are working for an insurance company, and they sell data to them, it could happen that they [insurance companies] propose different plans"* (P6). Not all of the participants associated these risks of material damage to location and financial situation - participants varied in their opinion about what information would bring each type of risk.

4.3.3 Risks for reputation. Some participants mentioned that there were parts of their life they prefer to keep private, because of the fear of how others perceive this: *"If it is not treatable then maybe it is something you want to keep it for yourself or share with only a limited amount of people"* (P5). Similarly, some mentioned being concerned about the possibility of being discriminated by their employers: *"I am afraid that in a potential data leak, my employer gains access and discriminate me in my employment"* (P11). However, participants often diverged on opinions about which data types they perceived as riskier for their reputation - their decision depended on their judgement of their own habits. Some participants said they would feel ashamed to disclose their drinking habits *"Alcohol, tobacco, I do not want people to know how much I drink because I drink too much"* (P8), but others said they were comfortable with it. Information perceived positively by society was often considered as more comfortable to share: *"Physical activities, yes, because I feel like is something positive"* (P6).

4.3.4 Invasion of others' privacy. Some participants were not comfortable with sharing data that did not belong to them: *"I think sharing calls and text content overlaps with the privacy of other people"* (P2). Sharing data from others was described as a betrayal by one of the participants: *"I do not want to share their data, it is kind of like going behind their backs"* (P4). They believed they were not in the position to decide: *"When it comes to the condition of my partner, it is tricky because I feel like it is their decision to share the information"* (P3). This concern was not mentioned by all of the participants, however.

4.3.5 The effort of the data collection. Sometimes, the issue was the burden of data collection. Most of the participants mentioned that collecting data can be an inconvenient responsibility: *"I don't want to share data when I'm low on my battery, or when I'm short of data. I wouldn't like this data to interfere with my daily life"* (P2); some also mentioned it can be emotionally difficult: *"If someone is really sad or depressed, I think it is harder to share that"* (P10). Some of the participants did not mention issues with data collection.

5 DESIGN GUIDELINES

Health research data platforms arrived to help science and healthcare to progress [6] and if they stick to this goal, public acceptance should be desirable. The interview study shed light on critical factors involved in young adults decision-making regarding (i) eventually contributing to health research platforms, (ii) and sharing specific types of information. The findings indicate that motivations to contribute (public and personal benefits) and indications that a platform can be trusted (who are the organisers, how data

is secured and used) are significant components for initial acceptance of the platform's terms. Acceptance is, however, not fully achieved before potential participants evaluate their willingness to share the specific data types in question. According to our findings, this decision depends on the relevance for research, the effort to data collection and risks for material and reputation damage. Even though previous works have identified similar factors, our findings are original because of the context under investigation. The literature on participating in research could perhaps enlighten some motivations to join, but it would be insufficient to describe the importance of the willingness to share each data type. Similarly, the literature on data sharing and privacy has not yet considered the case of contributing to a shared-access research platform. This empirical study allowed us to pinpoint the key factors for the acceptance of the emergent platforms for digital phenotyping research, which has not been done before.

In Figure 1, we show the conceptual model we introduce for acceptance in the context of health research data platforms, which includes a summary of our findings and our proposed design guidelines. These guidelines focus not only on the initial acceptance of the platform and its data types but also their *continuous acceptance*. As a recent review on technology acceptance highlights, health initiatives should be aiming for a long-term engagement with its users [101], which is also what we advocate.

5.1 Provide supervised health feedback as a motivation source

As mentioned in the review, the motivation to participate in research can come from many sources. Financial compensation, to start with, has been used in several clinical studies and existing platforms (e.g. All of Us). However, in the interview study, participants rarely mentioned this as a motivation source, which was also the case in previous reviews about bio-banks [96]. Even though it has been found that people might be willing to sell personal data for money [13], another work has questioned whether material incentives lead to increased adherence and data quality [76]. Such types of compensations were found to be more beneficial if combined with other motivations [56].

The reasons to contribute to a research repository mentioned by our participants included both altruistic and more individualistic ones. It is widely known that people are more open to sharing personal data if it is for helping patients [75], science [137] and society [78, 152]. However, it was observed in our study that getting direct benefits from the data was as important, which aligns with studies about sharing genetic information [70]. In the case research repositories which collect health and behavioural data, this direct benefit is related to the possibility of receiving feedback on personal health and well-being indicators. The All of Us Programme seems to offer this, as they state they will provide ways for volunteers to get access to the data they share and some of the results [108]. Allowing people to keep track of their results, could not only motivate people to participate but also stay adherent and see a meaning for the effort of data collection, mentioned in the interview.

However, platforms should be mindful of how they inform users about health management so that they can extract a correct meaning out of their data. There are many studies of digital tools for personal

health monitoring, [45, 68, 81, 87] and social sharing [30, 51, 66], which are useful inspirations on how to integrate personal care with self-tracking. However, an issue with self-administered tools is that there can be negative consequences, such as self-scrutiny [28, 100] and performance anxiety [81], which are particularly applicable to young adults [65]. Furthermore, when it comes to vulnerable patients of any age, self-monitoring is not recommended without the supervision of a specialist [11].

Therefore, research platforms should also consider how to offer proper clinical follow-up. Except for the COVID-19 testing, being offered by Project Baseline, large-scale platforms have not committed to health support. There are several studies about mobile applications which have combined personal health tracking with the close support of healthcare professionals [10, 93, 147]. Using health tracking with supervision and support could be a way to get participants motivated to contribute to research; however, it is also crucial to be mindful of the risks as mentioned above.

5.2 Allow personalised and flexible privacy choices

Our study elicited privacy concerns that could emerge if behavioural data is to be stored in a shared-access repository. Some were connected to the potential for personal harm, which was also found within data sharing for health treatments [106]; others were related to misalignment of personal values, such as data being used for hidden purposes, and a possible invasion to the privacy of others, which have been brought up in studies about genetic repositories [96]. Regarding differences across the different information types, even though quantitative studies could define some data types as more 'sensitive' [14], we observed that the reasons for such sensitivity depended a lot on the meaning people attribute to the different aspects of their lives and the importance they give to the different risks. Also, contrary to another study with young adults mobile sensors and privacy [121], our analysis showed that the same information often led to different types of concerns depending on the person, and no generalisations could be devised.

Such observation has implications for the use of behavioural monitoring for research repositories because if they intend to accommodate this pluralism of experiences, they need to allow participants to choose which data types are to be stored. Currently, consent forms follow an 'all or nothing' approach: participants who join a study are expected to share all the data required [116]; however, it can be difficult for some to disclose some parts of their life [113]. Participants in our interviews would like to stop tracking in some situations and select the level of detail visible to others, suggestions which were mentioned by older people (above 70) as well [106]. A way to achieve could be to provide participants with an overview of their data and allow them to exert some control over it, aligning with the increasing public interest on more flexible privacy options for personal data [1, 64, 74, 151]. User interfaces could be designed to allow interactive forms of selective disclosure [100] through continuous control of the data collected, as well as the possibility to not disclose parts of it. The consent process should also take into account these individual preferences, and volunteers should be given more flexibility about which information they want to share for research.

Figure 1: Conceptual model for the acceptance of health research data platforms based on our empirical findings.

Furthermore, our participants demonstrated a desire to have more control over the future of their data. Allowing participants to withdraw from studies is already required in recent regulations [144]; however, each platform has its approach when it comes to how much information and control participants have about the individual studies making use of their data. Participant-driven platforms (e.g. Open Humans) allow users to select precisely to which projects they want to share their data, but this can restrict data usage in the future. On the other hand, more significant initiatives (e.g. All of Us, Evidation) ask for broad consent so that they can use the data for any project they decide, but participants are not asked about anything else afterwards [70]. An attractive mid-term solution is, therefore, to allow participants to set-up rules or conditions under which any studies that fulfil them can access their data (tiered consent [20]). If participants would like to deny access to some receivers, this does not have to imply a complete withdrawal from the repository.

However, setting up privacy rules should not become an overwhelming process, in the way that consent forms became complex and lengthy documents [19, 84, 139]. The idea of moving away from the model of broad consent has been discussed in the context of bio-banks, but it is still an open question if permission should be requested for each new study [96]. Our findings indicate that it should be optional if permission is to be requested each time a new study needs the data - some might consider the obligation to provide consent as a burden while others might prefer to have full control. Providing individuals with more control over their data can be perceived as an indication of respect [98, 143], which can, in turn, become an enabler for trust and acceptance. The negotiation of privacy boundaries can be helped by more communication from the platform side, which is an aspect discussed next.

5.3 Use communication to increase engagement and acceptance

Our participants showed a positive attitude about sharing behavioural data for a research repository, as long as their data is safe, which aligns with the public views about genetic information as well [70, 96]. As discussed in the review, acceptance in this context is connected to the perceived usefulness of the purpose [14, 75]. Therefore, information about this purpose must be efficiently communicated to potential participants [152].

However, even though the repositories are providing a clear overall idea about how contributions can be useful, the ideal platform envisioned by most of our participants would be much more transparent about individual projects. As mentioned in the review of existing platforms, there is a lack of involvement of participants once data is collected. Participation in a repository is not only about the organisation who is behind it. If users want to know more about the on-going projects, they should be provided with complete reports about goals, methods and people involved, so that they know to what they are contributing. However, the communication between contributors and researchers in many cases consists only of 'publication updates' on a website.

For this reason, a continuous dialogue could be a path to fulfil the need for more transparency. Similar to our participants, young individuals in another study were intrigued about how some types of mobile sensors could be health indicators and if their identities could be revealed through them [121]. Platforms could offer more support and information when it comes to explaining how data is being used for research, going beyond public reports, and providing personalised content based on the participants' concerns. Besides, some of our participants were also interested in staying updated

with the research outcomes [39]. Platforms should consider how to keep volunteers updated more frequently about the on-going analysis. Such a continuous form of communication could include chat tools with the scientists, social forums with other participants [129] and open access virtual meetings, similar to what Open Humans has been offering, for example. Another aspect which deserves more attention is the fact that not everyone has the same understanding about the overall risks of sharing data with a repository [31]. Information about how data is secured and kept anonymous to the researchers should be appropriately conveyed [115] and tools for communication could be useful in this process.

Spaces for the active involvement of participants could correspond to the desire expressed by some the young adults in our interviews. Active engagement may involve suggesting new research questions, providing feedback and voting on decisions [21, 46, 130], which in turn can make people feel rewarded when the initiatives are successful [7]. Learning materials could be provided for participants interested in understanding the methods behind the investigations [150]. By considering and providing such options, the platforms of the future could move forward from the traditional relationship between ‘study subjects’ and ‘scientists’ [138].

From the review of the existing platforms, it could be observed that since the surge of the virus, there have been efforts to promote more participant-centred features, such open community calls (Open Humans) and weekly reports (Evidation). However, none of these examples has fully considered the directions presented in this section. The COVID-19 pandemic brought urgency to data sharing for research, which represents an excellent opportunity to reflect on how we should be building acceptance and trust. Our vision is that by bringing participants closer to the research process, their participation can become more engaging and respectful. Instead of imposing terms and conditions from the top, these could be negotiated and established based on participants’ input of how they would like the future of research repositories to be.

5.4 Limitations and future work

This paper presents a limited empirical study in terms of sample demographics, which could compromise broader generalisations. Besides, as the study happened before the COVID-19 pandemic, any possible changes in attitudes because of that are not represented. Future follow-up studies could investigate the perspective of other age groups, as well as the effect of the pandemic on their opinions. Quantitative approaches could also be used in the future to identify patterns across larger samples. It could also be interesting to expand this comprehension and consider how the emerging platforms are being conceived and perceived across the globe. Finally, further studies could investigate the impact of our design guidelines in-the-wild.

6 CONCLUSION

This paper provided an overview of health research data repositories from a participant-centred perspective. It reviewed the landscape of current platforms and brought together scattered literature to conceptualise the experience of participating in them. The acceptance to participate was found to depend on the trust in the

platform, individual motivations, and the willingness to share different data types. Based on the review and the empirical results, we formulated guidelines on how to enhance identified enablers and dissipate barriers for participation in an ethical way.

We argue that participants should not be considered passive study subjects, but rather be invited to get actively involved in the studies and control the shared data. Some existing initiatives have started moving towards this direction, but there are still several opportunities to design features to reshape relationships of power between data contributors and platform organisers. After all, the future of health research data platforms should be guided by the participants’ perspectives, because they are the ones putting themselves in a vulnerable position for the common good.

ACKNOWLEDGMENTS

This project is part of TEAM (Technology Enabled Mental Health for Young People), which is funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 722561. We thank each of the participants for their time and openness; Kevin Doherty and Darius Rohani for feedback; and our TEAM colleagues for the constant motivating support.

REFERENCES

- [1] Mohamed Abdelhamid. 2018. Greater patient health information control to improve the sustainability of health information exchanges. *Journal of biomedical informatics* 83 (2018), 150–158.
- [2] Achievement. 2019. “Achievement is my accountability buddy for logging food, activity, and water intake.” Retrieved May 2, 2020 from <https://www.myachievement.com/>
- [3] Afua Adjekum, Alessandro Blasimme, and Effy Vayena. 2018. Elements of Trust in Digital Health Systems: Scoping Review. *Journal of medical Internet research* 20, 12 (2018), e11254.
- [4] Emily E Anderson, Susan B Newman, and Alicia K Matthews. 2017. Improving informed consent: Stakeholder views. *AJOB empirical bioethics* 8, 3 (2017), 178–188.
- [5] Nicholas Anderson, Caleb Bragg, Andrea Hartzler, and Kelly Edwards. 2012. Participant-centric initiatives: tools to facilitate engagement in research. *Applied & translational genomics* 1 (2012), 25–29.
- [6] Euan A Ashley. 2015. The precision medicine initiative: a new national effort. *Jama* 313, 21 (2015), 2119–2120.
- [7] Mara Balestrini, Yvonne Rogers, Carolyn Hassan, Javi Creus, Martha King, and Paul Marshall. 2017. A city in common: a framework to orchestrate large-scale citizen engagement around urban issues. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2282–2294.
- [8] Gaurav Bansal, David Gefen, et al. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems* 49, 2 (2010), 138–150.
- [9] Rita Banzi, Steve Canham, Wolfgang Kuchinke, Karmela Krleza-Jeric, Jacques Demotes-Mainard, and Christian Ohmann. 2019. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials* 20, 1 (2019), 169.
- [10] Jakob E. Bardram and Mads Frost. 2017. Double-Loop health technology: Enabling socio-technical design of personal health technology in clinical practice. In *Designing Healthcare That Works: A Sociotechnical Approach*.
- [11] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2627–2636.
- [12] Scott Beach, Richard Schulz, Julie Downs, Judith Matthews, Bruce Barron, and Katherine Seelman. 2009. Disability, age, and informational privacy attitudes in quality of life technology applications: Results from a national web survey. *ACM Transactions on Accessible Computing (TACCESS)* 2, 1 (2009), 5.
- [13] Volker Benndorf and Hans-Theo Normann. 2018. The willingness to sell personal data. *The Scandinavian Journal of Economics* 120, 4 (2018), 1260–1278.
- [14] Jaspreet Bhatia and Travis D Breaux. 2018. Empirical measurement of perceived privacy risk. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 6 (2018), 34.

- [15] Barbara E Bierer, Rebecca Li, Mark Barnes, and Ida Sim. 2016. A global, neutral platform for sharing trial data. *New England Journal of Medicine* 374, 25 (2016), 2411–2413.
- [16] UK Biobank. 2020. *Online showcase of UK Biobank*. Retrieved May 2, 2020 from <http://biobank.ndph.ox.ac.uk/showcase>
- [17] AE Blandford. 2013. Semi-structured qualitative studies. Interaction Design Foundation.
- [18] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).
- [19] Jamie C Brehaut, Kelly Carroll, Glyn Elwyn, Raphael Saginur, Jonathan Kimelman, Kaveh Shojania, Ania Syrowatka, Trang Nguyen, Erica Hoe, and Dean Fergusson. 2012. Informed consent documents do not encourage good-quality decision making. *Journal of clinical epidemiology* 65, 7 (2012), 708–724.
- [20] Stefanie Broes, Denis Lacombe, Michiel Verlinden, and Isabelle Huys. 2018. Toward a tiered model to share clinical trial data and samples in precision oncology. *Frontiers in medicine* 5 (2018), 6.
- [21] Elizabeth Bromley, Lisa Mikesell, Felicia Jones, and Dmitry Khodyakov. 2015. From subject to participant: Ethics and the evolving role of community in health research. *American Journal of Public Health* 105, 5 (2015), 900–908.
- [22] Alison Burrows, David Coyle, and Rachael Gooberman-Hill. 2018. Privacy, boundaries and smart homes for health: An ethnographic study. *Health & place* 50 (2018), 112–118.
- [23] CACHET. 2019. *CARP Mobile Sensing*. Retrieved May 2, 2020 from <https://github.com/cph-cachet/carp.sensing-flutter>
- [24] Kelly Caine, Selma Šabanovic, and Mary Carter. 2012. The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 343–350.
- [25] Vivli center for global research data. 2019. *Our members*. Retrieved May 2, 2020 from <https://vivli.org/members/ourmembers/>
- [26] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1143–1152.
- [27] Delphine Christin, Andreas Reinhardt, Salil S Kanhere, and Matthias Hollick. 2011. A survey on privacy in mobile participatory sensing applications. *Journal of systems and software* 84, 11 (2011), 1928–1946.
- [28] Trudy Hui Hui Chua and Leanne Chang. 2016. Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior* 55 (2016), 190–197.
- [29] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, and Sean A Munson. 2015. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of medical Internet research* 17, 8 (2015), e203.
- [30] Chia-Fang Chung, Nanna Gorm, Irina A Shklovski, and Sean Munson. 2017. Finding the right fit: understanding health tracking in workplace wellness programs. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, 4875–4886.
- [31] Karin Clark, Matt Duckham, Marilys Guillemin, Assunta Hunter, Jodie McVeron, Christine O'Keefe, Cathy Pitkin, Steven Prawer, Richard Sinnott, Deborah Warr, et al. 2019. Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice. *Ethics and Information Technology* 21, 1 (2019), 59–73.
- [32] Deborah J Cohen, Sara R Keller, Gillian R Hayes, David A Dorr, Joan S Ash, and Dean F Sittig. 2015. Developing a model for understanding patient collection of observations of daily living: a qualitative meta-synthesis of the Project HealthDesign program. *Personal and ubiquitous computing* 19, 1 (2015), 91–102.
- [33] Francis S Collins and Harold Varmus. 2015. A new initiative on precision medicine. *New England journal of medicine* 372, 9 (2015), 793–795.
- [34] Eric Corbett and Christopher A Le Dantec. 2018. Exploring trust in digital civics. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 9–20.
- [35] Edward L Deci and Richard M Ryan. 2010. Intrinsic motivation. *The corsini encyclopedia of psychology* (2010), 1–2.
- [36] Pranammya Dey, Joseph S Ross, Jessica D Ritchie, Nihar R Desai, Sanjeev P Bhavnani, and Harlan M Krumholz. 2017. Data sharing and cardiology: platforms and possibilities. *Journal of the American College of Cardiology* 70, 24 (2017), 3018–3025.
- [37] Daniel Di Matteo, Alexa Fine, Kathryn Fotinos, Jonathan Rose, and Martin Katzman. 2018. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. *JMIR mental health* 5, 3 (2018), e56.
- [38] Ezgi Dogan, Christian Sander, Xenija Wagner, Ulrich Hegerl, and Elisabeth Kohls. 2017. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. *Journal of medical Internet research* 19, 7 (2017), e262.
- [39] Jan Domaradzki and Jakub Pawlikowski. 2019. Public Attitudes toward Biobanking of Human Biological Material for Research Purposes: A Literature Review. *International journal of environmental research and public health* 16, 12 (2019), 2209.
- [40] Juan Pablo Domecq, Gabriela Prutsky, Tarig Elraiyah, Zhen Wang, Mohammed Nabhan, Nathan Shippee, Juan Pablo Brito, Kasey Boehmer, Rim Hasan, Belal Firwana, et al. 2014. Patient engagement in research: a systematic review. *BMC health services research* 14, 1 (2014), 89.
- [41] Laura B Dunn and Dilip V Jeste. 2001. Enhancing informed consent for research and treatment. *Neuropsychopharmacology* 24, 6 (2001), 595–607.
- [42] Daniel A Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A Munson. 2016. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1109–1113.
- [43] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes Towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). ACM, New York, NY, USA, Article 494, 14 pages.
- [44] Evidation. 2019. *Radically changing medicine*. Retrieved May 2, 2020 from <https://evidation.com/>
- [45] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM, 216–223.
- [46] Rachael Fleurence, Joe V Selby, Kara Odom-Walker, Gail Hunt, David Meltzer, Jean R Slutsky, and Clyde Yancy. 2013. How the Patient-Centered Outcomes Research Institute is engaging patients and others in shaping its research agenda. *Health Affairs* 32, 2 (2013), 393–400.
- [47] Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J Oedegaard, and Jim Tørresen. 2018. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing* (2018).
- [48] Sunela Garg, Nidhi Bhatnagar, and Navya Gangadharan. 2020. A Case for Participatory Disease Surveillance of the COVID-19 Pandemic in India. *JMIR Public Health and Surveillance* 6, 2 (2020), e18795.
- [49] Heather H Gatny and William G Axinn. 2012. Willingness to participate in research during pregnancy: race, experience, and motivation. *Field methods* 24, 2 (2012), 135–154.
- [50] Katrina AB Goddard, K Sabina Smith, Chuhe Chen, Carmit McMullen, and Cheryl Johnson. 2009. Biobank recruitment: motivations for nonparticipation. *Biopreservation and biobanking* 7, 2 (2009), 119–121.
- [51] Lisa Graham, Anthony Tang, and Carman Neustaedter. 2016. Help me help you: Shared reflection for personal data. In *Proceedings of the 19th International Conference on Supporting Group Work*. ACM, 99–109.
- [52] Franz Gravenhorst, Amir Muaremi, Jakob Bardram, Agnes Grünerbl, Oscar Mayora, Gabriel Wurzer, Mads Frost, Venet Osmani, Bert Arnrich, Paul Lukowicz, et al. 2015. Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing* 19, 2 (2015), 335–353.
- [53] Bastian Greshake Tzovaras, Misha Angrist, Kevin Arvai, Mairi Dulaney, Vero Estrada-Galiñanes, Beau Gunderson, Tim Head, Dana Lewis, Oded Nov, Orit Shaer, et al. 2019. Open Humans: A platform for participant-centered research and personal data exploration. *GigaScience* 8, 6 (2019), giz076.
- [54] Mark Hamer, Ding Ding, Josephine Chau, Mitch J Duncan, and Emmanuel Stamatakis. 2020. Association between TV viewing and heart disease mortality: observational study using negative control outcome. *J Epidemiol Community Health* (2020).
- [55] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.
- [56] Gabriella M Harari, Sandrine R Müller, Varun Mishra, Rui Wang, Andrew T Campbell, Peter J Rentfrow, and Samuel D Gosling. 2017. An evaluation of students' interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science* 8, 5 (2017), 479–492.
- [57] Jean Hardy, Tiffany C Veinot, Xiang Yan, Veronica J Berrocal, Philippa Clarke, Robert Goodspeed, Iris N Gomez-Lopez, Daniel Romero, and VG Vinod Vydiswaran. 2018. User acceptance of location-tracking technologies in health research: implications for study design and data quality. *Journal of biomedical informatics* 79 (2018), 7–19.
- [58] Luke Hutton, Blaine A Price, Ryan Kelly, Ciaran McCormick, Arosha K Bandara, Tally Hatzakis, Maureen Meadows, and Bashir Nuseibeh. 2018. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. *JMIR mHealth and uHealth* 6, 10 (2018), e185.
- [59] Marcello Ienca and Effy Vayena. 2020. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature Medicine* (2020), 1–2.
- [60] Anton Ivanov, Raj Sharman, and H Raghav Rao. 2015. Exploring factors impacting sharing health-tracking records. *Health Policy and Technology* 4, 3 (2015), 263–276.
- [61] Maia L Jacobs, James Clawson, and Elizabeth D Mynatt. 2015. Comparing health information sharing preferences of cancer patients, doctors, and navigators.

- In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 808–818.
- [62] Adam N Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B Paine Schofield. 2010. Privacy, trust, and self-disclosure online. *Human-Computer Interaction* 25, 1 (2010), 1–24.
 - [63] Iris A Junglas, Norman A Johnson, and Christiane Spitzmüller. 2008. Personality traits and concern for privacy: an empirical study in the context of location-based services. *European Journal of Information Systems* 17, 4 (2008), 387–402.
 - [64] Jane Kaye, Edgar A Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. 2015. Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics* 23, 2 (2015), 141.
 - [65] Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 629–641.
 - [66] Elisabeth Kersten-van Dijk and Wijnand A IJsselstein. 2016. Design Beyond the Numbers: Sharing, Comparing, Storytelling and the Need for a Quantified Us. *IXD&A* 29 (2016), 121–135.
 - [67] Katherine K Kim, Jill G Joseph, and Lucila Ohno-Machado. 2015. Comparison of consumers' views on electronic data sharing for healthcare and research. *Journal of the American Medical Informatics Association* 22, 4 (2015), 821–830.
 - [68] Yoojung Kim, Eunyong Heo, Hyunjeong Lee, Sookyoung Ji, Jueun Choi, Jeong-Whun Kim, Joongseek Lee, and Sooyoung Yoo. 2017. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5787–5799.
 - [69] Yoojung Kim, Sookyoung Ji, Hyunjeong Lee, Jeong-Whun Kim, Sooyoung Yoo, and Joongseek Lee. 2016. My doctor is keeping an eye on me!: Exploring the clinical applicability of a mobile food logger. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5620–5631.
 - [70] Jennifer King. 2019. "Becoming Part of Something Bigger" Direct to Consumer Genetic Testing, Privacy, and Personal Disclosure. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–33.
 - [71] William R King and Jun He. 2006. A meta-analysis of the technology acceptance model. *Information & management* 43, 6 (2006), 740–755.
 - [72] René F Kizilcec. 2016. How much information?: Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395.
 - [73] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. 2009. Exploring privacy concerns about personal sensing. In *International Conference on Pervasive Computing*. Springer, 176–183.
 - [74] 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.
 - [75] Stephanie A Kraft, Mildred K Cho, Katherine Gillespie, Meghan Halley, Nina Varsava, Kelly E Ormond, Harold S Luft, Benjamin S Wilfond, and Sandra So-Jin Lee. 2018. Beyond consent: Building trusting relationships with diverse populations in precision medicine research. *The American Journal of Bioethics* 18, 4 (2018), 3–20.
 - [76] Nicola Lacetera, Mario Macis, and Robert Slonim. 2013. Economic rewards to motivate blood donations. *Science* 340, 6135 (2013), 927–928.
 - [77] Mark R Leary and Robin M Kowalski. 1990. Impression management: A literature review and two-component model. *Psychological bulletin* 107, 1 (1990), 34.
 - [78] Amy A Lemke, Wendy A Wolf, Jennifer Hebert-Beirne, and Maureen E Smith. 2010. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics* 13, 6 (2010), 368–377.
 - [79] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proceedings of the ninth symposium on usable privacy and security*. ACM, 7.
 - [80] Chen Li, Xiangdong Xu, Guanghua Zhou, Kai He, Tianliang Qi, Wei Zhang, Feng Tian, Qinghua Zheng, and Jianping Hu. 2019. Implementation of National Health Informatization in China: Survey About the Status Quo. *JMIR medical informatics* 7, 1 (2019), e12238.
 - [81] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 557–566.
 - [82] Melissa LoPresti, Geoff Appelboom, Olivier Bruyère, Jean-Yves Reginster, Eric Klug, and E Sander Connolly Jr. 2014. Patient engagement in clinical research through mobile technology. *Clinical Practice* 11, 6 (2014), 549.
 - [83] Pamela J Ludford, Reid Priedhorsky, Ken Reily, and Loren Terveen. 2007. Capturing, sharing, and using local place information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1235–1244.
 - [84] Ewa Luger, Stuart Moran, and Tom Rodden. 2013. Consent for all: revealing the hidden complexity of terms and conditions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2687–2696.
 - [85] Maulik D Majumdar, Lina Avancini Colucci, and Adam B Landman. 2015. The quantified patient of the future: Opportunities and challenges. In *Healthcare*, Vol. 3. Elsevier, 153–156.
 - [86] Miguel Malheiros, Sacha Brostoff, Charlene Jennett, and M Angela Sasse. 2013. Would you sell your mother's data? personal data disclosure in a simulated credit card application. In *The Economics of Information Security and Privacy*. Springer, 237–261.
 - [87] Meethu Malu and Leah Findlater. 2017. Sharing automatically tracked activity data: implications for therapists and people with mobility impairments. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 136–145.
 - [88] CJ Mann. 2003. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency medicine journal* 20, 1 (2003), 54–60.
 - [89] Kate E Mason, Neil Pearce, and Steven Cummins. 2018. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *The Lancet Public Health* 3, 1 (2018), e24–e33.
 - [90] Amy Maxmen. 2017. Google spin-off deploys wearable electronics for huge health study. *Nature News* 547, 7661 (2017), 13.
 - [91] Michelle L Mazurek, JP Arsenault, Joanna Bresee, Nitin Gupta, Iulia Ion, Christina Johns, Daniel Lee, Yuan Liang, Jenny Olsen, Brandon Salmon, et al. 2010. Access control for home data sharing: Attitudes, needs and practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 645–654.
 - [92] Jacob McPadden, Thomas JS Durant, Dustin R Bunch, Andreas Coppi, Nathaniel Price, Kris Rodgeron, Charles J Torre Jr, William Byron, Allen L Hsiao, Harlan M Krumholz, et al. 2019. Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform. *Journal of medical Internet research* 21, 4 (2019), e13043.
 - [93] Helena M Mennis, Anita Komlodi, Katrina Schrader, Michael Phipps, Ann Gruber-Baldini, Karen Yarbrough, and Lisa Shulman. 2017. Crafting a view of self-tracking data in the clinical visit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5800–5812.
 - [94] Nick Merrill, John Chuang, and Coye Cheshire. 2019. Sensing is Believing: What People Think Biosensors Can Reveal About Thoughts and Feelings. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, 413–420.
 - [95] Anna Middleton. 2018. Society and personal genome data. *Human molecular genetics* 27, R1 (2018), R8–R13.
 - [96] Anna Middleton, Richard Milne, Adrian Thorogood, Erika Kleiderman, Emilia Niemiec, Barbara Prainsack, Lauren Farley, Paul Bevan, Claire Steed, James Smith, et al. 2019. Attitudes of publics who are unwilling to donate DNA data for research. *European journal of medical genetics* 62, 5 (2019), 316–323.
 - [97] Alistair Morrison, Donald McMillan, and Matthew Chalmers. 2014. Improving consent in large scale mobile HCI through personalised representations of data. In *Proceedings of the 8th Nordic conference on human-computer interaction: Fun, fast, foundational*. ACM, 471–480.
 - [98] Daphne A Muller and Pierre Lévy. 2019. A design approach towards affording the trend of privacy. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 435–446.
 - [99] Patrick Murmann and Simone Fischer-Hübner. 2017. Usable transparency enhancing tools: A literature review.
 - [100] Elizabeth L Murnane, Tara G Walker, Beck Tench, Stephen Volda, and Jaime Snyder. 2018. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 127.
 - [101] Camille Nadal, Corina Sas, and Gavin Doherty. 2020. Technology acceptance in mHealth: a scoping review of definitions, models and measurement. *Journal of Medical Internet Research* (2020).
 - [102] Camille Nebeker, John Harlow, Rebeca Espinoza Giacinto, Rubi Orozco-Linares, Cinnamon S Bloss, and Nadir Weibel. 2017. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. *AJOB empirical bioethics* 8, 4 (2017), 266–276.
 - [103] Onicio Leal Neto, Oswaldo Cruz, Jones Albuquerque, Mariana Nacarato de Sousa, Mark Smolinski, Eduarda Ângela Pessoa Cesse, Marlo Libel, and Wayner Vieira de Souza. 2020. Participatory Surveillance Based on Crowdsourcing During the Rio 2016 Olympic Games Using the Guardians of Health Platform: Descriptive Study. *JMIR Public Health and Surveillance* 6, 2 (2020), e16119.
 - [104] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.
 - [105] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs* 41, 1 (2007), 100–126.

- [106] Leysan Nurgalieva, Alisa Frik, Francesco Ceschel, Serge Egelman, and Maurizio Marchese. 2019. Information Design in An Aged Care Context: Views of Older Adults on Information Sharing in a Care Triad. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 101–110.
- [107] National Institutes of Health (NIH). 2020. *All of Us Research Program Protocol*. Retrieved May 2, 2020 from <https://allofus.nih.gov/about/all-us-research-program-protocol>
- [108] National Institutes of Health (NIH). 2020. *Precision Medicine Initiative: Privacy and Trust Principles*. Retrieved May 2, 2020 from <https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-privacy-and-trust-principles>
- [109] Aisling Ann O'Kane and Helena Mentis. 2012. Sharing medical data vs. health knowledge in chronic illness care. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2417–2422.
- [110] Kirsten Osther, Svetlana Borodina, Rachel Conrad Bracken, Charles Lotterman, Eliot Storer, and Brandon Williams. 2017. Trust and privacy in the context of user-generated health data. *Big Data & Society* 4, 1 (2017), 2053951717704673.
- [111] Leysia Palen and Paul Dourish. 2003. Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 129–136.
- [112] Michael J Pencina, Darcy M Louzao, Brian J McCourt, Monique R Adams, Rehbar H Tayyabkhan, Peter Ronco, and Eric D Peterson. 2016. Supporting open access to clinical trial data for researchers: The Duke Clinical Research Institute-Bristol-Myers Squibb Supporting Open Access to Researchers Initiative. *American heart journal* 172 (2016), 64–69.
- [113] Justin Petelka, Lucy Van Kleunen, Liam Albright, Elizabeth Murnane, Stephen Volda, and Jaime Snyder. 2020. Being (In) Visible: Privacy, Transparency, and Disclosure in the Self-Management of Bipolar Disorder. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [114] Carolyn Petersen. 2016. The future of patient engagement in the governance of shared data. *eGEMS* 4, 2 (2016).
- [115] Carolyn Petersen. 2018. Through Patients' Eyes: Regulation, Technology, Privacy, and the Future. *Yearbook of medical informatics* 27, 01 (2018), 010–015.
- [116] James Pierce, Sarah Fox, Nick Merrill, Richmond Wong, and Carl DiSalvo. 2018. An Interface Without A User: An Exploratory Design Study of Online Privacy Policies and Digital Legalese. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 1345–1358.
- [117] Covid-19 Data Portal. 2020. *Accelerating research through data sharing*. Retrieved May 2, 2020 from <https://www.covid19dataportal.org/>
- [118] Abhishek Pratap, Ryan Allred, Jaden Duffy, Donovan Rivera, Heather Sophia Lee, Brenna N Renn, and Patricia A Areán. 2019. Contemporary Views of Research Participant Willingness to Participate and Share Digital Data in Biomedical Research. *JAMA network open* 2, 11 (2019).
- [119] Mika Raento and Antti Oulasvirta. 2008. Designing for privacy and self-presentation in social awareness. *Personal and Ubiquitous Computing* 12, 7 (2008), 527–542.
- [120] Fiona Riordan, Chrysanthi Papoutsis, Julie E Reed, Cicely Marston, Derek Bell, and Azeem Majeed. 2015. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *International journal of medical informatics* 84, 4 (2015), 237–247.
- [121] John Rooksby, Alistair Morrison, and Dave Murray-Rust. 2019. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 425.
- [122] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 217–226.
- [123] Anna Rudnicka, Anna L Cox, and Sandy JJ Gould. 2019. Why Do You Need This?: Selective Disclosure of Data Among Citizen Scientists. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 392.
- [124] Sundeep Sahay and Arunima Mukherjee. 2020. Where Is All Our Health Data Going? *Economic & Political Weekly* 55, 1 (2020), 47.
- [125] Pamela L Sankar and Lisa S Parker. 2017. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genetics in Medicine* 19, 7 (2017), 743.
- [126] Shruti Sannon and Dan Cosley. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 282.
- [127] Kate EA Saunders, Amy C Bilderbeck, Priyanka Panchal, Lauren Z Atkinson, JR Geddes, and Guy M Goodwin. 2017. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. *European Psychiatry* 41 (2017), 115–121.
- [128] Morten Schmidt, Lars Pedersen, and Henrik Toft Sørensen. 2014. The Danish Civil Registration System as a tool in epidemiology. *European journal of epidemiology* 29, 8 (2014), 541–549.
- [129] Siobhan Sharkey, Ray Jones, Janet Smithson, Elaine Hewis, Tobit Emmens, Tamsin Ford, and Christabel Owens. 2011. Ethical practice in internet research involving vulnerable people: lessons from a self-harm discussion forum study (SharpTalk). *Journal of medical ethics* 37, 12 (2011), 752–758.
- [130] Nathan D Shippee, Juan Pablo Domecq Garces, Gabriela J Prutsky Lopez, Zhen Wang, Tarig A Elraiyah, Mohammed Nabhan, Juan P Brito, Kasey Boehmer, Rim Hasan, Belal Firwana, et al. 2015. Patient and service user engagement in research: a systematic review and synthesized framework. *Health Expectations* 18, 5 (2015), 1151–1166.
- [131] Fatemeh Shirazi and Melanie Volkamer. 2014. What deters Jane from preventing identification and tracking on the Web?. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, 107–116.
- [132] Michael Siegrist. 2000. The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk analysis* 20, 2 (2000), 195–204.
- [133] Christian M Simon, Jamie L'heureux, Jeffrey C Murray, Patricia Winokur, George Weiner, Elizabeth Newbury, Laura Shinkunas, and Bridget Zimmerman. 2011. Active choice but not too active: public perspectives on biobank consent models. *Genetics in Medicine* 13, 9 (2011), 821.
- [134] Daniel J Solove. 2005. A taxonomy of privacy. *U. Pa. L. Rev.* 154 (2005), 477.
- [135] Jae W Song and Kevin C Chung. 2010. Observational studies: cohort and case-control studies. *Plastic and reconstructive surgery* 126, 6 (2010), 2234.
- [136] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 12, 3 (2015), e1001779.
- [137] Joshua Tan, Mahmood Sharif, Sruti Bhagavatula, Matthias Beckerle, Michelle L Mazurek, and Lujo Bauer. 2018. Comparing Hypothetical and Realistic Privacy Valuations. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. ACM, 168–182.
- [138] Sharon F Terry and Patrick F Terry. 2011. Power to the people: participant ownership of clinical trial data. *Science Translational Medicine* 3, 69 (2011), 69cm3–69cm3.
- [139] Susan Brown Trinidad, Stephanie M Fullerton, Julie M Bares, Gail P Jarvik, Eric B Larson, and Wylie Burke. 2012. Informed consent in genome-scale research: what do prospective participants think? *AJOB primary research* 3, 3 (2012), 3–11.
- [140] Yonatan Vaizman, Katherine Ellis, Gert Lankriet, and Nadir Weibel. 2018. Extratransparency app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 554.
- [141] Petar Velicković, Nicholas D Lane, Sourav Bhattacharya, Angela Chieh, Otmane Bellahsen, and Matthieu Vegreville. 2017. Scaling health analytics to millions without compromising privacy using deep distributed behavior models. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 92–100.
- [142] Giovanna N. Vilaza and Jakob E. Bardram. 2019. Sharing Access to Behavioural and Personal Health Data: Designers' Perspectives on Opportunities and Barriers. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (Trento, Italy) (PervasiveHealth'19)*. ACM, New York, NY, USA, 346–350.
- [143] Francesco Vitale, William Odom, and Joanna McGrenere. 2019. Keeping and Discarding Personal Data: Exploring a Design Space. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1463–1477.
- [144] Paul Voigt and Axel Von dem Bussche. 2017. The EU General Data Protection Regulation (GDPR). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).
- [145] Maiko Watanabe, Yusuke Inoue, Chiungfang Chang, Hyunsoo Hong, Izumi Kobayashi, Satoshi Suzuki, and Kaori Muto. 2011. For what am I participating? The need for communication after receiving consent from biobanking project participants: experience in Japan. *Journal of human genetics* 56, 5 (2011), 358.
- [146] Elissa R Weitzman, Skyler Kelemen, Liljana Kaci, and Kenneth D Mandl. 2012. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. *BMC medical informatics and decision making* 12, 1 (2012), 39.
- [147] Peter West, Richard Giordano, Max Van Kleek, and Nigel Shadbolt. 2016. The quantified patient in the doctor's office: Challenges & opportunities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3066–3078.
- [148] Richard Whiddett, Inga Hunter, Judith Engelbrecht, and Jocelyn Handy. 2006. Patients' attitudes towards sharing their health information. *International journal of medical informatics* 75, 7 (2006), 530–541.
- [149] Paul Wicks, Michael Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy Vaughan, Richard Bradley, and James Heywood. 2010. Sharing health data for better outcomes on PatientsLikeMe. *Journal of medical Internet research* 12, 2 (2010), e19.
- [150] Andrea Wiggins and John Wilbanks. 2019. The rise of citizen science in health and biomedical research. *The American Journal of Bioethics* 19, 8 (2019), 3–14.

- [151] Niels Wouters, Ryan Kelly, Eduardo Velloso, Katrin Wolf, Hasan Shahid Ferdous, Joshua Newn, Zaher Joukhadar, and Frank Vetere. 2019. Biometric Mirror: Exploring Values and Attitudes towards Facial Analysis and Automated Decision-Making. In *Conference on Designing Interactive Systems.*, Vol. 1145.
- [152] Martina Ziefle, Julian Halbey, and Sylvia Kowalewski. 2016. Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats. In *IoTBD*. 255–265.