CAMBRIDGE
UNIVERSITY PRESS

COMMENTARY

Emerging health data platforms: From individual control to collective data governance

Timothy Kariotis[1]* ⓘ, Mad Price Ball[2], Bastian Greshake Tzovaras[3], Simon Dennis[4], Tony Sahama[5], Carolyn Johnston[6], Helen Almond and Ann Borda[7]

[1]Melbourne School of Government, School of Computing and Information Systems, University of Melbourne.
[2]Open Humans Foundation.
[3]Center for Research and Interdisciplinarity (CRI) Université de Paris, Open Humans Foundation.
[4]Melbourne School of Psychological Sciences, University of Melbourne.
[5]School Health Information Science, Victoria University, BC, Canada. Faculty of Science, Engineering and IT, Federation University (Brisbane Campus), Australia.
[6]Melbourne Law School, University of Melbourne.
[7]Centre for Digital Transformation of Health, University of Melbourne.
*Corresponding author. E-mail: timothy.kariotis@unimelb.edu.au

Abstract

Health data have enormous potential to transform healthcare, health service design, research, and individual health management. However, health data collected by institutions tend to remain siloed within those institutions limiting access by other services, individuals or researchers. Further, health data generated outside health services (e.g., from wearable devices) may not be easily accessible or useable by individuals or connected to other parts of the health system. There are ongoing tensions between data protection and the use of data for the public good (e.g., research). Concurrently, there are a number of data platforms that provide ways to disrupt these traditional health data siloes, giving greater control to individuals and communities. Through four case studies, this paper explores platforms providing new ways for health data to be used for personal data sharing, self-health management, research, and clinical care. The case-studies include data platforms: PatientsLikeMe, Open Humans, Health Record Banks, and unforgettable.me. These are explored with regard to what they mean for data access, data control, and data governance. The case studies provide insight into a shift from institutional to individual data stewardship. Looking at emerging data governance models, such as data trusts and data commons, points to collective control over health data as an emerging approach to issues of data control. These shifts pose challenges as to how "traditional" health services make use of data collected on these platforms. Further, it raises broader policy questions regarding how to decide what public good data should be put towards.

Policy Significance Statement

The tension between data protection and using data for the public good, such as in health and medical research, is an ongoing policy issue. There are a growing number of data platforms, which aim to either ease data sharing between institutions and researchers or give greater control to individuals over their data. These platforms reflect a growing shift in data governance away from institutional control toward greater individual control. Looking

CrossMark

into the future, there is a growing interest in collective data governance of data for the public good. These emerging data platforms and the parallel shift in data governance pose broader challenges to both the health system and health and biomedical research with regard to data control.

1. Introduction

Health data, especially big health data, have been publicized as having the potential to improve health and medical research, health systems, and healthcare (Hripcsak et al., 2014; Kruse et al., 2016; Van Dijck and Poell, 2016; Sonja et al., 2018). However, because health data are often considered to be a particularly sensitive type of personal data, its use beyond individual care has been limited (Hafen et al., 2014; Hripcsak et al., 2014; Presser et al., 2015; Van Roessel et al., 2017). In addition to data collected in healthcare institutions (e.g., hospitals), there are also collections of biomedical research data (e.g., genomic data) that sit in data silos (Brand et al., 2016; Carbon et al., 2018). Finally, there are new sources of data related to individual health, such as person-generated health data from wearables and health devices or inferred from other data (e.g., social media posts). Equally, these new data sources are not readily accessible to individuals, health systems, or researchers (Van Kleek and OHara, 2014; Hodson, 2016; Verborgh, 2019). In this context, there is a growing number of platforms, which propose to "break down" the traditional data silos promoting both greater access to data for researchers, while also giving control of data back to individuals (Mählmann et al., 2017; Riso et al., 2017; Starkbaum and Felt, 2019). Alongside these platforms are new data governance strategies posing opportunities and challenges to current established practice.

This paper uses four case studies to explore the shifts in the control of health data ("Emerging Data Platforms" and "Data Control Dimensions" sections). Emerging models of data governance will also be considered, along with the potential impacts of these shifts on health systems ("Discussion" and "Conclusions").

1.1. Health data

Traditionally, health data included data collected within specific health services, such as tests, medical record, or biomedical research data. In recent years, what is considered as health data have expanded to include data as genomic data and patient-generated health data (Wood et al., 2015; Maglaveras et al., 2016; O'Doherty et al., 2016; Vayena et al., 2016). Patient-generated health data have been defined by Shapiro et al. (2012, p. 2) as "health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information-created, recorded, gathered, or inferred by or from patients or their designees (i.e., care partners or those who assist them) to help address a health concern." Purtova (2017) further makes the argument, in the age of data analytics, all data should be considered health data. This statement resonates due to the potential for health information to be gathered from any data. There is also a closing of the gap between research data and clinical care data. Traditionally, these two data sources have been treated in distinctly different ways. New data platforms are making this distinction porous, challenging how current regulations are applied to these platforms (Deverka et al., 2017).

1.2. Big data, data protection, and protecting health

Healthcare and health research have gradually moved toward greater data dependence (Weber et al., 2014; Starkbaum and Felt, 2019). The shift toward data-driven models of care and research is happening in tension with a greater focus on privacy and data protection, as captured by the European Union's General Data Protection Regulation (EU) 2016/679 (GDPR) or the United States Health Insurance Portability and Accountability Act (HIPAA). Although privacy and data protection are common policy themes, there are no agreed upon policy direction for the use of big data in healthcare (Blasimme et al., 2018). Mills (2019, p.14) describes this as a laissez-faire model of data governance, in which "individuals provide their data

on a pseudo-transactional basis in exchange for 'free' services." External regulation (e.g., HIPPA, GDPR) acts as a constraint on these transactions, setting standards regarding the transaction (e.g., notice and consent). However, beyond these constraints, there is little in the way of a data governance framework that involves individuals after the initial transaction. Driving toward a more data-driven health sector is the rise of big data and data analytic, which facilitates the collection, analysis, and linking of large sets of health data (Luo et al., 2016; Wang et al., 2018). There is great interest in linking together the different types of health data outlined above to provide new insights. For example, what social media activity can tell us about health behaviors (Young, 2014). Interestingly, much of these data are not under the control of those who generate it, as the quote below from the Obama's White House report into Big Data highlights (Executive Office of the President, 2014, p. 9).

> *"Certainly data is freely shared and duplicated more than ever before...The technological trajectory, however, is clear: more and more data will be generated about individuals and will persist under the control of others."*

2. Emerging Data Platforms

There are several existing platforms that promote giving control of health data back to individuals or dissolution of data silos. Rather than attempt to provide a comprehensive overview of these platforms and their features, this paper explores four case studies that provide insight into the challenges and opportunities these platforms raise. Each represent a way to approach the control of data generated in a traditional health institution and person-generated health data.

2.1. Case study 1: Open Humans and Nightscout Data Commons

Open Humans (www.openhumans.org) is a community-centric ecosystem. Launched in 2015, Open Humans now has over 8,000 members on the platform. Delivered as an online platform, it supports individuals importing and aggregating various types of personal data. These include personal health data such as wearable device data and genomic data—from various sources. Members registered with the platform are encouraged to decide how their data are used, for personal data exploration, as well as sharing with participant-led or academic research projects (Greshake Tzovaras et al., 2019). By aggregating personal health data from a variety of sources and putting the individual in control of the data, Open Humans aims to provide individuals access to their own data—breaking down silos—and increasing control over their data. Individuals can make granular decisions if and how to share those data.

The governance of the data held in Open Humans is handled at multiple levels. At the basic level, governance of data is performed by the individual, who has full control over who accesses any part of their data. Additionally, the larger community performs governance functions for the platform; individual projects invite individual members to share their data, these need to be approved before they are publicly listed to solicit data donations. The approval process is performed by community members reviewing prospective new projects. In addition, community members vote for the board members of the nonprofit foundation running the Open Humans platform.

When a member shares data, they give permission for diverse data re-use. This infers that Open Humans facilitate, the formation of "data commons" for specific data types—including by patient communities. Noteworthy examples of this are projects aggregating continuous glucose monitor (CGM) data, associated with the type 1 diabetes community: the *Nightscout Data Commons* (managed by the Nightscout Foundation) invites data donation from individuals using *Nightscout* open-source software for CGM data; users of Open Artificial Pancreas System (OpenAPS) for automated basal insulin regulation are invited to donate data to the *OpenAPS Data Commons* (managed by Dana Lewis, a lead OpenAPS developer). Lewis has managed the re-use of these data in further research and academic collaborators, for example, to assess glycemic control among OpenAPS users (Melmer et al., 2019).

2.2. Case study 2: Health Record Banking

Health Record Bank (HRB) is a patient-centric health information repository (Yasnoff, 2016). HRB has been described by the Health Record Banking Alliance (2012) as an "independent organizations that provide a secure electronic repository for storing and maintaining an individual's lifetime health and medical records from multiple sources and assure that the individual always has complete control over who accesses their information." The repository is governed by three key principles: (a) each patient's record should be functionally stored in one place (currently, not all patient records in the same place), (b) each patient should control who accesses to their medical records, and (c) medical records should be stored under the patient's control by a trusted organization. These principles facilitate and safeguard HRB storage, access, quality, control, security, and privacy. HRB proposes that an independent entity allows individuals to set up a HRB, which health institutions can then deposit health information into. The individual decides who can withdraw information from their account. While privacy has culturally dependent variables (Sahama et al., 2013), the HRB data model enhances patient empowerment, including sharing electronic health records with functional data interoperability and meaningful use to minimize the security and privacy concerns. The HRB data architecture does not support all patient records in a single repository. Thus, it is less vulnerable to loss of the entire dataset from a single unauthorized intrusion (Yasnoff, 2016). HRB is still in its infancy, but has garnered support from a number of stakeholders in the healthcare system. There have been pilots in the United States, which was discontinued due to low enrolment of people (Yasnoff and Shortliffe, 2014).

2.3. Case study 3: Unforgettable.me

Unforgettable Research Services Pty Ltd (www.unforgettable.me) provides a memory prosthetic and personal data brokerage service (Dennis et al., 2019). Unforgettable.me (2020) "has in excess of 2,675 registered participants… some of whom have been collecting for over four years. In November 2019 [the platform recorded] over 1.3 million events." It is currently in use as a research tool and not widely available to consumers. Participants can connect over 600 data streams to their account from: smart-phones, wearables, online services, and internet of things devices. These data streams derive from Unforgettable.me, If This Then That (IFTTT, ifttt.com), email servers and the SEMA3 ecological momentary assessment system (https://sema3.com/).

   Using keyword search and visualization interfaces, users are able to retrieve specific events, reflect on patterns across their daily activities and reminisce about their experiences. For compensation, they can also participate in projects posted by researchers. Users provide authorization for the use of their data on a project by project basis. This is considered a form of dynamic consent (Kaye et al., 2015). Projects posted to the system must have ethical clearance from approved research institutions, the clearance documentation is posted with the project. All data provided by users, either through passive data streams or through the completion of surveys or experiment, remain the property of the user. This is in contracts to typical university processes the data becomes the property of the research institution, allowing them to accumulate a personal data asset that they are free to license multiple times.

   To protect users' privacy and preserve the value of their data assets, researchers employ a privacy-preserving language called *Private*. This language allows analysis to be conducted without revealing raw data thus preventing researchers from seeing or copying it. All results provided by the language are tested to ensure that they are not too sensitive to the data provided by any given individual before they are released to the researcher (Dennis et al., 2019).

2.4. Case study 4: PatientsLikeMe

PatientsLikeMe (PLM; www.patientslikeme.com) is an online patient community with the vision of facilitating online interaction between individuals living with chronic health concerns (Frost and Massagli, 2008; PatientsLikeMe, 2019). PLM was founded in 2004 and remains in high use today. The platform has expanded to include thousands of patients and communities living with many different conditions. The main component of PLM, that participants interact with, is their profile. This is where

individuals can input and visualize their health data. PLM collects a variety of data, which is specific for each condition they have communities for. It includes data on disease progression, treatments, side-effects, medication, and physical activity; data is self-entered by the individual (Eichler et al., 2016). The data are transformed into visualizations, such as charts to track medications related to side-effects, or progression of disease. Data from all participants with the same condition are aggregated, allowing participants to relate their progression with other participants living with the same condition. Participants can also print out their data to share with their health providers (Wicks et al., 2010). Research is an important aspect of the PLM model, much of their funding comes from the selling of de-identified data back to industry (Morgan, 2019). PLM shares de-identified data with a number of groups, including the PLM community, PLM as a company, partners of PLM, and vendors (Patients Like Me, 2018). PLM has become a platform for crowdsourced research and patient-driven research with researchers engaging PLM participants in research trials, using PLM data in research, and using PLM data as the basis of research studies (Bradley et al., 2016).

3. Data Control Dimensions

To explore the different approaches to data control the four cases have taken, the authors draw on three dimensions of data control proposed by Vayena and Blasimme (2017). These are data use, data access, and governance. It is pertinent to mention at this point, the legal framework the four cases operate under only tell us the requirments for data processing, not how to implement the values of privacy and data protection in specific contexts. Thus, what the four cases have provided are examples of data governance models, which align with regulatory frameworks and expand the idea of data governance within a specific context.

3.1. Data use

Data use relates to both the primary and secondary uses of data, who decides these uses, and how they are legitimized and approved through data governance strategies. Traditionally implied or informed consent models are applied in healthcare in addition to data anonymization, so data can be used for secondary purposes (Vayena and Blasimme, 2017). Regulatory frameworks set the amount of control that an individual has over their data. In the GDPR, for example, consent is the primary mechanism for protection of personal data, and it must be freely given, specific, informed, and unambiguous. In the context of processing of health data, the data subject must have given explicit consent. However, both consent and anonymization approaches have come under scrutiny due to the ease by which data can be de-anonymized (Ohm, 2009), and more recently the plethora of uses of data beyond which individuals initially consent (Mittelstadt and Floridi, 2016). As already established, the current Laissez-Fair approach (Mills, 2017) to data governance does not give participants a voice in how their data are used, beyond the notice and consent procedure. This leaves the majority of control with the health system, healthcare providers, and clinicians. In relation to health data collected outside the health system, which usually resides either with the individual or with the companies that collect it, through notice and consent models. In the case studies we have outlined, some platforms increase individual's control over how data are used, and others decrease it. Unforgettable.me and HRB give greater control to individuals about how their data are used. For example, in unforgettable.me, people can choose exactly what projects can use their data, what types of data are used and retain ownership over their data. On the other hand, platforms such as PLM follow a traditional terms and conditions model, whereby signing-up to the platform individuals give away control over their data, as such it may be sold on to PLM partners. Finally, Open Humans shares control between the individual and the collective. However, individuals can decide how their data are used, the opportunities for data use are decided by the collective of people on the platform.

We have focused on individuals readily sharing their own data, but it is pertinent to consider what legal and ethical obligations parents and carers have in sharing their charge's health data, for example, the CGM data (mentioned "Case study 1: Open Humans and Nightscout Data Commons"). The term "sharenting" is used to describe parents sharing information about their children on social media, but any sharing of another's data comes with a moral obligation to act with appropriate discretion (Steinberg, 2016). If data

can be linked to an individual, this leaves a digital footprint, which can have foreseen (health insurance) and perhaps unforeseen implications.

3.2. Data access

Current challenges with both health data held in institutions and health data collected through other platforms, such as social media, is much of it is not easily accessible by the individual (Li et al., 2010). Thus, individuals have little ability outside institutional channels to gain benefit from that data. A potential benefit of the four case examples is that they aim to give individuals greater access to their data. HRB aims to ensure individuals can access and take their data with them. This represents what Vayena and Blassime (2017) desibe as a shift toward the individual as a distributor of their health data. In Open Humans, individuals and groups are also leading their own research but "traditional" health services and researchers will need to decide how to discern their findings (Vayena et al., 2018). An interesting point raised by Tempini and Del Savio (2019) is the risk of "digital orphans," individuals provide platforms with access to their data, however, this does not lead to broader accessibility for researchers. PLM data, for example, are accessible by the individual and PLM, but due to their business model, PLM became gatekeepers for others hoping to access that data (Tempini and Del Savio, 2019). A similar example is personal health apps, which though might provide some sharing options for data collected, usually putting constraints on the amount and granularity of data that can be accessed and shared (Kim et al., 2019).

Finally, does control mean active engagement? There is a risk that people do not, will not, or cannot actively engage with these platforms and it may be no better than leaving control in the hands of clinicians and health services (Steinsbekk et al., 2013). This phenomenon has been seen in the implementation of patient portals. Engagement with such portals is not guaranteed and may hinge on multiple elements, such as the design of the platform (Fylan et al., 2018; Dendere et al., 2019). Further, there is an implicit assumption, in some systems; people can take control of their health data. This assumption is made regardless of the fact that there are different levels of literacy, health literacy and digital literacy across any community. These factors can and will be a barrier to using such platforms (Bodie and Dutta, 2008).

3.3. Governance of system

Data governance is defined by Rosenbaum (2010, p. 1444) as "the process by which stewardship responsibilities are conceptualized and carried out, that is, the policies and approaches that enable stewardship." With the increasing potential of health data, data stewards and data governance approaches are under increasing scrutiny to define the appropriate uses of data in these new contexts (Hripcsak et al., 2014). Data governance is also one of the main challenges to the use of big data in healthcare (Kruse et al., 2016). There are a few broad ways these emerging platforms appear to address data governance. Platforms such as PLM require consent to the initial use of the platform, this then allows secondary use of de-identified data without further consent; thus, PLM becomes a data steward. HRBs propose a banking model, control sits with the individual and allows them to make decisions regarding who can withdraw health data from their individual "bank." This approach shifts the current control away from clinicians who traditionally tend to, with consent, make decisions on what data to share and when a person needs data about them shared (Kariotis et al., 2019). More broadly, what is seen in models, such as Open Humans is a move away from individuals just providing researchers access to their data, they are also able to define the projects that their data might contribute toward.

4. Discussion

4.1. What does the future look like?

In addition to the case studies, there are a number of emerging data governance approaches. These raise new considerations for health data; three examples which require further exploration are data trusts, data marketplaces, and data collectives.

4.1.1. Data trusts

Data trusts defined by the (Open Data Institute (2018, p. 1) are "a repeatable framework of terms and mechanisms that is mandated for use (or subject to scrutiny, or certification) in particular contexts to provide oversight of data access." Like any legal trust, it is envisioned that a trustee would be legally obliged to make decisions on data access and use in the best interest of stakeholders in a specific context (Open Data Institute, 2019). Data trusts have burst into popular policy discussion in recent years in part due to Alphabet (Google's parent company) proposing a data trust as a solution to concerns over data collected as part of a smart city project in Toronto (McDonald, 2019). There are examples, such as NightScout (outlined in "Case study 1: Open Humans and Nightscout Data Commons"), that act as a Data Trust as per the Open Data Institute (ODI) definition, even though it may not be labelled as such.

4.1.2. Data marketplaces

Data marketplaces put sole focus on the monetization of personal data. Examples include DataCoup (www.datacoup.com/) and CitizenMe (www.citizenme.com/). Both examples currently include the collection of health data, mainly collected from health devices, and personal wearable and readable devices. Data marketplaces, though they give control over data access to individuals, they do not, from the examples available, promote voice to individuals over how data should be used or serve any specific public good (Elvy, 2017). It is also questionable, who has the power to decide how valuable data are? and whether individuals will be able to negotiate this price. Individual data points may not seem valuable until they are linked together with other data points, many individuals may be unaware of this fact (Cooper and LaSalle, 2016).

4.1.3. Data cooperatives

Data cooperatives are membership organizations, individuals pay a fee to join and then can store their data in the cooperative, while also becoming part-owner and thus having a say in how the data are used (Hafen et al., 2014). One example of data cooperatives is MiDATA (https://www.midata.coop/), established in Switzerland. Individual account holders have logged access to their data, data cooperatives give complete decision-making control back to the individual whose data are stored, as well as how the revenues generated by sharing their data used (Hafe, 2019). Global and clinical research could be facilitated through secure national cooperatives (Hafen et al., 2014).

4.2. Shift from institutional to individual to collective and public good

It appears that there is a shift toward collective models of governance, as in Open Humans. Collective autonomy is an evolution of previous waves of data governance, which prioritizes individual autonomy (Evans, 2017). This reflects the evidence available on public attitudes for the use of health data beyond care; there is support for its use to further the common good (Skovgaard et al., 2019). Evans (2017, p. 1) suggests, we may be seeing "a popular uprising of regular people seeking a meaningful voice in establishing citizen-led ethical and privacy standards to advance big-data science while addressing the concerns people feel about the privacy of their health data." This shift toward collective control poses questions for privacy policies, which tend to focus on the individual right to privacy, rather than viewing privacy and control over data as a social good (Kasper, 2007).

The different approaches to data governance outlined in the case studies above are situated in a broader societal narrative around privacy of health data. There a number of recent examples of government supported electronic health records (e.g., My Health Record in Australia) or secondary data platforms (e.g., the proposed *care.data* in the Unite Kingdom), which have raised community concern of privacy and expectations of government compliance. Carter et al. (2015) consider that *care.data* experienced ongoing "challenge and contestation" resulting in its failed implementation because it did not meet societal expectations in relation to its proposed activities. Privacy of medical data is an important issue due
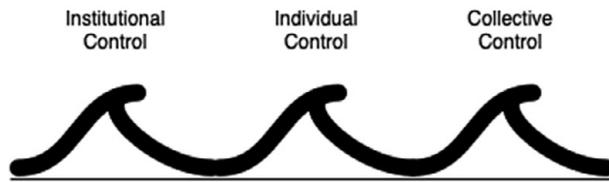
Figure 1 Waves of health data control.

to the sensitive nature of some information and the risks posed if the data "leaks" into other areas of life (e.g., insurance or the workplace). The risks of breaking confidentiality of medical data is especially great for some health conditions, which remain highly stigmatized (e.g., mental health conditions). However, at the same time, many service users and clinicians face barriers to receiving and providing good care because of the lack of information sharing.

The platforms considered have outlined three broad waves of data governance, that of control sitting with institutions, control sitting with individuals, and control being shared between individuals and the collective (Figure 1). The changes these data governance strategies pose go beyond the control of data and present challenges to the "traditional" health system.

4.3. How will this change healthcare?

The platforms described sit in contrast to the current workings of a traditional health system, where data within the system remains siloed, and data outside the health system lack legitimacy. These platforms can provide individuals with access to new data and new knowledge. This infers individuals are now entering a new era for health services. Clinicians and healthcare providers can be provided with more data than they may expect. The health system may not be prepared to use data collected from outside the health system, both in terms of individuals knowing how to bring their data into the health system and clinicians knowing what to do with that data (Chung et al., 2016; Zhu et al., 2017). There is also a risk that the shift toward data-driven healthcare facilitated by these platforms, may not be to the benefit of greater person-centred care which relies on the therapeutic relationship (Greenhalgh et al., 2015). Further, the OpenAPS glucose monitoring system (outlined in "Case study 1: Open Humans And Nightscout Data Commons") raises the issue of regulation in a health system where individuals and groups are sharing and learning new skills to manage their health. The rise of "do-it-yourself" systems, such as OpenAPS pose questions as to how will health services and health regulators manage this crowdsourced data and knowledge? Which, quite fairly, individuals may view as meeting their needs more than other options in the health system (Farrington, 2017).

This paper has explored the way the different platforms and how their associated data governance strategies may change healthcare in relation to information norms. Nissenbaum (2009) in her book "*Privacy in Context*" has pioneered an alternative way of considering why the implementation of new technologies into certain contexts leads to privacy concerns. Historically, privacy has been considered the individuals ability to limit the collection of personal information about themselves. Nissenbaum (2019) argues that privacy is the appropriate flow of information, what is appropriate is defined by contextual information norms. Nissenbaum (2019) outlines, in any context there are specific information norms relating to: the different actors (sender, receiver, subject); the type of information (e.g., health information or administrative information); and transmission principles (e.g., consent is required, by law, online/offline). New technologies, which breach these norms can lead to privacy backlashes and/or work-arounds by individuals attempting to realign to their expected norms. Although a technology, such as the proposed Care.data may align with legal requirements, it may not necessarily align with expected contextual norms (Carter et al., 2015). This identified a key weakness of the current Laissez-Faire model of data governance described by Mills (2017), it does not have any mechanisms built-in to grapple with contextual norms.

5. Conclusions

Health data are powerful resource that have the potential to improve healthcare, empower individuals, and advance health research. However, traditional control and access to health data have been limited, held in siloes by institutions. An emerging array of health data platforms are changing this siloed situation by opening up data to individual and collective uses. These new models pose new data governance opportunities and challenges. The four case studies explored represent a shift from institutional to individual control. However, there is also a current shift toward more collective approaches to data governance, such as data trusts and data collectives. These new emerging data platforms raise broader concerns regarding the management of privacy and integration into the traditional health system. A particular obstacle is how the knowledge accessed through the platforms, that is not part of the "traditional" health system, may create challenges for clinicians and health services in the translational use of this information.

References

Bastian Greshake Tzovaras, **Misha Angrist**, **Kevin Arvai**, **Mairi Dulaney**, **Vero Estrada-Galiñanes**, **Beau Gunderson**, **Tim Head**, **Dana Lewis**, **Oded Nov**, **Orit Shaer**, **Athina Tzovara**, **Jason Bobe**, **Mad Price Ball**. (2019) Open Humans: A platform for participant-centered research and personal data exploration. *GigaScience 8*(6), giz076. https://doi.org/10.1093/gigascience/giz076

Blasimme A, **Fadda M**, **Schneider M and Vayena E** (2018) Data sharing for precision medicine: Policy lessons and future directions. *Health Affairs 37*(5), 702–709. https://doi.org/10.1377/hlthaff.2017.1558

Bodie GD and Dutta MJ (2008) Understanding health literacy for strategic health marketing: EHealth literacy, health disparities, and the digital divide. *Health Marketing Quarterly 25*(1–2), 175–203. https://doi.org/10.1080/07359680802126301

Bradley M, **Braverman J**, **Harrington M and Wicks P** (2016) Patients' motivations and interest in research: Characteristics of volunteers for patient-led projects on PatientsLikeMe. *Research Involvement and Engagement 2*(1), 33. https://doi.org/10.1186/s40900-016-0047-6

Brand A, **Evangelatos N and Satyamoorthy K** (2016) Public Health Genomics: The essential part for good governance in public health. *International Journal of Public Health 61*(4), 401–403. https://doi.org/10.1007/s00038-016-0828-6

Carbon S, **Champieux R**, **McMurry J**, **Winfree L**, **Wyatt LR and Haendel M** (2018) A measure of open data: A metric and analysis of reusable data practices in biomedical data resources. *BioRxiv*, 282830. https://doi.org/10.1101/282830

Carter P, **Laurie GT and Dixon-Woods M** (2015) The social licence for research: Why care. Data ran into trouble. *Journal of Medical Ethics 41*(5), 404–409.

Chung C-F, **Dew K**, **Cole A**, **Zia J**, **Fogarty J**, **Kientz JA and Munson SA** (2016) Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.* pp. 770–786. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2818048.2819926

Cooper T and LaSalle R (2016) Guarding and Growing Personal Data Value. Available at https://www.accenture.com/t20160929T010202__w__/id-en/_acnmedia/PDF-32/Accenture-Guarding-and-Growing-Personal-Data-Value-POV-Low-Res.pdf October 10 2019.

Dendere R, **Slade C**, **Burton-Jones A**, **Sullivan C**, **Staib A and Janda M** (2019) Patient portals facilitating engagement with inpatient electronic medical records: A systematic review. *Journal of Medical Internet Research 21*(4), e12779. http://doi.org/10.2196/12779

Dennis S, **Yim H**, **Garrett P**, **Sreekumar V and Stone B** (2019) A system for collecting and analyzing experience-sampling data. *Behavior Research Methods 51*(4), 1824–1838. http://doi.org/10.3758/s13428-019-01260-y

Effy Vayena1, **Roger Brownsword**, **Sarah Jane Edwards**, **Bastian Greshake**, **Jeffrey P Kahn**, **Navjoyt Ladher**, **Jonathan Montgomery**, **Daniel O'Connor**, **Onora O'Neill**, **Martin P Richards**, **Annette Rid**, **Mark Sheehan**, **Paul Wicks**, **John Tasioulas** (2016) Research led by participants: A new social contract for a new kind of research. *Journal of Medical Ethics 42*(4), 216–219. https://doi.org/10.1136/medethics-2015-102663

Elvy SA (2017) Paying for privacy and the personal data economy. *Columbia Law Review 117*, 1369. Available at https://columbialawreview.org/content/paying-for-privacy-and-the-personal-data-economy/

Evans BJ (2017) Power to the people: Data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law 19*(2), 243. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5673282/

Executive Office of the President (2014) Big Data: Seizing Opportunities, Preserving Values. Available at https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf October 20 2019.

Farrington C (2017) Hacking diabetes: DIY artificial pancreas systems. *The Lancet Diabetes & Endocrinology 5*(5), 332. https://doi.org/10.1016/S2213-8587(16)30397-7

Frost J and Massagli M (2008) Social uses of personal health information within PatientsLikeMe, an online patient community: What can happen when patients have access to one another's data. *Journal of Medical Internet Research 10*(3), e15. https://doi.org/10.2196/jmir.1053

Fylan F, **Caveney L**, **Cartwright A and Fylan B** (2018) Making it work for me: Beliefs about making a personal health record relevant and useable. *BMC Health Services Research 18*(1), 445. https://doi.org/10.1186/s12913-018-3254-z

Gabriel S Eichler, **Elisenda Cochin**, **Jian Han**, **Sylvia Hu**, **Timothy E Vaughan**, **Paul Wicks**, **Charles Barr**, **Jenny Devenport**. (2016) Exploring concordance of patient-reported information on PatientsLikeMe and medical claims data at the patient level. *Journal of Medical Internet Research 18*(5), e110. https://doi.org/10.2196/jmir.5130

George Hripcsak, **Meryl Bloomrosen**, **Patti FlatelyBrennan**, **Christopher G Chute**, **Jim Cimino**, **Don E Detmer**, **Margo Edmunds**, **Peter J Embi**, **Melissa M Goldstein**, **William Ed Hammond**, **Gail M Keenan**, **Steve Labkoff**, **Shawn Murphy**, **Charlie Safran**, **Stuart Speedie**, **Howard Strasberg**, **Freda Temple, and Adam B Wilcox**. (2014) Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 Health Policy Meeting. *Journal of the American Medical Informatics Association 21*(2), 204–211. https://doi.org/10.1136/amiajnl-2013-002117

Greenhalgh T, **Snow R**, **Ryan S**, **Rees S and Salisbury H** (2015) Six 'biases' against patients and carers in evidence-based medicine. *BMC Medicine 13*(1), 200. https://doi.org/10.1186/s12916-015-0437-x

Hafen E (2019) Personal data cooperatives—A new data governance framework for data donations and precision health. In Krutzinna J and Floridi L (eds), *The Ethics of Medical Data Donation*. pp. 141–149. https://doi.org/10.1007/978-3-030-04363-6_9

Hafen E, **Kossmann D and Brand A** (2014) Health data cooperatives—Citizen empowerment. *Methods of Information in Medicine 53*(2), 82–86. https://doi.org/10.3414/ME13-02-0051

Health Record Banking Alliance (2012) Health Record Banking: A Foundation for Myriad Health Information Sharing Business Models. Available at http://www.healthbanking.org/uploads/9/6/9/4/9694117/hrba_white_paper_business_model_dec_2012.pdf September 25 2019.

Hodson H (2016) Put your data to work. *New Scientist 231*(3090), 22. https://doi.org/10.1016/S0262-4079(16)31650-5

Kariotis T, **Prictor M**, **Chang S and Gray K** (2019) Evaluating the contextual integrity of Australia's my health record. *Studies in health technology and informatics 265*, 213–218. https://doi.org/10.3233/SHTI190166

Kasper DV (2007) Privacy as a social good. *Social Thought & Research*, *28* 165–189. Available at https://www.jstor.org/stable/23252125

Kaye J, **Whitley EA**, **Lund D**, **Morrison M**, **Teare H and Melham K** (2015) Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics 23*(2), 141. https://doi.org/10.1038/ejhg.2014.71

Kieran C. O'Doherty, **Emily Christofides**, **Jeffery Yen**, **Heidi Beate Bentzen**, **Wylie Burke**, **Nina Hallowell**, **Barbara A. Koenig** & **Donald J. Willison** (2016) If you build it, they will come: Unintended future uses of organised health data collections. *BMC Medical Ethics 17*(1), 54. https://doi.org/10.1186/s12910-016-0137-x

Kim Y, **Lee B and Choe EK** (2019) Investigating data accessibility of personal health apps. *Journal of the American Medical Informatics Association 26*(5), 412–419. https://doi.org/10.1093/jamia/ocz003

Kruse CS, **Goswamy R**, **Raval YJ and Marawi S** (2016) Challenges and opportunities of big data in health care: A systematic review. *JMIR Medical Informatics 4*(4), e38. https://doi.org/10.2196/medinform.5359

Li M, **Yu S**, **Ren K and Lou W** (2010) Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings. In Jajodia S and Zhou J (eds), *Security and Privacy in Communication Networks*. Berlin, Heidelberg: Springer, pp. 89–106.

Luo J, **Wu M**, **Gopukumar D and Zhao Y** (2016) Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights 8*, 1–10. https://doi.org/10.4137/BII.S31559

Maglaveras N, **Kilintzis V**, **Koutkias V and Chouvarda I** (2016) Integrated care and connected health approaches leveraging personalised health through big data analytics. *Studies in Health Technology and Informatics 224*, 117–122. https://doi.org/10.3233/978-1-61499-653-8-117

Mählmann L, **Reumann M**, **Evangelatos N and Brand A** (2017) Big data for public health policy-making: Policy empowerment. *Public Health Genomics 20*(6), 312–320. https://doi.org/10.1159/000486587

McDonald S (2019) Reclaiming Data Trusts. Centre for International Governance Innovation website. Available at https://www.cigionline.org/articles/reclaiming-data-trusts (accessed 2 December 2019).

Melmer A, **Züger T**, **Lewis DM**, **Leibrand S**, **Stettler C and Laimer M** (2019) Glycemic control in individuals with type 1 diabetes using an open source artificial pancreas system (OpenAPS). *Diabetes, Obesity and Metabolism 21*(10), 2333–2337. https://doi.org/10.1111/dom.13810

Mills, Stuart, Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership (September 24, 2019). Available at SSRN: https://ssrn.com/abstract=3437936 or http://dx.doi.org/10.2139/ssrn.3437936

Mills S (2019) Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership. *Data Trusts, Data Commons, and the Future of Data Ownership.* Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3437936 (accessed 15 August 2019).

Mittelstadt BD and Floridi L (2016) The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics 22*(2), 303–341. https://doi.org/10.1007/s11948-015-9652-2

Morgan L (2019) How Does Patients Like Me Make Money? Patients Like Me website. Available at https://support.patientslikeme.com/hc/en-us/articles/201245750-How-does-PatientsLikeMe-make-money- (accessed 28 November 2019).

Nissenbaum H (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford California Stanford University Press.

Nissenbaum H (2019) Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law 20*(1) 221–256. Available at http://www7.tau.ac.il/ojs/index.php/til/article/view/1614

Ohm P (2009) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review 57*, 1701. Available at https://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/

Open Data Institute (2018) What is a Data Trust? Available at https://theodi.org/article/what-is-a-data-trust (accessed 1 December 2019).

Open Data Institute (2019) Data Trusts: Lessons from Three Pilots. Available at https://theodi.org/article/odi-data-trusts-report/ October 25 2019.

Patients Like Me (2018) Privacy Policy. Available at https://www.patientslikeme.com/about/privacy_full (accessed 28 November 2019).

Patients Like Me (2019) Featured Conditions at Patients Like Me. Available at https://www.patientslikeme.com/conditions (accessed 28 November 2019).

Patricia A. Deverka, **Mary A. Majumder**, **Angela G. Villanueva**, **Margaret Anderson**, **Annette C. Bakker**, **Jessica Bardill**, **Eric Boerwinkle**, **Tania Bubela**, **Barbara J. Evans**, **Nanibaa' A. Garrison**, **Richard A. Gibbs**, **Robert Gentleman**, **David Glazer**, **Melissa M. Goldstein**, **Hank Greely**, **Crane Harris**, **Bartha M. Knoppers**, **Barbara A. Koenig**, **Isaac S. Kohane**, **Salvatore La Rosa**, **John Mattison**, **Christopher J. O'Donnell**, **Arti K. Rai**, **Heidi L. Rehm**, **Laura L. Rodriguez**, **Robert Shelton**, **Tania Simoncelli**, **Sharon F. Terry**, **Michael S. Watson**, **John Wilbanks**, **Robert Cook-Deegan** & **Amy L. McGuire** (2017) Creating a data resource: What will it take to build a medical information commons? *Genome Medicine 9*(1), 84. https://doi.org/10.1186/s13073-017-0476-3

Paul Wicks, **Michael Massagli**, **Jeana Frost**, **Catherine Brownstein**, **Sally Okun**, **Timothy Vaughan**, **Richard Bradley**, **James Heywood** (2010) Sharing health data for better outcomes on PatientsLikeMe. *Journal of Medical Internet Research 12*(2), e19. https://doi.org/10.2196/jmir.1549.

Presser L, **Hruskova M**, **Rowbottom H and Kancir J** (2015) Care. Data and access to UK health records: Patient privacy and public trust. *Technology Science* 1–35. Available at https://techscience.org/a/2015081103

Purtova N (2017) Health data for common good: Defining the boundaries and social dilemmas of data commons. In *Under observation: The interplay between eHealth and surveillance.* Springer, pp. 177–210. https://doi.org/10.1007/978-3-319-48342-9

Riso B, **Tupasela A**, **Vears DF**, **Felzmann H**, **Cockbain J**, **Loi M**, **Rakic V**, *et al.* (2017) Ethical sharing of health data in online platforms—Which values should be considered? *Life Sciences, Society and Policy 13*(1), 12. https://doi.org/1186/s40504-017-0060-z

Rosenbaum S (2010) Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research 45*(5p2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x

Sahama T, **Simpson L and Lane B** (2013) Security and privacy in ehealth: Is it possible? In *2013 IEEE 15th International Conference on E-Health Networking, Applications and Services (Healthcom 2013).* pp. 249–253. https://doi.org/10.1109/HealthCom.2013.6720676

Shapiro M, **Johnston D**, **Wald J and Mon D** (2012) Patient-Generated Health Data. Office of Policy and PlanningOffice of the National Coordinator for Health Information Technology website. Available at https://www.rti.org/sites/default/files/resources/patientgeneratedhealthdata.pdf

Skovgaard LL, **Wadmann S and Hoeyer K** (2019) A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good. *Health Policy 123*(6), 564–571. https://doi.org/10.1016/j.healthpol.2019.03.012

Sonja M, **Ioana G**, **Miaoqing Y and Anna K** (2018) Understanding value in health data ecosystems: A review of current evidence and ways forward. *Rand Health Quarterly 7*(2), 3. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5798965/

Starkbaum J and Felt U (2019) Negotiating the reuse of health-data: Research, Big Data, and the European General Data Protection Regulation. *Big Data & Society 6*(2), 1–12. https://doi.org/10.1177/2053951719862594

Steinberg SB (2016). Sharenting: Children's privacy in the age of social media. *Emory Law Journal 66*, 839.

Steinsbekk KS, **Myskja BK and Solberg B** (2013) Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? *European Journal of Human Genetics 21*(9), 897. https://doi.org/10.1038/ejhg.2012.282

Tempini N and Del Savio L (2019) Digital orphans: Data closure and openness in patient-powered networks. *BioSocieties 14*(2), 205–227. https://doi.org/10.1057/s41292-018-0125-0

Unforgettable.me (2020) Research Services. Available at https://www.unforgettable.me/research-services/ (accessed 8 June 2020).

Van Dijck J and Poell T (2016) Understanding the promises and premises of online health platforms. *Big Data & Society 3*(1), 1–11. https://doi.org/10.1177/2053951716654173

Van Kleek M and OHara K (2014) The future of social is personal: The potential of the personal data store. In *Social Collective Intelligence.* Springer, pp. 125–158 10.1007/978-3-319-08681-1.

Van Roessel I, **Reumann M and Brand A** (2017) Potentials and challenges of the health data cooperative model. *Public Health Genomics 20*(6), 321–331. https://doi.org/10.1159/000489994

Vayena E and Blasimme A (2017) Biomedical big data: New models of control over access, use and governance. *Journal of Bioethical Inquiry 14*(4), 501–513. https://doi.org/10.1007/s11673-017-9809-6

Vayena E, **Dzenowagis J**, **Brownstein JS and Sheikh A** (2018) Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, *96*(1), 66. https://doi.org/10.2471/BLT.17.197426

Verborgh R (2019) How We Regain Control of Personal Data. Towards Data Science website. Available at https://towardsdatascience.com/ruben-verborgh-on-data-privacy-accf91d280c9 (accessed 25 November 2019).

Wang Y, **Kung L and Byrd TA** (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change 126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Weber GM, **Mandl KD and Kohane IS** (2014) Finding the missing link for big biomedical data. *JAMA 311*(24), 2479–2480. https://doi.org/10.1001/jama.2014.4228

Wood WA, **Bennett AV and Basch E** (2015) Emerging uses of patient generated health data in clinical research. *Molecular Oncology 9*(5), 1018–1024. https://doi.org/10.1016/j.molonc.2014.08.006.

Yasnoff WA (2016) A secure and efficiently searchable health information architecture. *Journal of Biomedical Informatics 61*, 237–246. https://doi.org/10.1016/j.jbi.2016.04.004

Yasnoff WA and Shortliffe E (2014) Lessons learned from a health record bank start-up. *Methods of Information in Medicine 53*(02), 66–72. https://doi.org/10.3414/ME13-02-0030

Young SD (2014) Behavioral insights on big data: using social media for predicting biomedical outcomes. *Trends in Microbiology 22*(11), 601–602. https://doi.org/10.1016/j.tim.2014.08.004

Zhu H, **Colgan J**, **Reddy M and Choe EK** (2017) Sharing patient-generated data in clinical practices: An interview study. In *AMIA—Annual Symposium proceedings. AMIA Symposium, 2016.* pp. 1303–1312. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333267/