Original Research Article

BIG DATA & SOCIETY

Big Data & Society
January–June 2017: 1–11

DOI: 10.1177/2053951717704673
journals.sagepub.com/home/bds
SAGE

Trust and privacy in the context of user-generated health data

Kirsten Ostherr, Svetlana Borodina, Rachel Conrad Bracken, Charles Lotterman, Eliot Storer and Brandon Williams

Abstract

This study identifies and explores evolving concepts of trust and privacy in the context of user-generated health data. We define "user-generated health data" as data captured through devices or software (whether purpose built or commercially available) and used outside of traditional clinical settings for tracking personal health data. The investigators conducted qualitative research through semistructured interviews (n = 32) with researchers, health technology start-up companies, and members of the general public to inquire why and how they interact with and understand the value of user-generated health data. We found significant results concerning new attitudes toward trust, privacy, and sharing of health data outside of clinical settings that conflict with regulations governing health data within clinical settings. Members of the general public expressed little concern about sharing health data with the companies that sold the devices or apps they used, and indicated that they rarely read the "terms and conditions" detailing how their data may be exploited by the company or third-party affiliates before consenting to them. In contrast, interviews with researchers revealed significant resistance among potential research participants to sharing their user-generated health data for purposes of scientific study. The widespread rhetoric of personalization and social sharing in "user-generated culture" appears to facilitate an understanding of user-generated health data that deemphasizes the risk of exploitation in favor of loosely defined benefits to individual and social well-being. We recommend clarification and greater transparency of regulations governing data sharing related to health.

Keywords

Big data, health data, terms and conditions, trust, privacy, sharing

Introduction

As the field of medicine has begun to embrace big data, a problematic truism has taken hold: more data equals more knowledge equals better health outcomes. Ubiquitous environmental and lifestyle data from wearable technologies and mobile apps promises to uncover new indicators of health and illness from outside of traditional clinical settings (Steinhubl et al., 2015). While the often-cited "Four V's" of big data, "volume, variety, velocity, and veracity," all hint at the complexity of deriving straightforward insights from these new sources of data (Raghupathi and Raghupathi, 2014), the governing logic of many business and research enterprises holds that the unfettered flow of data will yield real value as soon as it is coupled with appropriate analytics. But what new kinds of knowledge might these insights reveal and for whom might they improve outcomes? The novel achievements of user-generated health data rely heavily on participants' willingness to share their data, even when doing so may not serve their own best interests (Leaf, 2015). The question of who benefits from big health data is therefore entangled with questions about data ownership, sharing, trust, and privacy. This study explores how concepts of trust and privacy are changing in the context of user-generated health data and

Rice University, USA

Corresponding author:
Kirsten Ostherr, Department of English, Rice University, 6100 Main St. MS-30, Houston, TX 77005, USA.
Email: kostherr@rice.edu

analyzes how researchers, start-up companies, and members of the general public interact with and understand the value of user-generated health data as a key component of big health data.

Members of the general public, including patients, have begun to play a newly important role in collecting data about health and disease (Sarasohn-Kahn, 2014). With the rise of the mobile web and the growth of smartphone use (Rainie and Wellman, 2014), citizens' daily lives have become experiments "in the wild," whose digital traces offer new opportunities and challenges to researchers seeking to gather information about human behavior and exposures outside of the controlled settings of lab-based studies. This phenomenon has emerged with the rise of "user-generated content" (UGC), defined as content that "comes from regular people who voluntarily contribute data, information, or media that then appears before others in a useful or entertaining way, usually on the Web—for example, restaurant ratings, wikis, and videos" (Krumm et al., 2008; Van Dijck, 2009). As researchers and marketers began to mine UGC for insights and predictors of user behavior in the early 2000s, the relevance to health of what might be considered incidental data, such as global positioning system (GPS) or social media data, quickly became apparent. In addition, the growing popularity of wearable health and wellness trackers, such as the Fitbit, Jawbone UP, the Apple watch, and others, has created an abundance of user-generated health data. Like the incidental health data derived from GPS or social media, user-generated health data is produced, shared, and exploited under poorly defined privacy and ownership policies (Lupton, 2016; Neff and Nafus, 2016).

In light of the growing importance of patients and consumers in the life cycle of big health data creation and exploitation, the need for clarity around the role of user-generated health data in commercial and scientific enterprises is pressing. When the Precision Medicine Initiative (PMI) was launched in the United States in 2015, it was described as "a new way of doing research that fosters open, responsible data sharing with the highest regard to participant privacy, and that puts engaged participants at the center of research efforts" (NIH, 2015). The premise of "open, responsible data sharing" rests upon the assumption that future uses of PMI datasets would not produce harmful unintended consequences for data donors, yet legal scholars and data scientists have shown that data privacy is virtually impossible to ensure (Ohm, 2010; Pasquale and Ragone, 2014). Moreover, little is known about how and why participants engage in data sharing, what privacy means to those participants, what individuals think researchers and businesses can and should do with their data, and what users think they might gain (or lose) by sharing their data beyond their personal social networks. This study contributes to the growing body of research on the role of big data, personal data, and data sharing in healthcare by illuminating how members of the general public, health researchers, and health information technology start-up companies understand the meaning and value of user-generated health data. While this study has global implications, it is primarily focused on the effects of policies governing health and social data in the United States.

We began this study with the broad question: "how is user-generated health data transforming ideas about health, both within and beyond medical contexts?" However, our research quickly identified the concepts of trust and privacy as particularly critical for shaping the value of user-generated health data, so we narrowed the focus of our interviews to prioritize those terms. We define "user-generated health data" as data captured through devices or software (purpose built or commercially available) and used outside of traditional clinical settings for tracking personal health data (such as wearable heart rate monitors, step-counters, and sleep trackers). For the purposes of this paper, we define "medical contexts" (used interchangeably with "traditional clinical settings") as those sites where formal doctor–patient interaction is governed by health law such as the Health Information Portability and Accountability Act of 1996 (HIPAA) and the U.S. Food and Drug Administration (FDA) approval process governing the use of medical devices, including some digital health tools. Our research explores how distinctions between clinical and nonclinical spaces and practices are changing in the context of mobile health technologies and user-generated health data. Therefore, while the blurring of boundaries between the clinical and the nonclinical, or between medical and health/wellness domains, may seem to suggest that these distinctions are becoming less relevant (Fiore-Gartland and Neff, 2015), we nonetheless recognize that existing rules define regulatory boundaries between consumer-facing software applications and devices (which do not require FDA approval and are not governed by HIPAA), and clinical-facing apps and devices (which are regulated by FDA and HIPAA). When considering how user-generated health data travels through social and information networks, the boundaries between the nonclinical and the clinical remain quite relevant, with significant implications for our study.

After the description of our research methods, the first section of this paper describes how the mobile technologies that have facilitated the rise of "user-generated data" have enabled new forms of autonomy for patients and new processes of health datafication, raising important questions about the meaning of privacy and sharing in this new context. The remaining

sections of the paper describe and analyze the key results of our interviews. In the second section, we discuss how the concept of trust shapes users' attitudes not only toward sharing their health data, but also toward their assessment of the significance of the data itself. The third section explains how concepts of privacy have become more flexible in relation to evolving attitudes about the value of user-generated health data, with significant consequences for users' willingness to agree to device and software "terms and conditions." In contrast, in the fourth section we analyze the growing unwillingness of individuals to participate as "human subjects" in health data science, despite their willingness to agree to corporate terms of use that entail commodification of their privacy, and the implications for models of data sharing, privacy, and trust.

Methods

We first conducted a literature review of academic, journalistic, and gray literature focused on key concepts in user-generated health data, such as quantification, big data, mobile technology, and digital health. The results highlighted the interconnections between industry and the academy, as health researchers are adapting consumer-facing wearable technologies in their work, while health technology companies are drawing on behavioral science to validate and promote the claims of their devices. Widespread consumer adoption of health-tracking devices has demonstrated the acceptance of these relatively new technologies outside of clinical settings. Our aim was to characterize and critically interrogate how different groups of stakeholders understood the concepts of trust and privacy through a methodological approach that bridged discourse analysis, ethics, and science and technology studies. The study protocol was approved by an Institutional Review Board.

On the basis of our literature review we identified three target populations to interview, according to the following inclusion criteria: participants must be healthy adults and either: researchers who interact with user-generated data, employees of a business that interacts with user-generated data, or members of the general public who interact with user-generated data. Researchers included behavioral and computational scientists, businesses included health information technology start-up companies (including software and device developers), and members of the general public included individuals who use wearable devices or apps to capture their own health data. Together, these three groups cover the spectrum of actors who, through their professional and everyday activities, shape the ideas around and practices of using technologies that produce user-generated health data.

We initially conducted informal, unstructured interviews with three researchers and three start-up companies to help us further identify core issues for these groups. On the basis of those interviews, we developed semistructured interview scripts for each cohort. The interview questions for the researchers and start-ups were closely aligned and focused on what kinds of data our interlocutors used in their research or business; what role user-generated data played in their work, whether they had to develop novel consent procedures or terms of use for user-generated data; how they saw this type of data as different from other forms of data, whether there were new business or research challenges that arose from user-generated data, whether new privacy and security issues emerged from this type of data; and what they saw as the major benefits of working with this new kind of data. We recruited participants from September 2015 to January 2016 by networking with local experts to identify 10 researchers and 10 start-up companies to interview. We completed nine researcher and six start-up interviews over four weeks from January to February 2016, mostly in person at their offices, occasionally over the phone.

The interview questions for end users (members of the general public) were shaped by published literature reporting users' attitudes toward health-tracking devices, as well as informal ethnography with users in the local community. The questions asked were what kind of device was used for tracking health data, what they used it for, when and why they started using it, how they use it on a daily basis, whether they see this kind of health data as different from data they might receive in a clinical setting, whether they share their data with anyone else, why or why not, whether they think anyone else has access to their data, whether they read the terms and conditions for the device, and what they like or dislike about using the device. Through convenience sampling over four weeks in February and March 2016, we conducted 17 interviews with an average length of 20 minutes each by approaching members of the general public in three highly trafficked urban parks. Our participation rate was approximately 80%. Upon completion of 32 interviews, the recordings were transcribed and the interviews were manually coded by six members of the research team, using an inductive approach to identify latent themes in the data.

Autonomy and health datafication in an age of user-generated culture

The emergence of user-generated health data—as distinct from clinical health data—is part of a larger zeitgeist of "user-generated culture" that has captured the attention of individuals, corporations, hospitals, and

governments within the past decade (Füller, 2016). Entities like Uber, AirBnB, and Tinder are keystone examples of how devices and the data they produce have transformed their respective industries through new patterns of digital intermediation (Benghozi and Paris, 2016). While user-generated health data appears to be part of a larger cultural trend in mobile device integration, healthcare is a unique domain with a specific set of histories, demands, and stakes that do not necessarily apply to rideshare networks, real estate tourism, or romantic match-making services.

Mobile health technologies now enable users to accrue large volumes of real time and longitudinal health data, using methods not typically possible in traditional clinics or "analog" self-tracking journals (Cortez et al., 2014). These user-driven practices generate new types of health data that avoid many of the infrastructures and actors traditionally involved in healthcare and health decision-making. As improved methods for collecting, processing, and storing large datasets are developed, the big health data generated by individual patients may redefine our conceptions of health, disease, and what it means to be a patient (Fox and Duggan, 2013; Topol, 2015). The practices surrounding user-generated health data do not merely convey information; they mediate medical knowledge and help to construct meaning that bridges health and medical domains (Neff and Nafus, 2016; Ostherr, 2013). These practices of technomediation provide an important context for understanding what contemporary scholars have called "datafication," a process of "rendering into data aspects of the world not previously quantified" (Kennedy et al., 2015). Practices of datafication also involve the transformation of existing data into actionable forms that generate diverse and unevenly distributed forms of value for their producers and consumers (Van Dijck, 2014). Contemporary practices of health datafication occur both within and beyond clinical settings, posing challenges to traditional understandings of agency and ownership of medical data (Health Information and the Law Project, 2015).

Alongside and overlapping user-generated health data's relation to "user-generated culture" is the emerging phenomenon of patient-generated health data (PGHD). Like user-generated health data, PGHD often relies on mobile devices to generate health data. Unlike user-generated health data, PGHD is typically enfolded within traditional healthcare ecosystems that include existing privacy infrastructure governed by HIPAA, the Common Rule, and other federal and state regulations (Deering et al., 2013; Thorpe and Gray, 2015). With user-generated health data and the mobile devices that produce them, issues related to privacy and data sharing do not simply evolve within an existing healthcare ecosystem, but potentially formulate an entirely new type of healthcare.

Importantly, user-generated health data from commercial devices are not easily integrated into clinical settings (Chung and Basch, 2015; Luxton et al., 2012). Most patients cannot simply bring their Fitbit data to their cardiologist and expect to receive recommendations based on those data. While a provider could "prescribe" the use of a commercial tracking device for a patient to monitor her cardiovascular activity, incorporating the data from that device into the patient's electronic health record (EHR) would pose significant legal and regulatory challenges. With few exceptions, user-generated health data presently has no place within formal EHR-based medical documentation systems, rendering it invisible in the majority of doctor–patient encounters (Kish and Topol, 2015). Conversely, health-related device and software companies operating outside of hospitals, clinics, and other HIPAA-protected zones face few restrictions on their exploitation of users' data, as consumers must agree to "terms and conditions" to activate and use the app. In many cases, those terms of use permit the parent company to sell users' health data to third parties, including marketers, advertisers, and other types of data brokers (Shklovski et al., 2014).

While the users who generate health data outside of clinical settings may be vulnerable to third-party exploitation, many see self-tracking tools that put health measurement and quantification into the hands of ordinary users as a democratizing force that challenges traditional doctor–patient knowledge hierarchies. Activists engaged in the Quantified Self and e-patient movements (Ferguson et al., 2007; Nafus and Sherman, 2014) seek to transform the process of health datafication into a process of health data making (Pybus et al., 2016) that generates value for the individuals whose bodies generate the data, rather than solely for the corporations who manufacture those devices or provide formal healthcare services to those bodies (Van Dijck and Poell, 2016). Ironically, concern for the need to protect patient health information through overly cautious adherence to HIPAA guidelines has constrained the expansion of patient autonomy into clinical domains, as new methods for sharing patient data, enabled by electronic communication technologies, have raised concerns regarding ownership, confidentiality, and control (Strauss, 2012; Wilkes, 2015).

Paradoxically, some users are more willing to share their health data on an app than with their healthcare provider (Wortham, 2016). This may result from the device's sociotechnical infrastructure: the social networking capacity that enables users to share their health data is often a key feature in product design and a central marketing component for many

health-related apps on mobile devices. With Fitbit, for example, social media connectivity allows users to compare their data and "compete" within their social networks (Nakhasi et al., 2014). Thus, the barriers to sharing user-generated health data within formal healthcare settings are elided by the seeming openness of consumer-facing health apps designed to cultivate unrestricted data sharing (Kim, 2014) outside of clinical settings. The contradiction between the restrictive view of data sharing within medicine and the permissive view of data sharing outside of medicine cultivates a sense of uncertainty among users about the value of privacy and trust on one hand, and openness and sharing on the other. The asymmetry of opportunities for user-generated data to serve the goals of patients inside versus outside of clinical settings points to the conflicting conceptual models that characterize these ecosystems today. These contradictions are giving rise to new attitudes toward privacy and sharing as well as new understandings of the meaning and value users can derive from quantified health data.

Trusting and sharing numerical data

A core tenet of science and technology studies is that empirical evidence is social and situated, rather than objective and neutral (Latour and Woolgar, 1979). Numerical data, in particular, are not to be trusted absolutely but instead considered as contingent outcomes of the social practices that yield them (Porter, 1996). Thus, all data are "user generated." Notably, our diverse sets of interviewees seemed to share this viewpoint as they reflected upon the importance of situating data within networks in order to divine the significance of given numbers. Each group enacted distinct practices to materialize user-generated data as a social object.

Following our preliminary interviews, the topic of trust in numbers, and trust in data, became an animating concern that directed the course of our study. Researcher–interviewees highlighted the fact that user-generated data came from diverse sources "in the wild" and as a result, they were less secure as evidence than data gathered through controlled experiments. By collecting user-generated data from novel sites, researchers expanded the scope of their work; however, the new methodologies raised concerns about how these new streams of data were to be interpreted and trusted. As one researcher told us: "The data we have now has surpassed our conceptual model's abilities to tell us exactly what to do." New models are needed in order to put the numbers into scholarly narratives. Interdisciplinary alliances that brought together diverse genres of expertise facilitated this practice.

In interviews with end users, the topic of trust centered on the submission of their personal data into worlds beyond their technological hardware. Individuals who were less professionally trained in the interpretation of numerical data gained insights into the personal and social significance of their data by sharing it with others. Like science and technology studies scholars, our interviewees situated themselves and their devices within networks (Haraway, 1988). When we asked how they understood the meaning of the term "user-generated data," our interlocutors emphasized its emergence from multiple origins and its circulation through multiple domains. User-generated data, they said, is marked by its immediacy and ubiquity, its "bigness," and "speed," as well as its travels. Encounters with user-generated data are organized through relations across scales and domains, from personal to institutional collaboration and from behavioral strategies to epistemological maneuvers. The particularities of these distinct assemblages guided users' management and interpretation of their data.

Surprisingly, our interviewees expressed little concern about sharing their user-generated health data with corporate actors. They expressed much greater interest in the ways that their data was purposefully shared with known members of their social networks. Several Fitbit users described sharing their daily step counts with others, and emphasized that viewing others' data inspired them to walk more. Their network was composed of themselves, their devices, and the friends that they shared their data with. Numbers were relative. By relating one's personal number to the number of a friend who they could socially situate—as a person of a certain age, with a certain job, in a certain location—these end users measured the significance of their own data. By interpreting their daily step count within the context of this network, they drew motivation that propelled their physical body onward.

In contrast, the researchers we interviewed interpreted data with attention to the diverse genres of expertise that made their research agenda possible. Every investigator who was involved in projects concerning user-generated data was part of a collaborative and interdisciplinary team. As one researcher said:

> I think modern science is all about teams now. It's like mapping the human genome happened because we threw really large, smart teams of people at that problem to be able map it. It's the same way now with a lot of the new stuff. [...] I think the old way of people toiling away solitarily in their lab are generally going away. I collaborate on—all my current grants have electrical engineers on them. They have computer scientists. They have computational scientists on them. They often have geneticists. I don't know what to do

> with any of those. [. . .] We work as a team and actually using big data, but there are specific people that actually do the computational models because they're far beyond me.

Bioengineers, software developers, psychologists, and others combined their complementary expertise in order to enact their research design. Each recognized the involvement of their collaborators as essential, often admitting that they were unqualified to perform that work themselves. In this way, the difficulty of trust in numbers materialized through user-generated data collection is resolved through trust in collaborators who together establish the viability of this type of data as evidence.

Businesses, on the other hand, situated data within networks composed of hospitals, physicians, patients, government regulators, and hackers—each with their own perspectives and capacities. They, too, entered relations with other genres of expertise to manage their enterprise. As an employee from one healthcare start-up said: "you have to get a consultant who's familiar with what you're collecting and familiar with that landscape in order to come and help you understand any regulation around it [. . .] there are social, moral, other considerations as well." Unlike other interviewees, the primary goal of start-ups was not to materialize numbers that could be put toward self-realization or scholarly argument; rather, their goals were financial. As such, they strategically managed the complexities of the networks they worked within. This task centered on the enclosure of data to ensure patient privacy and proprietary rights, and as a result the social voyages of data became less of an enactment of meaning and more of a threat to be managed. Despite the enhanced technological features of cloud computing, for example, an employee from another healthcare start-up warned that "pushing the data outside the hospital is a challenge, so if you wanted to store and process data in the cloud it's just not gonna happen right now because hospitals don't want to put their data outside the networks." Another employee noted that "instead of building things ourselves, we will use as much premade items as possible to reduce any of our risk." Indeed, while they stated that the emergence of large and comprehensive datasets could make healthcare more efficient and effective, they identified the barriers that privacy advocates and competing business enterprises placed upon the circulation of data as a hindrance.

Because of its immensity and immediacy, user-generated data offers unique possibilities to those who encounter it. However, these traits also make it difficult for any one individual to interpret this data in isolation from other individuals and other datasets. By locating user-generated data within networks, its significance came into sharper focus as points of contrast and genres of expertise were brought to bear on disconnected and incomplete numbers.

Privacy as flexible cultural artifact

Data privacy and security have become major topics of concern in the post-Snowden era (Pybus et al., 2016), with special emphasis on the vulnerability of health data (Dockery, 2016; Ornstein, 2015). As one interlocutor noted, health and financial data constitute sensitive objects that need careful management and protection. However, we found that concerns about privacy in healthcare differ in substance, depending on the actor's position and stakes in the chain of data collection, storage, and use. Thus, health data privacy is not a stable natural object that has value regardless of the subjects who enact it; rather, health data privacy is a multifaceted cultural artifact that becomes assembled and maintained within a complex ecology of alliances and disconnections.

A recent survey by the Pew Research Center found that most Americans "strongly agree" that maintaining privacy and confidentiality in their everyday activities is important (Madden and Rainie, 2015). Yet, we found that very few individuals we interviewed held these concerns. When asked whether they thought that their data was being used by anyone for purposes that they were unaware of, one respondent replied: "They might be. I don't really care if they do or not . . ." Almost all our interviewees agreed that they might have shared their health data with third parties, without being fully aware. Some assumed that corporations such as Apple collect their data automatically, with the purpose of producing more technologically sophisticated—and thus, "better"—services and devices. Most did not actively think about how their data was viewed by the companies that manufacture the devices and apps. When pressed, most felt that the manipulation of this data by other parties was innocuous, since it was likely only valuable in the aggregate, in their view.

While users were generally aware that consenting to a company's terms of use constitutes a legal contract, very few reported actually reading those agreements before consenting to them. One participant commented: "Do I ever read 'terms of use'? Did I actually read the consent form I just signed? No. I just agree to everything like I do for all of my Apple updates. Agree. Agree. Done. So, no." Attitudes like this one appear to be the norm, and they highlight the contrast between the widespread concern captured by the Pew survey described above and the casual attitudes associated with informal, social settings for user-generated health data sharing.

Similarly, a survey of over 10,000 users in 20 different countries (Internet Society, 2012) asked: "What are the main reasons you accept the terms and conditions as offered, without reading them?" A full 42% of respondents noted the length of the document, while 19% of respondents indicated that the legal terminology was difficult to understand, and 11% selected "I don't have a choice if I want to complete an activity that I need to complete." These responses indicate that users feel they have no agency in controlling access to their data. Notably, the inclusion criteria for this study selected for interviewees who already use some form of software or app to capture health data, and therefore, we could only include individuals who have already consented to the provider's terms of use. Indeed, the widespread use of self-tracking devices and smartphone apps offers a proxy measure of public willingness to agree to terms and conditions to facilitate participation in digital health.

But if personal privacy is as important as public debate and the experiences of researchers (described below) would suggest, further explanation of this behavior is needed. In his discussion of "digital market manipulation," legal scholar Ryan Calo describes the cognitive overload that users experience when faced with the prospect of reading through the multiple pages of "legalese" that constitute the average terms of use document:

> ... too much or extraneous information is said to underlie a host of departures from rational decision-making. For example, 'information overload' causes consumers to rely on heuristics or rules of thumb, shortcuts which are sometimes faulty. The phenomenon of 'wear out,' which suggests consumers tune out messages they see too often, renders product warnings less effective. (Calo, 2014: 1012)

The length, complexity, and ubiquity of these agreements may be leading end users to forego control over their health data because the "heuristics or rules of thumb" that they operate under lead them to believe that the entity who is collecting the data will not use it maliciously. It may surprise many users to know that the major wearable technology companies all reserve the right to share personal, identifiable data in the process of business deals (Fitbit, 2015; Garmin, 2014; Jawbone, 2014; Misfit, 2015). Furthermore, these companies have virtually no restrictions on the ways that they may use and sell aggregated, unidentifiable data that they collect on the users of their technologies. "[T]he law's always behind technology," one researcher commented, indicating that legal categories available today and newly developed technological solutions do not neatly map onto each other.

However, some users proactively reframe their concerns about privacy by emphasizing their acquired capacity to monitor their own health and fitness levels, practice preventative self-care, and thereby potentially avoid costly medical services. Although some expressed laughingly that they probably should care more about third-party use of their health data, the majority of our interviewees said they did not think about it much. Given the terms of use, interviewees indicated that they value using their health-tracking apps more than they value their data privacy. As several interviewees suggested, attitudes toward privacy in healthcare are changing, under the conditions of rapid technological change and its impact on patterns of sociality. Another interviewee observed,

> What we never thought we would post is being posted by the people who thought that they would never post- ... the definition of privacy will completely change as we move forward. ... We're actually quite adaptive. It will change. Maybe that's actually the reason why as a species we are very successful, because we change with what we feel has value for us.

Our research suggests that the concept of privacy itself is undergoing change in the public consciousness, and the legal system has not kept pace.

Human subjects in data science

Participants in data science research appear as sources of data and as protected legal subjects. Researchers working with user-generated health data thus require sophisticated technical knowledge and skills to deidentify collected data, police access to those data, and ensure that any public appearance or use of the data is in full compliance with the law. Yet these assurances have done little to persuade participants that their privacy will be protected in research settings. Several researchers engaged in the development of new health technologies reported considerable difficulty engaging study participants:

> When I would mention, hey, here is the type of data we collect. It immediately puts some people in a very alert, semi-panic mode. This is way too much information you are collecting about people. I think the reason is that we hear a lot of stories of how perhaps different companies know a lot about us. Or maybe government knows a lot about us. I try to tell them there's a difference between the two approaches. One is being, you are being tracked without you being told, and without you knowing who is likely looking at it ... However, in this situation the way we collect data actually is completely different. It's a fully informed situation where the

> participant or the patient is actually told what information will be collected, right. Even the method of collection. At the same time, they're also told who will likely look at it, and what they plan to do with it. Then, in fact, they are given guarantees that this data will be sandboxed to the point that only these two or three people can actually look at it.

The reticence to consent to research governed by Common Rule and HIPAA regulations contrasts sharply with our interviewees' reported comfort with data collection by for-profit companies who are not beholden to HIPAA guidelines at all (see Comstock, 2016).

This contrast reflects larger contradictions around issues of privacy, sharing, and trust as health data crosses boundaries between clinical and social domains. As Metcalf and Crawford (2016) and Zwitter (2014) have argued, the field of big data research is rapidly outpacing the ability of institutional ethics regulations to keep pace, leading to major disputes over the meaning of ethical human subjects research in data science. ''[H]ow a particular patient feels versus how the general public feels about the same data'' matters, as one researcher mentioned, gesturing at different emotional responses and conceptual vocabularies around privacy that he encounters in his work with user-generated health data.

The cultivated loyalty to privacy as a weapon in the cyberwar over personal data prompts broader, more systemic thinking about the conditions in which health becomes expressed as an individual responsibility and concern. As various thinkers (Foucault, 2009; Rabinow and Rose, 2006) have argued, health and biological vitalities are major analytics for governing individuals and populations in contemporary states. In neoliberal environments with receding welfare provisions and strong emphasis on personal autonomy and independence, resilient health and vitality become matters of individual responsibility and choice (Rose, 2006). Instead of addressing larger conditions that produce toxic environments, systemic poverty, and inadequate social resources, contemporary discourses of corporate and state care nurture the ideas of health and healthcare as primarily individual concerns and responsibilities (Jain, 2012). In this context, the imperative to track and take control of one's health, as a privacy that must be defended, solidifies the idea that individuals must take full credit for maintaining and investing in their own health and wellness.

At the same time, the connective capacities of digital technologies create novel opportunities for alternative ''data-making'' practices (Pybus et al., 2016) that challenge the idea of personal health data as private property that only provides individual value, security, and wealth (Agus, 2016). One example is *PatientsLikeMe*, an online platform that enables sharing of user-generated health data with a network of strangers committed to the idea that patient-generated knowledge would benefit the community by making health management more accessible, more supported, and less isolating. Engaging user-generated data as a tool to connect and relate to others, to offer encouragement, or to foster competitive spirit gives room to a different kind of sociality in which data is not a threat but a component of the very social fabric. However, as Van Dijck and Poell (2016) have argued, data use is loosely regulated on many online health platforms, allowing for commodification and exploitation by actors with less community-minded goals, raising once more the question of who truly benefits from big health data.

Conclusion

One of the surprising results of our interviews was that, despite the constant reporting of large-scale data breaches around the world (Comey, 2016), our interlocutors felt little concern about sharing their user-generated health data with corporations. Why? Some interviewees suggested that the transactional nature of their consent overrode any concerns about privacy; individuals had already decided that they wanted to use a device or piece of software, so they consented to the terms of use in exchange for access to the product they desired. Campbell and Carlson (2002) note that the commodification of privacy is often presented as a necessary feature of consumer access to popular platforms such as Facebook, and the social sharing features of technologies that produce user-generated health data further encourage users to see ''health'' as a commodified benefit of the exchange of personal data. Turow et al. (2015) have called this ''the tradeoff fallacy,'' noting that most Americans feel it is impossible to limit access to their data, and instead see digital profiling as inevitable. Our research also suggests that the ''black box'' (Pasquale, 2015) surrounding these transactions may obscure the true nature of the exchange, with potentially harmful results, including the widespread perception that it is impossible for users to opt out of participation in surveillance practices (Elmer, 2003). As Wilbanks and Topol (2016) have argued, ''undisclosed algorithmic decision-making'' based on user-generated health data could lead to ''discriminatory health actions'' against the very users who willingly shared their own data.

Entangled with the emergent understanding of privacy as flexible and contextual, our research also identified a new concept of communities of data sharing. Many of our interviewees expressed a willingness to create and share personal health data with other users. The rhetoric of personalization and sharing

appears to facilitate an understanding of user-generated health data that deemphasizes the risk of exploitation in favor of loosely defined benefits to individual and social well-being. In this model, the concept of personalization emerges from the purposeful creation of networks of family members or friends with whom individuals share data to motivate or make sociable their data-tracking activities. The sense that each of these networks is highly personal to its creator seems to override awareness of the other, less benevolent entities with whom the data is being shared.

An interesting corollary was the idea expressed by several researchers that data sharing by the general public would lead to greater improvements in health outcomes than previously possible through lab-based research. The rhetoric of openness and sharing has begun to frame data exchanges among researchers just as it has shaped the practices of casual users in online health platforms. The complex nature of user-generated health data research has also given rise to the formation of multidisciplinary research teams whose members must participate in "sharing" across traditional disciplinary boundaries to accomplish their research objectives. In this sense, the new model of communal data sharing can be seen as having a transformative effect on the conduct of scientific big health data research as well.

Importantly though, many of the data scientists we interviewed described significant challenges in recruiting research participants, despite the relaxed attitudes individuals expressed about consenting to terms and conditions that enable corporations to freely exploit their users' data. Our study suggests that when data privacy is explicitly foregrounded in the process of obtaining verbal consent, and participants are addressed as individuals rather than as anonymous consumers, the risk of malevolent data exploitation appears significantly more threatening. Ironically, researchers who are required to participate in ethics review procedures and follow explicit protocols for data privacy, security, and storage are subject to considerably more suspicion by members of the general public than are the corporations that overtly participate in data profiling with far less ethical supervision. Under these asymmetrical circumstances, researchers are penalized for raising public awareness of the procedures required ethically to conduct health data research, while businesses are free to benefit without restriction.

What emerges from these seemingly contradictory attitudes about sharing user-generated health data is a fluid, contextually specific, and social conception of privacy. A major implication of this finding is that there is a significant disconnection between the regulatory policies governing the sharing of health data for research and patient care, on one hand, and those policies governing corporate practices on the other. Moreover, there is an additional disconnection between public discourse on threats posed to personal privacy by data piracy, security leaks, and identity theft on one hand, and public interest in the terms and conditions that actually govern access to their user-generated data on the other. These contradictions suggest that the regulatory frameworks for managing the risks of sharing user-generated health data need an overhaul that brings them into closer conformity with the current attitudes of the general public. At the same time, the casual permissiveness that characterized many of our interlocutors' responses to our terms and conditions question suggests that, in addition to updating our legal frameworks for protecting and sharing user-generated health data, we also need to engage in a more robust public dialog about the potential benefits and harms of openly sharing health data. With recent studies demonstrating the harms embedded in artificial intelligence algorithms that replicate racial and other biases of their human programmers, as well as the growing intermediation of data sources that together might be capable of revealing sensitive personal data (Crawford and Calo, 2016), there is a clear need for regulations that offer consumers easily comprehensible terms of use, with opportunities to opt out of surveillance. Future research on user-generated health data that identifies and explains the effects of participating in this ecosystem—both outside and inside of clinical settings—will provide much needed guidance to policymakers and patients as regulations governing data sharing attempt to catch up with practices in the wild.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Agus DB (2016) Give up your data to cure disease. *The New York Times*, 6 February.

Benghozi P and Paris T (2016) The cultural economy in the digital age. *City, Culture and Society* 7(2): 75–80.

Calo R (2014) Digital market manipulation. *The George Washington Law Review* 82(4): 995–1051.

Campbell JE and Carlson M (2002) Panopticon.com: Online surveillance and the commodification of privacy. *Journal of Broadcasting and Electronic Media* 46(4): 586–606.

Chung AE and Basch EM (2015) Potential and challenges of patient-generated health data for high-quality cancer care. *Journal of Oncology Practice* 11(3): 195–197.

Comey J (2016) Humility, adaptability, and collaboration: The way forward in cyber security. *Speech delivered at FBI/Fordham University international cyber security conference,* New York City, New York, 27 July 2016. Available at: https://www.fbi.gov/news/speeches/humility-adaptability-and-collaboration-the-way-forward-in-cyber-security (accessed 1 August 2016).

Comstock J (2016) How consumer health, fitness devices reveal HIPAAS's blurry lines. Available at: http://mobihealthnews.com/content/how-consumer-health-fitness-devices-reveal-hipaas-blurry-lines (accessed 28 May 2016).

Cortez NG, Cohen IG and Kesselheim AS (2014) FDA regulation of mobile health technologies. *New England Journal of Medicine* 371(4): 372–379.

Crawford K and Calo R (2016) There is a blind spot in AI research. *Nature* 538(7625): 311–313.

Deering MJ, Siminerio E and Weinstein S (2013) Issue brief: Patient-generated health data and health IT. *Office of the National Coordinator for Health Information Technology*, Washington, D.C., 20 December 2013, pp. 1–11. Washington, D.C: Office of the National Coordinator for Health Information Technology.

Dockery S (2016) The morning risk report: Study shows deep flaws in health-care cybersecurity. *The Wall Street Journal*, 29 June.

Elmer G (2003) A diagram of panoptic surveillance. *New Media and Society* 5(2): 231–247.

Ferguson T, et al. (2007) *e-Patients: How they can help us heal healthcare*. Report for Robert Wood Johnson Foundation. Available at: http://www.e-patients.net/e-Patients_White_Paper.pdf (accessed 1 August 2016).

Fiore-Gartland B and Neff G (2015) Communication, mediation, and the expectations of data. *International Journal of Communication* 9: 1466–1484.

Fitbit (2015) Terms of service. Available at: https://www.fitbit.com/legal/terms-of-service (accessed 28 May 2016).

Foucault M (2009) *Security, Territory, Population*. New York: Picador.

Fox S and Duggan M (2013) *Tracking for health*. Report for Pew Research Center. Available at: http://www.pewinternet.org/2013/01/28/tracking-for-health/ (accessed 1 August 2016).

Füller J (2016) The power of community brands. In: Harhoff D and Lakhani KR (eds) *Revolutionizing Innovation*. Cambridge: MIT Press, pp. 353–376.

Garmin (2014) Terms of use. Available at: http://www.garmin.com/en-US/legal/terms-of-use (accessed 28 May 2016).

Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.

Health Information and the Law Project (2015) Who owns medical records: 50 state comparison. Report for George Washington University, Hirsh Health Law and Policy Program. Available at: http://www.healthinfolaw.org/comparative-analysis/who-owns-medical-records-50-state-comparison (accessed 1 August 2016).

Internet Society (2012) Global internet user survey 2012. Available at: http://www.internetsociety.org/surveyexplorer/online-privacy-and-identity/what-are-the-main-reasons-you-accept-the-terms-and-conditions-as-offered-without-reading-them-14/ (accessed 28 May 2016).

Jain SL (2012) Cancer butch. *Cultural Anthropology* 22: 501–538.

Jawbone (2014) UP terms of use. Available at: https://jawbone.com/legal/up/terms (accessed 28 May 2016).

Kennedy H, Poell T and Van Dijck J (2015) Data and agency. *Big Data & Society* 2(2): 1–7.

Kim N (2014) Three's a crowd: Towards contextual integrity in third party data sharing. *Harvard Journal of Law and Technology* 28(1): 325–347.

Kish L and Topol E (2015) Unpatients: Why patients should own their medical data. *Nature Biotechnology* 33(9): 921–924.

Krumm J, Davies N and Narayanaswami C (2008) User-generated content. *IEEE Pervasive Computing* 7(4): 10–11.

Latour B and Woolgar S (1979) *Laboratory Life*. Beverly Hills, CA: Sage.

Leaf C (2015) The biggest share in the sharing economy. *Fortune*, 7 August. Available at: http://fortune.com/2015/08/07/digital-health-data/ (accessed 1 August 2016).

Lupton D (2016) *The Quantified Self*. London: Polity.

Luxton DD, Kayl RA and Mishkind MC (2012) mHealth data security. *Telemedicine and e-Health* 18(4): 284–288.

Madden M and Rainie L (2015) Americans' attitudes toward privacy, security, and surveillance. Report for Pew Research Center. Available at: http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/ (accessed 2 May 2016).

Metcalf J and Crawford K (2016) Where are human subjects in big data research? *Big Data & Society* 3(1): 1–14.

Misfit (2015) Misfit terms of use. Available at: http://misfit.com/legal/terms_of_use (accessed 28 May 2016).

Nafus D and Sherman J (2014) This one does not go up to 11: The quantified self movement as an alternative big data practice. *International Journal of Communication* 8: 11.

Nakhasi A, Shen AX, Passarella RJ, et al. (2014) Online social networks that connect users to physical activity partners. *Journal of Medical Internet Research* 16(6): e153.

National Institutes of Health, Precision Medicine Initiative (2015) About the precision medicine initiative cohort program. Available at: https://www.nih.gov/precision-medicine-initiative-cohort-program (accessed 1 August 2016).

Neff G and Nafus D (2016) *Self-tracking*. Cambridge: MIT Press.

Ohm P (2010) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 57: 1701–1777.

Ornstein C (2015) Your health records are supposed to be private. They aren't. *The Washington Post*, 30 December.

Ostherr K (2013) *Medical Visions: Producing the Patient Through Film, Television, and Imaging Technologies*. New York: Oxford University Press.

Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.

Pasquale F and Ragone TA (2014) Protecting health privacy in an era of big data processing and cloud computing. *Stanford Technology Law Review* 17: 595–653.

Porter TM (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Pybus J, Coté M and Blanke T (2016) Hacking the social life of big data. *Big Data & Society* 2(2): 1–10.

Rabinow P and Rose N (2006) Biopower today. *BioSocieties* 1: 195–217.

Raghupathi W and Raghupathi V (2014) Big data analytics in healthcare. *Health Information Science and Systems* 2(3): 1–10.

Rainie L and Wellman B (2014) *Networked: The New Social Operating System*. Cambridge: MIT Press.

Rose N (2006) *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-first Century*. Princeton, NJ: Princeton University Press.

Sarasohn-Kahn J (2014) Here's looking at you: How personal health information is being tracked and used. Report, California Health Care Foundation, July 2014.

Shklovski I, Mainwaring SD, Skúladóttir HH, et al. (2014) Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In: *Proceedings of the 32nd annual ACM conference on human factors in computing systems-CHI '14*, Toronto, Canada, 26 April–1 May, pp.2347–2356. New York, NY: Association for Computing Machinery.

Steinhubl SR, Muse ED and Topol EJ (2015) The emerging field of mobile health. *Science Translational Medicine* 7(283): 283rv3.

Strauss LJ (2012) Patient privacy—Then and now. *Journal of Health Care Compliance* 61: 19–61.

Thorpe JH and Gray EA (2015) Big data and public health: Navigating privacy laws to maximize potential. *Public Health Reports* 130(2): 171–5.

Topol E (2015) *The Patient Will See You Now: The Future of Medicine Is In Your Hands*. New York: Basic Books.

Turow J, Hennessy M and Draper N (2015) The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation. Report, Annenberg School for Communication, University of Pennsylvania. Available at: https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf (accessed 29 July 2016).

Van Dijck J (2009) Users like you? Theorizing agency in user-generated content. *Media, Culture and Society* 31(1): 41–58.

Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society* 12(2): 197–208.

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

Wilbanks J and Topol E (2016) Stop the privatization of health data. *Nature* 535: 345–348.

Wilkes JJ (2015) The creation of HIPAA culture: Prioritizing privacy paranoia over patient care. *BYU Law Review* 5(7): 1213–1249.

Wortham J (2016) We're more honest with our phones than with our doctors. *The New York Times Magazine,* 23 March.

Zwitter A (2014) Big data ethics. *Big Data & Society* 1(2): 1–6.