

## ARTICLE

### Privacy, Data Mining, and Digital Profiling in Online Patient Narratives

Kirsten Ostherr  
Rice University  
kostherr@rice.edu

#### Abstract

Practices of health datafication and inadequate privacy policies are redefining the meaning of online patient narratives. This article compares patient-driven illness narratives and clinic-driven illness narratives to uncover a set of unrecognized assumptions about trust and privacy in health discourses. Specifically, I show how the open sharing of patient stories in social media, blogs, and other public domains collides with privacy regulations and normative assumptions in the US health care system that prevent integration of those stories into electronic health record (EHR) systems. I argue that publicly told stories based on personal experiences of illness are valuable sources of health care information in part because they are subjective, richly detailed, and open ended. Yet, precisely because of their public nature, these patient stories are unprotected sources of data that are barred from integration into health care data ecologies where clinical action takes place. Consequently, an impermeable barrier exists between the officially sanctioned accounts in the clinical record and the contextual richness of patient stories on the social web. The tensions between these two approaches to

Ostherr, K. (2018). "Privacy, Data Mining, and Digital Profiling in Online Patient Narratives". *Catalyst: Feminism, Theory, Technoscience*, 4 (1), 1-24.  
<http://www.catalystjournal.org> | ISSN: 2380-3312 © Kirsten Ostherr, 2018 | Licensed to the Catalyst Project under a Creative Commons Attribution Non-Commercial No Derivatives license

narrative and data create an opening for exploitative digital profiling practices that can — and already do — harm patients. Examples are drawn from Hugo Campos and Medtronic, PatientsLikeMe, Apple Health Records, Google Health, Microsoft Health Vault, IBM Watson Health, and OpenNotes.

## Introduction

Digital technologies are redefining health, communication, and what counts as human connection in medicine through practices of datafication that quantify formerly intangible aspects of human experience (Mayer-Schonberger & Cukier, 2013; Van Dijck, 2014). Amid the growth of virtual clinics and remote health care, these developments give rise to an urgent question: How does health datafication impact our understanding of embodied knowledge in medicine? Clinically-driven digital health initiatives draw on the rhetoric of disembodiment, suggesting that many of the formerly “human” roles in health care can be performed by artificial intelligence (AI) or virtual telehealth encounters. In contrast, patient-driven digital health initiatives suggest that individual illness stories form the embodied foundation for practices of do-it-yourself (DIY) biohacking, crowdsourcing, and open data sharing that can facilitate new models of peer-to-peer health care. Entangled in these divergent views of the future of medicine are conflicting claims about the locus of knowledge in health care: Does meaning reside in the data, or in the stories? By exploring the different kinds of illness narratives that circulate inside and outside clinical settings, this essay will uncover a critically important but unexplored set of assumptions about trust and privacy in health discourses. Specifically, I will show that patient communities online assume an ethos of openness, trust, and sharing of stories and data with little expressed concern about privacy. In contrast, clinical environments value privacy over sharing of stories and, consequently, emphasize data over narrative, as a more secure and objective mechanism for transmitting information. The tensions between these two approaches to narrative and data create an opening for

exploitative digital profiling practices that can — and already do — harm patients.

The networked mediation of illness narratives takes place through the circulation of digital signals in a wide range of domains. In this essay, I want to focus on the “trading zones” (Galison, 1997) between two critically important discursive fields: patient-driven illness narratives and clinic-driven illness narratives. Specifically, I will explore how the proliferation and open sharing of patient stories in social media, blogs, and other public domains collides with privacy regulations and normative assumptions in the US health care system that prevent integration of those stories into electronic health record (EHR) systems. I will argue that publicly told stories based on personal experiences of illness are valuable sources of health care information in part because they are subjective, richly detailed, and open ended. Yet, precisely because of their public nature, these patient stories are considered to be unprotected sources of data that are barred from integration into health care data ecologies where clinical action takes place. Instead, contemporary EHRs quantify disease through biomedical datapoints such as blood pressure, weight, height, pulse, temperature, oxygen saturation, respiratory rate, pain level, medications, procedures, and so forth (Zhou et al., 2016). An impermeable barrier exists between the officially sanctioned accounts in the clinical record and the contextual richness of patient stories on the social web. In the final section of this essay, I will present a discussion of alternative practices, concluding with a warning about the risks of opening up patients’ detailed personal narratives to be mined and exploited through digital profiling practices that might harm their narrators as much as they help them.

### **Part I: Patient-Driven Illness Narratives**

Patients have shared their illness narratives for centuries, but the new platforms for storytelling in the digital age have radically expanded the scale of the audience and, consequently, the impact of individual stories.

As a result, patient stories that previously would have circulated privately, among family and friends, are now open to any person with an Internet connection. Notably, medical spaces are excluded from these publics by design. Although patient stories have been recognized as the foundation of diagnosis since the days of Osler (1901), they claim ever diminishing space in contemporary clinical documentation. Federal incentives have recently pushed more clinicians to adopt EHRs, leading to widespread complaints about the role of these technocratic interfaces in dehumanizing the clinical encounter, for both doctors and patients (Hsiao, Chun-Ju & Hing, 2012; Shanafelt et al., 2016). As I will discuss below, EHRs are widely seen as incapable of capturing important experiential and affective dimensions of the patient's story and, consequently, are seen as a threat to the practice of narrative medicine. Therefore, while patient-driven illness narratives are circulating more openly than ever online, the clinical repositories for these narratives are becoming more constrained and restrictive in their communicative abilities.

Patients' illness narratives have emerged as vivid counternarratives to official accounts, challenging traditional models of biomedical expertise as they circulate through robust networks of engaged citizens concerned with health care, often termed "e-patients" (Ferguson & Frydman, 2004). Members of these groups have developed extensive online activist communities, demand access to their own health data, conduct crowd-sourced experiments, and "hack" health problems that traditional medical experts have failed to solve. In this way, e-patients can be seen as practicing a form of "citizen science" (Hand, 2010) in medical contexts as a means of creating alternative models of patient-physician partnership. Critically, e-patients claim a unique form of expertise based on the embodied experience of living with illness (as a patient or a caregiver), and they emphasize the value of openness and sharing for themselves and other e-patients as a means of empowerment. As this movement extends earlier, pre-digital models of patient activism (Dumit, 2006; Epstein, 1996), it becomes enmeshed in networks of health

datafication that blur the boundaries of commercial, clinical, and everyday health activities (Ruckenstein & Schüll, 2017).

The expansion of platforms for sharing patient stories is part of a larger transformation of technologies of representation rooted in the emergence of the social web. While networked publics existed and communicated through analogue technologies long before the rise of the Internet (Jenkins, 2009), the possibility for subjects to broadcast their own narratives is a fundamentally important feature of the digital age that builds on the concept of the “open Internet” (Rainie & Wellman, 2012). While many scholars have described both the liberatory as well as the repressive features of global information and communication technology (ICT) networks (Tufekci, 2017), the aspect I want to highlight here is the role of the Internet in circulating new models for understanding the value of intimate stories as alternative forms of data-making that can be used to build things, such as counter-publics centered around healing and care.

Participants in the Quantified Self and e-patient movements take an active stance toward curating their own stories and health data (Ferguson and the e-Patient Scholars Working Group, 2007), transforming the process of health datafication into a process of health “data-making” (Nafus & Sherman, 2014; Pybus, Coté, & Blanke, 2016) that aims to benefit the people whose bodies generate the data rather than those who manufacture their device or act as data brokers (Van Dijck & Poell, 2016). Storytelling in these online patient communities can be seen as a means of pursuing greater patient autonomy that challenges paternalistic models of health care.

The emergence of patient stories online through tweetchats (e.g., #hcsn), personal blogs (such as The Liz Army), illness communities (e.g., Association of Cancer Online Resources, or ACOR), and “health journal” sites (e.g., CaringBridge) has been widely recognized as critically important, even life-saving, to the patients and caregivers who write and read them (deBronkhardt, 2013; Han & Wiley, 2013). Indeed, the explosion of patient stories online has given rise to digital health makerspaces,

including collaborative platforms for data collection and exploration (e.g., Crohnlology and PatientsLikeMe), and biomedical device hacking collectives (e.g., Do-It-Yourself-Pancreas-System, or DIYPS, and Nightscout) that can challenge exploitative models of health datafication. In these e-patient ecosystems, implicit trust in authors and audiences is underscored through celebration of openness and transparency across the sites. As e-patient blogger Liz Salmi (2017) notes, “Unlike a traditional journal, my blog has no privacy settings. People from around the world can read and learn, through my experience, about what it is like having brain cancer.” She goes on to explain the value of that openness: “Through logging my health and interactions with health professionals I am able to document my own medical history, and have been able to publically show that a frightening diagnosis has developed over time into something more akin to a chronic condition.” Here, personal health data, presented in narrative form, provides an intimate, experiential account of living with illness that is grounded in contextual detail and rooted in the author’s perspective on her illness, rather than in the biomedical model of disease.

While many e-patients turn to self-narration as a source of empowerment, there is widespread recognition in the community that control over documentation — the conversion of narrative into data — is a tremendous source of power within traditional medical organizations that e-patients have been barred from accessing. Building on citizen science models of engagement, e-patients have demanded access to their own medical datasets, viewing them as sources for telling their stories in new ways that can fundamentally alter the knowledge hierarchies associated with proprietary, private, closed biomedical data ecosystems. One example is the case of Hugo Campos, who has fought to gain access to the data generated by his implantable cardioverter-defibrillator (ICD), a small electronic device designed to shock his heart in the event of an arrhythmia. While his ICD sends data to his doctor on a daily basis, Campos can only view the data by making an appointment with his doctor and requesting a printout (Standen, 2012). Campos

sought open access to his data to take greater control over his own health, but he encountered major barriers from the device manufacturer, Medtronic (Singer, 2011; Marcus & Weaver, 2012). After years of activism on this issue, Campos was honored by the Obama White House in 2015 as a “Champion of Change” for “data liberation” (White House, 2015) but he is still fighting to gain full, open access to his ICD data (Campos & Sebastian, 2015).

Campos and other e-patients such as Regina Holliday (“Regina Holliday’s Medical Advocacy Blog,” 2009-) and Dave deBronkart (“e-Patient Dave: Toward a Science of Patient Engagement,” 2009-) often frame their critique of closed medical documentation systems in terms of privacy and ownership, asking: Who owns the data? Who controls access to the data? In whose interest is “privacy” invoked — the patient’s or the manufacturer’s? In Campos’s case, it is the story of his ongoing struggle, not any insights from the data itself, that earned him recognition from President Obama. The “Biomedical TechnoService Complex, Inc.” (Clarke et al., 2010: 57) responded to Campos’s demands by framing big health data as resistant to narrative translation. As David Steinhaus of Medtronic explained, “In principle, the company is fine with giving patients their data directly. The problem is how to format it so it’s useful to everyone. As is, the raw data is highly technical. Would patients understand it? You want to do it in a way that makes sense for the patient, that they can interpret correctly, so they don’t generate a lot of angst and difficulty for the physician or anyone else” (Standen, 2012, n.p.). This example demonstrates how the expression of affect through illness narratives can be used to frame patients as overly emotional and therefore incapable of objective, rational interpretation of data. Advocates of narrative medicine have long argued that we need access to personal, descriptive details to recover the important social, emotional, lived dimensions of experience that are crucial to the patient’s story and healing. However, aligning those often-ignored elements of patient experience with narrative, rather than with data, reproduces a linkage of affective storytelling with unreliable subjectivity, while reaffirming the purported objectivity of “pure” data

(Daston & Galison, 2007). Regulatory structures that keep patient narratives out of EHR systems reproduce this false opposition, increasing the need for patients to tell their illness stories in online publics, while decreasing the value of those stories within corporate health care systems.

And yet, as the possibilities for personal stories to morph into large, richly detailed databases become apparent through online sites, data donors and researchers are beginning to explore new models for thinking about narrative through a quantitative lens. For example, the site PatientsLikeMe has functioned since 2004 as an open forum for patients to describe symptoms, side effects, treatment regimens, and other experiential variables like mood, sleep, stress, and activity, so that participants can search for others with their condition and learn from their peers. The site quickly evolved into a vast archive of narratives that could be datafied and mined for larger-scale insights. Despite ongoing expressions of concern by medical experts about patients using “Dr. Google” to self-diagnose (Swaminath, 2016), participants found the PatientsLikeMe site to be a powerful resource that could be used to inform or challenge traditional scientific studies (Eichler et al., 2016; Okun & Goodwin, 2017).

PatientsLikeMe promises to be “honest, open, and transparent” about their uses of community members’ personal health data. They argue that “patients know what it’s like to live with disease, and they can teach us more than we can teach them,” while framing patient stories as a form of data (“About PatientsLikeMe,” n.d.). The site emphasizes the importance of personal narratives that, when documented on the platform, can convert experience into a form of evidence (“Our Science,” n.d.). Yet, this patient-centered counternarrative exists in a tenuous relationship to biomedical models of value and risk. The very conditions of possibility for the e-patient movement to emerge overlap with expanding practices of datafication that digital health initiatives, especially those concerning population health, seek to exploit (Ostherr, 2018). As a result, patient narratives are easily datafied online (even while

being denigrated as subjective), while individuals' own health data are not readily available to them.

The emphasis on privacy in regulations such as the Health Insurance Portability and Accountability Act (HIPAA) that aim to protect patient health information has limited the expansion of patient voices and autonomy in clinical domains, inadvertently pushing these stories out into public discourse instead.<sup>1</sup> In this sense, privacy norms and regulations for clinical contexts are indirectly responsible for putting patient data at greater risk for digital profiling and exploitation. Consequently, new methods for sharing patient data such as PatientsLikeMe generate unevenly distributed forms of value that do not always benefit the data "donors" as much as they should (Ostherr et al., 2017). Some users of the site have raised concerns about the partnerships between PatientsLikeMe and health care stakeholders, such as pharmaceutical companies AstraZeneca, Avenir, BioGen, Boehringer Ingelheim, Bristol-Myers Squibb, Genentech, Janssen, Novartis, Pfizer, and others, as well as health insurance providers Aetna, Bupa, CoPatient and Healthy Business Group (PatientsLikeMe, "Partners," n.d.). As one site user commented, "I did sign up, but how do I know drug companies or others won't use it to try and sell me something? What about my insurance company or employer getting at this data somehow? You're using my full name right here. It would seem a simple thing for someone I don't want to search the site and find out intimate details about my health record. Am I wrong?" (Allison, 2016). While PatientsLikeMe insists that it is transparent about these partnerships and its role in supporting its patient-centered business model, researchers have noted that data use is loosely regulated on many online health platforms (Agus, 2016; Van Dijck, 2014; Van Dijck & Poell, 2016). The lack of restrictions allows for commodification and exploitation by actors with less community-minded goals, raising the important question of who truly benefits from online sharing of personal health data.

## Part II: Clinic-Driven Illness Narratives

The role of illness narratives in clinical settings has long been recognized by practitioners of narrative medicine. The field has existed since the early 1970s, pioneered by scholars who published groundbreaking research demonstrating the value and clinical relevance of telling and hearing stories of illness (Charon, 2006; Charon et al., 2017). Clinician advocates of narrative-based medicine challenge the reductionism that can emerge from data-driven practices of evidence-based medicine, recognizing instead “the narrative features of *all* data and the evidentiary status of *all* clinical text” (Charon & Wyer, 2008: 297, emphasis in original). When narrative techniques are taught as part of medical school curricula, they are used to cultivate new approaches to communication among doctors and other health care providers, rather than among patients. Yet, they share a recognition voiced by e-patients: meaningful personal health data emerges from intimate stories rooted in the embodied experience of illness. Without the contextual richness found in these stories, the explanatory power of biomedical data is greatly diminished. However, under current privacy regulations, clinical EHRs only capture narrowly defined, quantitative biomedical data (Zhou et al., 2016). Despite the abundance of narrative details streaming through social media networks, the boundaries between the open web and HIPAA-protected zones cannot presently be crossed.

Since the early 2000s, technology companies such as Microsoft and Google have embraced a version of narrative medicine through strategic acknowledgement of the value of online patient stories, though with different objectives in mind. Because they operate in the public sphere, these companies can access the multi-layered “social life” of users’ health narratives by integrating diverse sources of data from non-health-related domains. Early efforts to construct consumer-facing EHR platforms like Microsoft Health Vault (2007) and Google Health (2008) can be seen as attempts to bring together the contextual data captured through online search and related activities with more traditional

biomedical markers of health and disease. Although these initial forays failed, with Google Health closing shop in 2011, and dwindling use of Microsoft's still-functioning platform (Thorp-Lancaster, 2016), the vision they pursued shapes the health data industry and its privacy policies today. Notably, Google researchers are at the forefront of investigations into development of deep learning methods for mining EHR data to build predictive models for patient health outcomes. While such studies promise to "drive personalized medicine and improve health care quality" (Rajkomar et al., 2018: 1), they also threaten to aggregate data from across the social web and the life span, thereby producing new levels of digital profiling that threaten new harms to patients' privacy.

Further threats arise with the recent announcement of Apple's update to its iPhone operating system that provides a purportedly "effortless solution" to integrating EHR data from multiple health care providers on the patient's iPhone (Apple, 2018). Media coverage of this new app noted the failures of Microsoft and Google to achieve widespread adoption of their own health data platforms, noting that Apple may succeed where others failed for two major reasons: the intimacy of smartphones as personal technologies, and the company's reputation for industry-leading privacy standards (Johnson, 2018). Indeed, Apple's website emphasizes its neutral role in EHR data transactions, stating, "Apple is not creating, receiving, maintaining, or transmitting protected health information for or on behalf of a covered entity or business associate" ("Healthcare," 2018, n.p.). While intended as an assurance of the privacy and security of its app, in essence, this legalese means that Apple is not subject to HIPAA regulations, leaving open the possibility for less overt uses of patient data through indirect means. Health information stored on Apple smartphones may be protected, but once the data is backed up on the iCloud platform, it can be mined not only by Apple (Pagliery, 2016), but also by Google, Microsoft, and Amazon when the data are backed up on servers operated by those companies (Axon, 2018; Novet, 2018).

Indeed, these threats are not merely hypothetical. A recent study

of what happens to the data we enter on popular mobile apps and their associated websites found that personally identifying information such as name, email address, location data, and other behavioral information is widely shared without users' knowledge. The third-party domain that receives sensitive data from the most apps is Google, followed by Apple and Facebook (Zang, Dummit, Graves, Lisker, & Sweeney, 2015: 3). Risks of exposure of sensitive information range from apps using unencrypted connections that could be intercepted and exploited to third-party advertisers and analytics services mining the content of users' data to target personal advertisements, set financial rates, and offer or withhold services. Commenting on the lack of regulations governing the mobile app ecosystem, the Privacy Rights Clearinghouse advocacy group has reported, "This is a particular concern with health and fitness apps, which often collect both demographic and health (or medical-like) information that does not fall under the protections of any health privacy laws. When you use the apps, you often create a record of data such as your diet, daily exercise, glucose readings, pregnancy, and/or menstrual cycle" (Njie, 2013: 4). Here we see the monetization of the rich contextual data that patient stories can provide when captured in open sources online and through health and wellness apps. Since we already know that Google and others in the data business seek to capture as much personal information about their users as possible, why would we assume that their intent for EHR mining is any different, or that their use of those data would be purely benevolent?

In contrast to the openness (and vulnerability) of patient communities online, the EHR — the central repository of patient information in American medicine — is almost universally seen as a closed system incapable of absorbing the unstructured data that patients share through their "metaclinical" social networks.<sup>2</sup> Comparative studies have shown how EHRs eliminate important contextual information about the patient's experience of illness by framing clinical narratives solely in terms of biomedical disease models (Patel, Arocha, & Kushniruk, 2002). Some narrative medicine skeptics have expressed concern that patient

stories in the form of unstructured data can become lengthy and unusable by clinicians who often need quick, succinct summaries to make treatment decisions (Morris, 2008). While artificial intelligence (AI) and machine learning are seen as potential solutions to this problem, even proponents acknowledge that few insights will emerge if the contextual data is never captured in the first place. As Deborah DiSanzo of IBM Watson Health recently observed, “Seventy percent of our determinants in health are not in our clinical or genomic data. We analyzed 650 million data points from one activity. We could find, really, no good actionable insights [because] it’s very difficult to get good social determinants in health data” (Muio, 2017: n.p.). The need to supplement clinical data with social and behavioral nuance leads researchers and data brokers alike back to scraping users’ personal information from the social web, yet critical code researchers have shown that big data analytics tend to reproduce the biases of their programmers and the society from which they emerge (Crawford & Calo, 2016; Eubanks, 2017; Noble, 2018; O’Neil, 2016; Sweeney, 2013; Wachter-Boettcher, 2017). The likelihood that algorithms designed to mine big health data will harm those already suffering from structural racism, poverty, and health disparities suggests that greater protection of users’ data within contexts of social sharing is urgently needed.

Under current regulation of these sociotechnical systems, patient stories and health data are disconnected forms of knowledge and meaning that exist in separate realms and are shaped by the affordances of their host platforms: open Internet for patient stories, closed EHR for patient data. A recent study of patient reviews of doctors on the online rating platform Yelp argued that the user-driven character of the content on this open-access site produces comments about fundamentally different aspects of the patient experience as compared to hospital-generated patient satisfaction questionnaires and surveys (Merchant, Volpp, & Asch, 2016; Ranard et al., 2016). The authors note, “Online platforms are democratizing in ways that answering pre-assigned questions can never be — because giving voice to patients also means

giving them the opportunity to select the topics” (Merchant, Volpp, & Asch, 2016: 2484). Some researchers have noted the unfortunate role of privacy restrictions in perpetuating the separation of unstructured narratives (e.g., Yelp) and quantitative medical datasets (e.g., EHRs) (Dohan, Garrett, Rendle, Halley, & Abramson, 2016). Yet, when considered in light of widespread third-party data mining, the restrictions governing clinical data sharing must also be seen as protective, insofar as they successfully prevent data brokers from accessing patients’ sensitive personal health information. However, for patients to benefit fully from sharing their illness narratives online, the digital health ecosystem must move beyond a binary opposition of paternalistic protectionism versus open source exploitation.

### **Conclusion: Closed Networks and OpenNotes**

The risks of third-party exploitation of data shared in online patient communities are real. However, the necessary response is not to firewall open patient resources online, thereby undermining their benefits alongside their harms, but rather to open up clinical EHR networks to more diverse forms of input, with greater privacy protections. No good models exist to date, but the OpenNotes project is a promising initiative that allows patients to see and comment on the few spaces where physicians actually include unstructured text: their clinical notes. While the project is explicitly focused on giving patients access to their own data, not on narrative medicine, OpenNotes nonetheless emphasizes narrative as a site of exchange that can improve patient care and well-being (NYS Health Foundation, 2017). Indeed, participants in the pilot study saw striking improvements in trusting doctor-patient relationships, patients’ feelings of control over their care, patients’ greater knowledge and recall about their medical conditions, and greater “adherence” to medication regimens (Gerard, Fossa, Folcarelli, Walker, & Bell, 2017; Klein et al., 2016). Populations that reported a particular benefit from OpenNotes included people with serious co-morbid or chronic

conditions, and patients who were elderly, non-white, or had less than a high school education (Bell et al., 2017).

With over 6 million users between 2010 and 2017, the organization is developing a second iteration of the project, called “OurNotes,” in which patients, families, and clinicians collaboratively generate notes (Commonwealth Fund, 2014). In this version, patients can not only review but also contribute to the clinical notes in their EHR. This phase explicitly centers on the power of storytelling to improve health outcomes and quality of life through patient-generated health histories and goals of care. Importantly, this project aims to integrate some of the social and affective dimensions of illness into clinical care, where it might impact the medical treatments available to a patient, or the clinician’s understanding of the patient’s priorities. OurNotes acknowledges and works within the existing framework of biomedical care while extending the boundaries of that frame to include more patients’ and caregivers’ voices. Yet, by preserving the centrality of the EHR, the traditional clinical context of care, and a limited social network, this initiative fundamentally accepts the model of clinic-driven definitions of disease, and in that way, reproduces the limitations of doctor-centered structures of knowledge (Lee, Walker, Delbanco, & Elmore, 2016).

The great strength of patient stories in platforms outside traditional clinical settings is their rich detail, their unrestrained narration, the varied and unfiltered affective states they register, the wild heterogeneity of sources and types of information they include, and the multiplicity of voices they present. Most of these variables continue to be excluded from EHRs, even those running OpenNotes or OurNotes. Medical institutions may not be receptive to the voluminous, messy “big health data” that floats freely across online platforms, but other organizations (like Google, Microsoft, Amazon, and Facebook) are paying attention (Robbins, 2016; Wingfield, Thomas, & Abelson, 2018). Just like all of the other data about Internet users that third parties can access (shopping, social media, GPS, music and video streaming, and so on), patients’ personal accounts of their daily lives are vectored through analytics

engines by companies seeking to monetize their data. The resulting digital profiles invisibly shape the communications and transactions that surround us. By maintaining rigid boundaries between public-sphere ambitions for health and wellness and private-sphere practices of medicine and delivery of health care, current US policies allow private interests to freely exploit our data, potentially harming us socially while ignoring our deeper medical needs.

### Acknowledgements

I would like to thank the editorial team at *Catalyst* for their valuable insights, especially David Serlin, Mara Mills, and the editors of this special issue, Beza Merid and Tamara Kneese.

### Notes

<sup>1</sup> In sum, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) governs the use and disclosure of individuals' protected health information (PHI) by covered entities, namely, health plans, health care clearinghouses, and any health care provider who transmits health information in electronic form in connection with business transactions. PHI includes demographic data that relates to the individual's past, present or future physical or mental health or condition; the provision of health care to the individual; or the past, present, or future payment for the provision of health care to the individual (Office for Civil Rights, 2013).

<sup>2</sup> Elsewhere I define "metaclinical" spaces as "those sites constituting the vast ecosystem outside of traditional clinical settings where consumer-patients engage in behaviors that may be directly or indirectly related to self-management of health and disease, whose digital traces can be captured and incorporated into data-driven frameworks for health surveillance and intervention" (Ostherr, 2018).

## References

Agus, D.B. (2016, February 6). Give up your data to cure disease. *The New York Times*. <https://www.nytimes.com/2016/02/07/opinion/sunday/give-up-your-data-to-cure-disease.html>

Allison, J. (2016, December 30). [User comment, PatientsLikeMe blog, "The Value of Openness."] <http://blog.patientslikeme.com/2007/12/13/the-value-of-openness/>

Apple for Healthcare. (2018). Empower your patients with Health Records on iPhone. <https://www.apple.com/healthcare/health-records/>

Apple Newsroom. (2018, January 24). Apple announces effortless solution bringing health records to iPhone. <https://www.apple.com/newsroom/2018/01/apple-announces-effortless-solution-bringing-health-records-to-iphone/>

Axon, S. (2018, February 26). Your Apple iCloud data is now stored on Google servers — surprised? *Ars Technica* <https://arstechnica.com/gadgets/2018/02/your-apple-icloud-data-is-now-stored-on-google-servers-surprised/>

Bell, S.K., Mejilla, R., Anselmo, M., Darer, J.D., Elmore, J.G., Leveille, S., ... Walker, J. (2017). When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. *BMJ Quality & Safety*, 26(4), 262-270. <http://dx.doi.org/10.1136/bmjqs-2015-004697>

Campos, H. & Sebastian, A. (2017). Data liberación: the Path to Precision Medicine. <https://obamawhitehouse.archives.gov/blog/2015/07/09/data-liberacion-path-precision-medicine>

Charon, R. (2006). *Narrative medicine: Honoring the stories of illness*. New York, NY: Oxford University Press.

Charon, R., DasGupta, S., Hermann, N., Irvine, C., Marcus, E.R., Colon, E.R.,... Spiegel, M. (2017). *The principles and practice of narrative medicine*. New York, NY: Oxford University Press.

Charon, R. & Wyer, P. (2008, January 26). Narrative evidence based medicine. *Lancet*, 371 (9609), 296-7.

Clarke, A.E., Shim, J.K., Mamo, L., Fosket, J.R., & Fishman, J.R. (2010). "Biomedicalization: Technoscientific transformations of health, illness, and

U.S. biomedicine.” In A.E. Clarke, L. Mamo, J.R. Fosket, J.R. Fishman, & J.K. Shim (Eds.), *Biomedicalization: Technoscience, health, and illness in the U.S.* (pp. 47-87). Durham, NC: Duke University Press.

Commonwealth Fund, The. (2014). OurNotes: A New Strategy to Improve Care for Complex Patients. <http://www.commonwealthfund.org/grants-and-fellowships/grants/2014/nov/ournotes-new-strategy-to-improve-care-for-complex-patients>

Crawford, K. & Calo, R. (2016, October 20). There is a blind spot in AI research. *Nature*, 538, 311–313. doi:10.1038/538311a

Daston, L.J. & Galison, P. (2007). *Objectivity*. Cambridge, MA: MIT Press.

deBronkhardt, D. (2013). How the e-patient community helped save my life. *BMJ* 346. doi: <https://doi.org/10.1136/bmj.f1990>

deBronkhardt, D. (2009-). e-Patient Dave: Toward a New Science of Patient Engagement. [Blog post] <http://www.epatientdave.com/>

Dohan, D., Garrett, S.B., Rendle, K.A., Halley, M., Abramson, C. (2016). The importance of integrating narrative into health care decision making. *Health Affairs* (Millwood), 35(4), 720-5. doi:10.1377/hlthaff.2015.1373

Dumit, J. (2006). Illnesses you have to fight to get: facts as forces in uncertain, emergent illnesses. *Social Science and Medicine*, 62(3), 577-590.

Eichler, G.S., Cochin, E., Han, J., Hu, S., Vaughan, T.E., Wicks, P.,... Devenport, J. (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. doi: 10.2196/jmir.5130

Epstein, S. (1996). *Impure science: AIDS, activism, and the politics of knowledge*. Berkeley and Los Angeles, CA: University of California Press.

Eubanks, V. (2017). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin's Press.

Ferguson, T., & the e-Patient Scholars Working Group. (2007). E-patients: How they can help us heal healthcare. <http://e-patients.net/>

Ferguson, T. & Frydman, G. (2004). The first generation of e-patients. *BMJ* 328, 1148-1149. doi: 10.1136/bmj.328.7449.1148

Galison, P. (1997). *Image and logic: A material culture of microphysics*. Chicago, IL: University of Chicago Press.

Gerard, M., Fossa A., Folcarelli, P.H., Walker, J. & Bell, S.K. (2017). What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. *Journal of Medical Internet Research*, 19(7): e237. doi:10.2196/jmir.7212

Han, J. & Wiley, J. (2013). Digital illness narratives: A new form of health communication. *Transactions of the International Conference on Health Information Technology Advancement*, 2(1). Paper 18.

Hand, E. (2010). People power: Networks of human minds are taking citizen science to a new level. *Nature*, 466 (7307), 685-687.  
<https://www.nature.com/news/2010/100804/full/466685a.html>

Holliday, R. (2009-) Regina Holliday's Medical Advocacy Blog.  
<http://reginaholliday.blogspot.com/>

Hsiao, C.-J. & Hing, E. (2012). Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2012. *NCHS data brief*, 111. Hyattsville, MD: National Center for Health Statistics.

Jenkins, H. (2009). What Happened Before YouTube. In J. Burgess & J. Green (Eds.), *YouTube: Online video and participatory culture* (pp. 109-125). Cambridge, UK: Polity Press.

Johnson, C.Y. (2018, January 24). Apple wants you to put your medical records on the iPhone. *The Washington Post*.  
[https://www.washingtonpost.com/news/the-switch/wp/2018/01/24/apple-wants-you-to-put-your-medical-records-on-the-iphone/?utm\\_term=.7d6059e5e594](https://www.washingtonpost.com/news/the-switch/wp/2018/01/24/apple-wants-you-to-put-your-medical-records-on-the-iphone/?utm_term=.7d6059e5e594)

Klein, J.W., Jackson, S.L., Bell, S.K., Anselmo, M.K., Walker, J., Delbanco, T., & Elmore, J.G. (2016). Your patient is now reading your note: Opportunities, problems, and prospects. *The American Journal of Medicine*, 129(10), 1018-1021.

Lee, B.S., Walker, J., Delbanco, T., & Elmore, J.G. (2016). Transparent electronic health records and lagging laws. *Annals of Internal Medicine*, 165(3): 219-220.

Marcus, A.D. & Weaver, C. (2012, November 28) Heart gadgets test privacy-law limits. *Wall Street Journal*.

<https://www.wsj.com/articles/SB10001424052970203937004578078820874744076>

Mayer-Schonberger V, Cukier K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.

Merchant, R.M., Volpp, K.G., & Asch, D.A. (2016). Learning by listening: Improving health care in the era of Yelp. *JAMA*, 316 (23), 2483-2484.

Morris, D.B. (2008). Narrative medicines: Challenge and resistance. *Permanente Journal*, 12(1), 88-96.

Muoio, D. (2017 December 5). IBM Watson manager, academics describe challenges, potential of healthcare AI. *mobihealthnews*.  
<http://www.mobihealthnews.com/content/ibm-watson-manager-academics-describe-challenges-potential-healthcare-ai>

Nafus, D. & 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, 1784-1794.

Njie C. (2013). Mobile health and fitness apps: What are the privacy risks? *Privacy Rights Clearinghouse*.<https://www.privacyrights.org/consumer-guides/mobile-health-and-fitness-apps-what-are-privacy-risks>

Noble, S.U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: New York University Press.

Novet, J. (2018, February 26). Apple confirms it uses Google's cloud for iCloud. *CNBC* online. <https://www.cnn.com/2018/02/26/apple-confirms-it-uses-google-cloud-for-icloud.html>

NYS Health Foundation. (2017, October 12). Full List of Selected Recipients for "Spreading OpenNotes in New York State" RFP. OpenNotes website. <https://www.opennotes.org/news/full-list-of-selected-recipients-for-spreading-opennotes-in-new-york-state-rfp/>

Office for Civil Rights, U.S. Department of Health & Human Services. (2013, July 26). Summary of the HIPAA Privacy Rule.

<https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>.

Okun, S., & Goodwin, K. (2017). Building a learning health community: By the people, for the people. *Learning Health Systems*, 1(3), e10028.  
<https://doi.org/10.1002/lrh2.10028>

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.

Osler W. (1901). The natural method of teaching the subject of medicine. *JAMA*, 36, 1673-9.

Ostherr, K. (forthcoming 2018). Risk media in medicine: The rise of the metaclinical health app ecosystem. In B. Ghosh & B. Sarkar (Eds.), *The Routledge companion to media and risk*. New York, NY and London, UK: Routledge.

Ostherr, K., Borodina, S., Bracken, R.C., Lotterman, C., Storer, E., & Williams, B. (2017). Trust and privacy in the context of user-generated health data. *Big Data & Society*, 4(1). <https://doi.org/10.1177/2053951717704673>.

Patel, V.L., Arocha, J.F. & Kushniruk, A.W. (2002). Patients' and physicians' understanding of health and biomedical concepts: Relationship to the design of EMR systems. *Journal of Biomedical Informatics*, 35(1), 8-16.

PatientsLikeMe. (n.d.). About PatientsLikeMe.  
<http://news.patientslikeme.com/about/background>

PatientsLikeMe. (n.d.). Partners.  
<https://www.patientslikeme.com/about/partners>

PatientsLikeMe. (n.d.). Research: Our Science.  
<https://www.patientslikeme.com/research/science>

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

Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system*. Cambridge, MA: MIT Press.

Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Liu, P.J., ... Dean, J. (2018, January 26). Scalable and accurate deep learning for electronic health records. *arXiv:1801.07860v2 [cs.CY]*

Ranard, B.L., Werner, R.M., Antanavicius, T., Schwartz, H.A., Smith, R.J., Meisel, Z.F., ... Merchant, R.M. (2016). Yelp reviews of hospital care can

supplement and inform traditional surveys of the patient experience of care. *Health Affairs* (Millwood), 35(4), 697-705.

Robbins, R. (2016, November 1). Get ready for more drug ads: Facebook is making a bid for pharma dollars. *STAT*.  
<https://www.statnews.com/2016/11/01/facebook-pharma-drug-ads/>

Ruckenstein, M. & Schüll, N.D. (2017). The datafication of health. *Annual Review of Anthropology*, 46, 261-78. <https://doi.org/10.1146/annurev-anthro-102116-041244>

Salmi, L. (2017, April 7). Tomorrow's patients will demand greater transparency and openness. *The BMJ: Opinion*.  
<http://blogs.bmj.com/bmj/2017/04/07/liz-salmi-tomorrows-patients-will-demand-greater-transparency-and-openness/>

Shanafelt, T.D., Dyrbye, L.N., Sinsky, C., Hasan, O., Satele, D., Sloan, J., & West, C.P. (2016, July). Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clinic Proceedings*, 91(7), 836-848.

Singer, E. (2011, November 22). Getting health data from inside your body. *MIT Technology Review*.  
<https://www.technologyreview.com/s/426171/getting-health-data-from-inside-your-body/>

Standen, A. (2012, May 28). Patients crusade for access to their medical device data. KQED/NPR online. <https://www.npr.org/sections/health-shots/2012/05/28/153706099/patients-crusade-for-access-to-their-medical-device-data>

Swaminath, A. (2016, August 29). Stop asking Dr. Google for advice. *New York Times*. <http://www.nytimes.com/roomfordebate/2016/08/29/are-medical-websites-like-webmd-healthy/stop-asking-dr-google-for-advice>.

Sweeney, L. (2013, May). Discrimination in Online Ad Delivery. *Communications of the ACM*, 56(5), 44-54.  
<https://dataprivacylab.org/projects/onlineads/>

Thorp-Lancaster, D. (2016, Aug 12). Fitbit shutting down Microsoft HealthVault integration on September 12 [Update]. Windows Central blog.  
<https://www.windowscentral.com/fitbit-shutting-down-microsoft-healthvault-integration-september-12>

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. New Haven, CT: Yale University Press.

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

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

Wachter-Boettcher, S. (2017). *Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech*. New York, NY: W.W. Norton.

White House Archives, President Barack Obama. (2015). Champions of Change: Precision Medicine: Hugo Campos.  
<https://obamawhitehouse.archives.gov/champions/precision-medicine/>

Wingfield, N., Thomas, K., & Abelson, R. (2018, January 30). Amazon, Berkshire Hathaway and JPMorgan team up to try to disrupt health care. *New York Times* <https://www.nytimes.com/2018/01/30/technology/amazon-berkshire-hathaway-jpmorgan-health-care.html>.

Zang, J., Dummit, K., Graves, J., Lisker, P., & Sweeney, L. (2015). Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps. *Technology Science*. 2015103001.

Zhou, L., Collins, S., Morgan, S.J., Zafar, N., Gesner, E.J., Fehrenbach, M. & Rocha, R.A. (2016). A decade of experience in creating and maintaining data elements for structured clinical documentation in EHRs. *AMIA Annual Symposium Proceedings 2016*, 1293-1302.

## Bio

**Kirsten Ostherr**, PhD, MPH is the Gladys Louise Fox Professor of English at Rice University, in Houston, Texas, where she is a media scholar, health researcher, and founding director of the Medical Humanities program. She is the author of *Medical Visions: Producing the Patient through Film, Television and Imaging Technologies* (Oxford, 2013) and *Cinematic Prophylaxis: Globalization and Contagion in the Discourse of World Health* (Duke, 2005). She is the editor of *Applied Media Studies* (Routledge, 2018), and co-editor of *Science/Animation*, a special issue of the journal *Discourse* (2016). Her current research is on information and communication technologies in medicine, patient narratives, trust and privacy in digital health ecosystems, and the role of simulation as a mediator between human and technological

forms of medical expertise. Her current book project is called *Quantified Health: Learning from Patient Stories in the Age of Big Data*.