Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Big data governance of personal health information and challenges to contextual integrity

Jenifer Sunrise Winter, School of Communications, University of Hawaii, Honolulu, Hawaii, USA

Elizabeth Davidson, Shidler College of Business, University of Hawaii, Honolulu, Hawaii, USA

CONTACT INFORMATION

Jenifer Sunrise Winter, School of Communications, University of Hawaii, Honolulu, 2550 Campus Road, Crawford Hall 325, Hawaii 96822, USA; EMAIL: jwinter@hawaii.edu

Citation

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Abstract

Pervasive digitization and aggregation of personal health information (PHI), along with artificial intelligence (AI) and other advanced analytical techniques, hold promise of improved health and healthcare services. These advances also pose significant data governance challenges for ensuring value for individual, organizational, and societal stakeholders as well as individual privacy and autonomy. Through a case study of a controversial public-private partnership between Royal Free Trust, a National Health Service hospital system in the United Kingdom, and Alphabet's AI venture DeepMind Health, we investigate how forms of data governance were adapted, as PHI data flowed into new use contexts, to address concerns of contextual integrity, which is violated when personal information collected in one use context moves to another use context with different norms of appropriateness.

Introduction

Ever-increasing digital stores of personal health information (PHI) hold the promise of improving efficiencies, efficacy, and precision of clinical medicine (Murdoch and Detsky 2013) via big data analytics and artificial intelligence (AI)/machine learning. On the other hand, there are growing concerns about protection of individual privacy and the security of the data, while also facilitating their use to enhance clinical research and societal welfare (e.g., Anderson and Agarwal 2011; Belanger and Xu 2015; Diamond, Mostashari, and Shirky 2009; Hripcsak et al. 2014; Rosenbaum 2010; Ross et al. 2014). Such complexities are compounded by differing positions of stakeholders (patients, healthcare providers, third-party insurers, marketers, and others) on issues such as data ownership, access, and privacy (Rosenbaum 2010).

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

In this article, we examine PHI data governance challenges that arise when the data are moved across different use contexts. Here data governance entails not only the institutional structures and processes for managing a data store or data domain, but also data stewardship – a fiduciary (trust) relationship between the steward and the individuals or entities whose data are involved (Rosenbaum 2010). Following Nissenbaum (2010), we consider forms of data governance with regard to their potential for ensuring contextual integrity. Through a case study of a controversial public-private partnership between Royal Free Trust, a National Health Service (NHS) hospital system in the United Kingdom, and Alphabet's AI venture DeepMind Health (DMH), we investigate how forms of data governance were adapted, as PHI data flowed into new use contexts, to address concerns of contextual integrity, which is violated when personal information collected in one use context moves to another use context with different norms of appropriateness. This case highlights the complexities in how PHI data might be governed to promote healthcare innovation while also protecting privacy and serving the public good.

The rest of the article proceeds as follows. The next section discusses the theoretical foundations of our analysis, focusing on Nissenbaum's (2010) notion of contextual integrity and our earlier study where we identified five analytic dimensions for understanding forms of PHI data governance (Winter and Davidson 2017). The subsequent sections discuss our research design and methods, findings, and analysis. The last section offers our concluding thoughts.

Theoretical foundations

We first consider how shifting use contexts of PHI may challenge their contextual integrity (Nissenbaum 2010). We then consider how forms for PHI data governance (Winter and Davidson 2017) change, as PHI data flow into new use contexts.

"Big data" and PHI privacy

As more transactional processes are digitized, and as everyday objects are redesigned to include digital sensors, computing power, and communication capabilities, the scope and volume of data generated offer opportunities to "mine" value from them to improve organizational performance (Davenport, Barth, and Bean 2012) and for entrepreneurial ventures (Lycett 2013). These trends are evident in the healthcare sector, which faces increasing costs and rising societal expectations for health services along with limited public and private funding to meet the demand (Deloitte 2016). Application of big data analytics tools and techniques to digitized health data holds the promise of improving efficiencies (Blumenthal 2010) and to "greatly expand the capacity to generate new knowledge" (Murdoch and Detsky 2013, 1351).

Security of some PHI[1] are subject to government regulation because of potential harms to individuals or groups, depending on how these data are (or could be) used, e.g. discriminate against individuals based on their health status. The growing concerns about the security of health data (e.g. prevention of unauthorized use) and management of data (e.g. ensuring that data

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

stores are accurate and complete) has fueled interest in data governance (Belanger and Xu 2015; Data Governance Institute n.d.; Elliott et al. 2013; Holmes et al. 2014; Hripcsak et al. 2014; Khatri and Brown 2010; Ladley 2012; Rosenbaum 2010; Zuboff 2015). These discourses range from a plethora of practical advice, such as how to make data governance work, to less frequent discussions of philosophical and societal implications of data stewardship.

Data aggregation and analysis create conflict between social values related to privacy, personal autonomy, and liberty, attributable in part to conflicts in various stakeholders' values and interests in how such data are used in socioeconomic activities (Nissenbaum 2010; Winter 2014). Beyond PHI data aggregation, algorithms are also used to predict future behavior and make judgments about individuals or groups, raising new concerns about privacy. As Mai (2016) has observed, with the widespread diffusion of big data analytics and machine learning techniques, our views of privacy must expand to consider the use of predictive analytics on existing data and the socioeconomic context in which they are used. For instance, in healthcare organizations, predictive analytics are being used to forecast individuals' health status changes for the purposes of delivering healthcare services, as well as to assess their likely consumption of healthcare resources (Bates et al. 2014; Wagner 2016). While these dual uses might be reconciled for societal and individual benefit, individual-level predictions could also be used to exclude high-cost individuals or populations from access to health services (Eyal 2013).

Contextual integrity and PHI data use and reuse

Privacy is often cited as a human right, but there are many conflicting views on what it entails (Acquisti, Brandimarte, and Loewenstein 2015; Solove 2010). Moreover, such conceptions of privacy are not able to address the radical changes and opportunities brought about by new technologies. Nissenbaum (2010) argues that privacy should not be conceived of as a right to secrecy or control, but as "appropriate flow of personal information" within particular social contexts (127). Each context of information generation, storage, and use is characterized by an array of actor roles, activities, values, and norms about the appropriateness of information sharing or use in that particular context. Contextual integrity is violated when personal information collected in one context moves to another use context with different norms of appropriateness.

Acknowledging that norms may be incomplete and complex, Nissenbaum focuses on "countervailing values" to highlight potential conflicts and tradeoffs when personal information is moved to a different use context. She acknowledges that the notion of contextual integrity is largely conservative of the status quo, that is, preserving the existing and entrenched norms in a context, but she allows for the reconsideration of norms based on higher social values, goals, and norms. Higher values include prevention of informational harms or informational inequality, and preservation of autonomy and freedom, human relationships, and democracy. Thus, the notion of contextual integrity addresses privacy via "appropriate information flows" and by comparing "entrenched normative practices against novel alternatives or competing practices on the basis of how effective each is in supporting, achieving, or promoting relevant contextual values" (Nissenbaum 2010, 166).

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

The notion of contextual integrity is valuable for examining the implications of large-scale personal health data collection, aggregation, use and reuse enabled by ICT (information and communications technologies), which facilitate the flow of personal information from one use context to another with different norms, often embedded in ICT systems and information practices. As data digitization, aggregation, and AI applications expand rapidly in healthcare, the potential for contextual integrity violations is substantial. For instance, norms for PHI flows in the patient-health care provider relationship clash with the norms for use of personal data in advertising. At the same time, PHI data flows such as transfer of individuals' clinical data into public health, medical research or health system policy contexts might be justified on the grounds of supporting higher social value. Nonetheless, any change in use context of PHI data is a potential violation of contextual integrity, and thus merit consideration by policy makers and health system stakeholders.

Forms of PHI data governance

Despite many calls for governance of PHI so as to protect individual privacy while also making it available for health system research and innovation, effective governance remains an elusive goal (British Academy and the Royal Society 2017; Elliott et al. 2013; Hripcsak et al. 2014; OECD, 2017; Rosenbaum 2010; Ross et al. 2014). Laws and regulations for patients' privacy limit the degree to which some forms of PHI can be shared or sold. Other PHI, such as data generated by individuals using consumer electronics, is largely unregulated (U.S. Department of Health and Human Services 2018). Beyond protecting individual privacy, the proprietary claims that organizations stake on health data generated or maintained by their own ICT can limit data sharing for societal purposes such as research or system improvements at inter-organizational level (Rosenbaum 2010). The complexity and limited standardization of health data, along with a plethora of health ICT systems lacking interoperability, also contribute to technical hurdles (Eden et al. 2016).

A variety of governance forms are emerging to address the challenges and opportunities of health data governance. They are identifiable by prototypical arrangements of core properties including goals, authority relations, technologies, and markets served (Scott 2001). In a study of PHI data governance (Winter and Davidson 2017), we identified five analytic dimensions that characterize various PHI governance forms and developed a preliminary taxonomy of forms arising from different arrangements of these dimensions. Table 1 summarizes the analytical dimensions.

[Insert Table 1 about here]

Data domains (dimension 1) are areas and purposes for which data are collected. PHI data domains may include individuals' medical histories, clinical data stored in electronic health record systems (EHRs) at hospitals, laboratories, and private medical practices; data related to prescriptions; data from medical devices (e.g. glucose monitors, personal health trackers like Fitbit); medical claims data; and genomic data in medical or commercial databanks. Additional PHI data domains can be constructed when data are analyzed to infer one's health status or

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

behaviors, even if those data are not classified as medical (e.g. retail purchases, Internet searches for information on health conditions) (Bates et al. 2014; Libert 2015; Wagner 2016).

Stakeholders (dimension 2) are individuals, organizations, and groups with an interest in the (potential) value of PHI data. They include patients, doctors, hospitals, pharmacies, third-party insurers, government agencies, data aggregators, and technology firms (Arndt 2018; Davidson, Østerlund, and Flaherty 2015). Stakeholders hold different values and interests with regard to PHI data and have different norms and practices for their use. Some values are widely shared, such as ensuring safety and quality in health services delivery, but there are also conflicting interests (Rosenbaum 2010).

The *value afforded by PHI data* (dimension 3) refers to how PHI data can create stakeholder value. PHI data can be applied to more precisely deliver healthcare services to individuals, to improve overall health system efficiency, to support an organization's competitive strategies, or to generate revenue through resale, to name just a few. Barrett, Oborn, and Orlikowski (2016) identified six different types of value that different stakeholders gained from interacting on an online health community platform – financial, epistemic, ethical, service, reputational, and platform. However, Tempini (2017) notes data reuse to enhance one stakeholder group's value may make data less valuable to others, because the processes of data generation, use, and reuse, and infrastructure development activities are not easily separated. Moreover, shifting data from one context to another may lessen value for some stakeholders. For instance, removing personally identifiable elements from a PHI data flow, before transferring data for secondary uses, may preserve individual privacy, but loss of individual-level information make PHI data less valuable for researchers looking for disease patterns, policy makers looking for health system inefficiencies, or marketers seeking to promote a product or service.

Governance goals (dimension 4) are values-based objectives for governing a data domain. Commonly cited governance goals are protecting privacy, ensuring data security from unauthorized access, facilitating ease of data access to legitimate users, and ensuring public trust in governance (British Academy and the Royal Society 2017). Other possible goals include giving individuals and groups voice (e.g. through personal health record systems), supporting innovation, and protecting intellectual property. Some governance goals may be broadly acknowledged or even shared within and across use contexts, but there could be conflict over others.

Finally, *governance forms* (dimension 5) refer to organizational units, practices, policies and regulations, and technologies that carry out governance goals (Rosenbaum 2010). For instance, privacy laws and regulations are governance forms that specify how identifiable PHI must be managed and the entities that are responsible for doing so (e.g. health service providers, pharmacies, insurance companies) in order to attain privacy goals. Algorithms for de-identification of PHI data are governance forms intended to address privacy goals and also to also make PHI more accessible for exchange or sale by reducing regulatory oversight (Hripcsak et al. 2014). A data access committee is a governance form that may set policies and evaluate requests for PHI data use within or between organizations. Understanding the relationships

among stakeholders, PHI data value, and governance goals is important to evaluate the effectiveness of data governance forms overall and their efficacy in managing data flows across contexts so as to maintain contextual integrity.

Two examples illustrate how these dimensions characterize various forms of PHI data governance[2]. The most common form is organizational-level governance, in which a hospital or other clinical entity serves as the steward for and primary user of PHI data generated on its own ICT. The hospital maintains individual-level clinical data along with operational data such as records of clinician-provided services and financial reimbursements (data domains). These PHI data are "owned" by the organization, though government regulations may require data to be shared with accreditation agencies, researchers or patients (governance goals and forms). The organization is responsible for protecting patient privacy (as regulated) and data security, while also making data easily accessible to clinicians as they carry out their duties and analysts to assess the efficiency and quality of service delivery (governance goals, data value), generally through ICT such as electronic health record systems or data repositories (governance forms). Data may also be used to stratify patients according to disease states or hospital resource consumption in order to design intervention strategies (data value). Thus, individual patients and their care providers as well as the hospital, its funders, and the public served by the hospital are all stakeholders within the domain of this governance form.

A much different governance form is individual-level PHI data governance. Here, individuals generate data by using consumer electronics such as wearable activity trackers and glucose monitors (Cortez et al. 2018; Montgomery, Chester, and Kopp 2018) or by entering information about their health-related activities into mobile apps (data domains). Individuals share data governance rights and responsibilities with the IT firms providing the devices or the data aggregation services (stakeholders), and data typically reside on the IT vendor's cloud-based infrastructure and in the individual's mobile devices such as a smart phone (governance form). In the U.S., these data are not covered by federal PHI regulations but instead are governed by the privacy policies of the IT firm (U.S. Department of Health and Human Services 2016) (governance form). For the individual, these data hold value as sources of health monitoring and improvement, whereas to the IT firms, these detailed personal data represent potential profit, by selling advertising access to consumers, selling PHI data, or using data resources to develop advanced data analytics capabilities (data value) (Sankin 2017).

These and other governance forms develop around ICT-generated health data and are adjusted when data shift from one context into another, particularly where informational practices, norms, and stakeholder values differ from those in the original setting, e.g. firms seeking to merge consumer-generated health data and protected clinical health data onto their own technology platforms (Agency for Healthcare Research and Quality 2017, Arndt 2018; Deering 2013). Governance conflicts are possible among traditional health system stakeholders as well, for instance, between independent policy researchers, focused on reducing health systems costs through big data analytics, and clinical organizations, whose economic and competitive interests may not be served by PHI-enhanced policy research (Eden et al. 2016; Rodwin 2009; Tanner 2016).

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

With this theoretical grounding, we now return to our research questions of how data governance forms are adapted as PHI flows into new use contexts and the implications for PHI contextual integrity of these adaptations. Our five analytical dimensions provide insight into whether or how adaptations in governance forms occur and how violations of contextual integrity are addressed through new or altered forms.

Research design and methods

We analyze, applying our five analytic dimensions, the controversy surrounding a public-private data sharing arrangement in the UK – the DeepMind Health (DMH)-Royal Free Trust agreement.

DMH is a technology firm dedicated to developing AI applications for healthcare and is owned by Google's parent company, Alphabet. The company states that it operates independently from Google and has the goal of "proving that [AI technologies] could have positive social impact" (DeepMind 2017a). Royal Free Trust is a hospital system operated by the National Health Service (NHS) of the UK government.

In the first stage of data analysis, we gathered and read an array of sources related to the DMH-Royal Free Trust agreement, including materials produced by DMH, Royal Free Trust/NHS, and the United Kingdom's Information Commissioner's Office (ICO). They included data sharing agreements, policy reports, news stories, and public interviews with stakeholders. In each case, the related web domain (DMH, NHS and Royal Free Trust, ICO) was searched for original documents, including formal statements, blogs, contracts/service agreements, FAQs (frequently asked questions), and related reports. We also collected secondary materials, largely from news outlets covering the case, for analysis. They included additional comments from stakeholders weighing in on the case. From this analysis, we created a detailed case narrative with important dates, actions, and actors, as well as a timeline of key events (Table 2) to represent these activities.

[Insert Table 2 about here]

In the second phase, we applied our five analytical dimensions (Table 1) to identify tensions and conflicts arising from the migration of PHI data from one context (data governance form at Royal Free Trust) to a novel context (data governance form at DMH). To highlight the implications of these moves, we first outlined the existing practices at the Royal Free Trust (prior to DMH-Royal Free Trust agreement). We then examined new informational practices arising from the agreement and identified where existing informational norms and values conflicted with the new ones emerging from changing sociotechnical arrangements for data governance as a result of the agreement. We discuss our findings and analysis in the following section.

Findings and analysis

Initiating PHI data sharing between Royal Free Trust and DeepMind Health

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Prior to Royal Free Trust's partnership with DMH, its PHI data were governed primarily within the context of healthcare services provided it provided to its patients — data collected and used in the actual delivery of services (e.g. orders for laboratory tests) and in administrative work (e.g. billing). NHS data sharing policies specified how patient data might be shared with health researchers.

The data domain (clinical health data) included personally identifiable information on specific individuals treated at the Royal Free Trust hospitals. It included a variety of information stored in electronic health records such as diagnoses, laboratory test results, and medical imaging. The key actors involved in handling of PHI included various employees directly engaged in patient care (e.g. nurses, technicians, physicians, clerical staff), clinical and public health researchers, and administrators. Other stakeholders included government policymakers (who analyzed the data in aggregate form), individual patients who sought health care at Royal Free Trust, and citizens of the UK, who collectively fund and seek services from the Royal Free Trust and also other parts of the NHS.

The value of the PHI from all stakeholders' perspectives included improving individual treatment and health outcomes. Additionally, Royal Free Trust, NHS, and government policymakers saw value in the data to improve operational efficiency, reduce expenditures, and to improve quality of care through data analytics. Key goals of governance included maintaining the privacy and security of PHI, efficient organizational access to PHI for delivery of care to patients, and analysis of PHI for innovation, as well as for evaluation of organizational operations and improvements. Data flows were governed by existing laws and regulations (e.g. Data Protection Act) and institutional policies (e.g. Royal Free Trust's privacy statement).

On September 29, 2015 Royal Free Trust signed a contract ("Information sharing agreement" 2015) with DMH and thereafter transferred the PHI of 1.6 million patients to it. This contract enabled DMH to access five years' worth of patient data ("Information sharing agreement" 2015; Hawkes 2016) and "a wide range of healthcare data … This will include information about people who are HIV-positive, for instance, as well as details of drug overdoses and abortions" (Hodson 2016, 3). These data were used to test DMH's first application, Streams (DeepMind 2017a). The Streams application provides intelligent alerts, clinical notes, and task management for patients with acute kidney injury to physicians and nurses via iPads and iPhones. DMH reports that data are transmitted with end-to-end encryption from Royal Free Trust to an NHS-approved data center in the UK. Once at the DMH site, patient data from a range of hospital IT systems are mined using DMH's proprietary algorithms in real-time. DMH reviews the data for health issues, and if problems are found, DMH "sends an urgent secure smartphone alert to the right clinician to help, along with information about previous conditions so they can make an immediate diagnosis" (DeepMind 2017e). This, and similar applications developed with DMH, have the potential to improve healthcare delivery and services to patients while also measuring and monitoring organizational performance to reduce undesirable variability in health delivery quality (DeepMind 2017e; Independent Review Panel 2017).

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Emergence of a new PHI organizational governance form

In this case, PHI data were migrated from a context with a well-established and legitimated data governance form, in which the health services organization creates and manages its own PHI data (i.e., Royal Free Trust acting as data steward) to a novel form of PHI data governance, in which a global IT firm (Alphabet through its subsidiary DMH) assumed stewardship responsibilities and rights over this same data. Royal Free Trust and DMH have argued that DMH acted only as an agent of the former (i.e., an outsourced data analysis organization), working within its data governance framework. However, our analysis indicates that the data sharing agreement engendered a new and distinct data governance form and therefore PHI data migration from Royal Free Trust to DMH violated the contextual integrity of the data and the adjustments as yet have only partially mitigated these violations.

Expanded data domains

While data sharing arrangement with DMH provided access to five years' worth of patient data ("Information sharing agreement" 2015), most of this data was not directly relevant to the Streams application developed by DMH (Powles and Hodson 2017), as it included records of many people not directly benefiting from Streams. Moreover, the data analytic techniques used in Streams were not clearly explained. Beyond the use of PHI for the Streams application, how it would be utilized by DMH in development of other applications was not specified in the agreement. This raised concerns from patient advocacy groups, who questioned why DMH needs so much data and noted that data could be used for most anything (Hawkes 2016). Moreover, predictive analytics applied to PHI could generate new types of data about individuals by categorizing them according to health conditions or behavioral risks.

DMH stated that PHI would not be combined with other data sources (such as Google's Internet search data or retail data), but the firm's intention to develop AI applications leaves open the possibility that the PHI could help it expand to other non-health-related domains. Moreover, the long-term implications of sharing PHI data with a subsidiary of Alphabet remain a concern. An independent panel of experts set up to advise DeepMind recommended in 2018 that DeepMind should clarify its business model and relationship with Alphabet, as the AI subsidiary "would eventually have to prove its value either by sharing algorithms and data, or by making money" (Ram and Waters 2018, 3).

New stakeholders

Although there had always been multiple stakeholders (e.g. clinicians, laboratory technicians, administrators, researchers) accessing the Royal Free Trust patient data, the DMH partnership introduced diverse actors and goals. DMH describes itself simply as a "data processor … a third party that must only process this data in strict accordance with the instructions of the data controller and the law" (DeepMind 2017c). However, stakeholders with interests in this data and its governance now include DMH executives, data scientists, and

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

employees and also Alphabet corporate management. The addition of these corporate stakeholders brings new values and interests to stewardship of the Royal Free Trust patients' PHI, as patient data becomes subject to analytics beyond their original healthcare context and that too by for-profit corporations. For instance, while DMH developed an application — Streams — for Royal Free Trust and its constituents, DMH's AI researchers may be primarily focused on improving algorithms in general, with a broader goal of exporting the DMH's AI work (if not specifically the data) to other contexts.

Transfer of PHI data to a subsidiary of Alphabet, which monitors Internet activity of billions of individuals, provoked a strong opposition from public and policy makers to the data sharing arrangement (Independent Review Panel 2017). Despite assurances that Royal Free Trusts patients' PHI would not be combined with other data, questions about who might benefit from data-sharing partnerships between corporations and trusted public institutions contributed to a "trust gap" (Sharon 2016).

Value in the PHI data

The migration of PHI to new contexts for new value propositions creates conflicts between individual and corporate interests, highlighting the "unresolved tension that emerges when altruistic modes of behavior and financial profit-seeking overlap; and this in ways that are often not transparent" (Sharon 2016, 568). Originally PHI was collected to improve health outcomes of individuals and the effectiveness of the overall healthcare system but with the introduction of DMH as a stakeholder the value of PHI took on new hue: opportunities for development of profitable applications, services, and algorithms.

These values are not necessarily contradictory, but data use and reuse is not seamless (Tempini 2017). Dr. Julian Huppert, head of an Independent Review Panel instituted by DMH to provide independent oversight and ensure public accountability for its health endeavors, noted that the NHS PHI data were not as optimized for reuse as DMH had expected: "DeepMind could use AI to help with healthcare, but I think that it found that the state of data in the NHS was not as good as it had hoped so it had to step back from this" (Wakefield 2017, 12). Further, he noted that, "a huge amount of work is needed to make the NHS more digital. We would get much more value from NHS data if it was in a secure, centrally managed system" rather than distributed across hundreds of databases that "don't talk to each other" (Saran 2017, 7-8). While few would argue against increased digitization of health data through interoperable information systems, designing such systems to optimize their utility for research versus for use by patients or clinicians could reduce the value of PHI to them and also increase workload (Tempini 2017).

Whether the initial machine learning projects have been successful or not, DMH clearly intended to deploy such analytical techniques to PHI obtained from Royal Free Trust sources. Although DMH claimed that AI was not used to develop the Streams application, the revised 2016 services agreement between DMH and Royal Free Trust outlined a list of collaborative goals and noted that DMH is a technology company specializing in "understanding and developing software and intelligent agents through the combination of cutting edge techniques

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

from machine learning and systems neuroscience – artificial intelligence ('AI'), in order to build powerful general-purpose learning algorithms" (Royal Free London NHS Foundation Trust 2016, 1), and also that DMH's goals is to gain "data for machine learning research under appropriate regulatory and ethical approvals" (Royal Free London NHS Foundation Trust 2016, 4; "Services agreement" 2016). Moreover, DMH would retain intellectual property rights to such techniques, which could be used beyond their original healthcare context.

Tensions over the value in the data emerged even within the NHS itself as new technologies enable new possibilities to extract value from PHI. Martin Severs, medical director at NHS Digital, division of NHS responsible for "transforming health and care through technology," thus conceptualized the value of the PHI:

> All the data the NHS holds is funded by the British taxpayer. Any use of that data should generate benefits back to the taxpayer. While we should open up enough data as possible for specific research and use cases, within those data-sharing agreements there should be a return on investment on that data. There is billions of pounds' worth of value in this data. We need to encourage innovation and allow failure at low costs, but there needs to be a return on investment of that data back into the NHS. (quoted in Saran 2017, 14)

Alternately, others in the NHS have argued that PHI should only be used for AI when medical innovations provide societal benefit and data are "de-identified and constrained by a legal contract to balance the benefits of data and the risk to privacy" (Saran 2017, 11).

New data governance goals

The shift from narrow use of PHI by Royal Free Trust clinicians (governed by regulations such as the Data Protection Act) to the broad use by DMH in search of opportunities for development of AI applications gave rise to conflicts in data governance goals.

In fact DMH-Royal Free Trust partnership may have tried to sidestep individual patient's consent for PHI use (Powles and Hodson 2017). In July 2017, ICO ruled that the DMH-Royal Free Trust trial had not complied with data protection laws, noting in particular that patients were not adequately informed that their data would be used as part of the test (Information Commissioner's Office 2017):

> My investigation has determined that under the terms of the agreement with the Royal Free, DeepMind processed approximately 1.6 million partial patient records for the purpose of clinical safety testing without those patients being informed of this processing. I was not satisfied that the Royal Free had properly evidenced a condition for processing that would otherwise remove the need to obtain the informed consent of the patients involved and our concerns in this regard remain (Information Commissioner's Office 2017, 4).

DMH emphasized its role as a "data handler" though a secure data sharing arrangement, and Royal Free Trust noted that patient consent was assumed because PHI data use was for direct

patient care (Hawkes 2016). However, this interpretation limited the patients' voice in deciding how their personal health data will be used. Additionally, despite DMH's assurance that the contract was "no different" from other data-sharing agreements with the NHS, "this first contract had differed from the standard ones the NHS signed with third-parties" (Wakefield 2017, 5)[3]. Elizabeth Dunham, the UK's Information Commissioner, said:

> There's no doubt the huge potential that creative use of data could have on patient care and clinical improvements, but the price of innovation does not need to be the erosion of fundamental privacy rights. Our investigation found a number of shortcomings in the way patient records were shared for this trial. Patients would not have reasonably expected their information to have been used in this way, and the Trust could and should have been far more transparent with patients as to what was happening. We've asked the Trust to commit to making changes that will address those shortcomings, and their co-operation is welcome. The Data Protection Act is not a barrier to innovation, but it does need to be considered wherever people's data is being used. (5-7)

New data governance forms

As stakeholders' values and interests with regards to PHI data diverged, new governance *forms* developed. Broadly, these governance structures fall into following three categories:

1. *Organizational governance forms.*

New organizational governance forms such as the governmental review boards were established to provide ethical oversight of PHI data and to determine what access will be granted and to whom. In the case of DMH-Royal Free Trust partnership, ICO appointed an oversight committee for all uses of the PHI. However, panel members are not civil servants and thus are not subject to government (or NHS) oversight and accountability. DMH also sought to establish a new organizational governance form to oversee the use of the data in this public-private partnership, possibly in response to public concerns and anticipated regulatory intervention. In February 2016, as described by DMH, it assembled an Independent Review Panel comprised of:

> respected public figures to act in the public interest as unpaid Independent Reviewers of DeepMind Health. They meet four times a year to scrutinize our work with the NHS, and will publicly issue an annual statement outlining their findings after reviewing our data sharing agreements, our privacy and security measures, and our product roadmaps. (DeepMind 2017d, 2-3)

This panel first met on June 14, 2016 and issued its first annual report (Independent Review Panel 2017) on July 1, 2017.[4] In October 2017, DeepMind (parent company of DMH) created a new ethics unit, the DeepMind Ethics & Society Fellows, to conduct research on privacy, transparency, fairness, and other such issues (DeepMind 2017b). Lomas (2017, 5) called this a

"gigantic conflict of interest" because here we have "a commercial AI giant researching the ethics of its own technology's societal impacts" (Lomas 2017, 5).

New organizational governance forms also included the revised data sharing agreement in November 2016 and DMH's creation of a series of patient engagement workshops, intended to increase transparency and patient buy-in for new developments related to health data and AI (DeepMind Health 2017a). Further, all parties have agreed that Alphabet or other corporations will not have access to the PHI data.

2. Regulatory governance forms.

This case also highlighted the limitations of the existing NHS policies on PHI data. On April 29, 2016, the *New Scientist* published the original "Information sharing agreement" (2015) that its correspondent had obtained through a Freedom of Information (FOI) request (Hodson 2016). In May 2016, ICO opened an investigation into the data sharing arrangement. At that time, it noted at least one (unspecified) complaint from a member of the public. During this investigation, which concluded on July 3, 2017, DMH and the Royal Free Trust revised their data sharing agreement ("Services agreement …" 2016). However, the ICO ruled that DMH and the Royal Free Trust had failed to comply with the Data Protection Act.

A specific regulatory concern was that patients were not adequately informed that their data would be used as part of the DMH project (Information Commissioner's Office 2017). Identifiable PHI data on *all* patients seen at Royal Free Trust were migrated without consent of or notification to individual citizens (Powles and Hodson 2017). Patients were not given a meaningful path to opt out or to engage in critical debate about the use of their data. This lack of disclosure limited a key regulatory provision – the right of individuals to opt out of data reuse – that supports governance goals of transparency and patient voice. Policies and procedures adopted by Royal Free Trust and DMH did not trigger debate among stakeholders until after issues were raised publicly (Independent Review Panel 2017; Powles and Hodson 2017). Although Royal Free Trust later indicated that patients could opt out of this data sharing via the Trust's website, whether this constitutes a meaningful option for consent or withdraw consent is questionable, as patients were not were not individually notified of this option.[5]

Just days before ICO's ruling (July 1, 2017), the Independent Review Panel had concluded its first annual report and noted concern about the regulatory compliance and recommended that DMH "should respond positively to any recommendations that result from the ICO investigation" and "set as a firm policy that all future contracts with the public sector should also be published openly, with minimal or no redactions" (Independent Review Panel 2017, 11). After the ICO decision, Royal Free Trust was directed to halt all future trials until an adequate legal basis under the Data Protection Act (or GDPR) was in place. Further, it was required to explain in advance how it will "comply with its duty of confidence to patients in any future trial involving personal data" (Information Commissioner's Office 2017, 11), create a privacy impact assessment that ensures transparency, and share audits of future trials with ICO.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

The ICO and Independent Review Panel also expressed concern that data sharing agreements involving PHI may conflict with the GDPR (Saran 2017). The Independent Review Panel recommended convening a group to design a new model of regulation that addresses the conflicting goals of information technology companies, government medical services, regulatory authorities, and the public:

> That tech providers, the Department of Health and the Information Commissioner should discuss together a new system which protects patient data whilst allowing innovation and that collaborative discussions should take place in safe places, similar to Research Council "sandpits" in order to create a new model for regulation (Independent Review Panel 2017, 11).

3. *Technical/algorithmic governance forms.*

In response to widespread criticism, DMH announced it is developing an automated audit of PHI health data access using technology similar to blockchain to address concerns (DeepMind 2017f; Hern 2017). This "Verifiable Data Audit" is an innovative form of sociotechnical governance intended to provide a real-time audit and verification of data access and use, enabling appropriate authorization for access of PHI that is consistent with patient consent. DMH notes that this audit includes:

> giving our partner hospitals an additional real-time and fully proven mechanism to check how we're processing data … For example, an organization holding health data can't simply decide to start carrying out research on patient records being used to provide care, or repurpose a research dataset for some other unapproved use. In other words: it's not just where the data is stored, it's what's being done with it that counts. We want to make that verifiable and auditable, in real-time, for the first time (DeepMind, 2017f, 8).

That is, while this digital ledger is intended to focus largely on assuring that data is private and secure (i.e., who has access to the data, where it moves), DeepMind claimed that it will also enable transparency of data use (i.e., how it is processed). This latter function is as yet unsubstantiated, as the machine learning algorithms that actually make decisions about whether to alert hospital staff or not are still a black box.

Discussion

The DMH-Royal Free Trust case highlights how the interests of different stakeholders in PHI data diverge and often conflict. On the one hand, the Streams application is reportedly benefiting individual patients such as those undergoing treatment for acute kidney injury at the Royal Free Trust (DeepMind 2017e; Lydall 2017; Royal Free Hospital 2017). On the other hand, this broad PHI data-sharing arrangement has generated widespread concern and criticism (Hodson 2016; Independent Review Panel 2017; Powles and Hodson 2017; Ram and Waters 2018). Consequently, policymakers and regulators have examined the details of this data sharing arrangement and debated post hoc what the roles, responsibilities, and

allowable actions are for different parties (Lydall 2017; Powles and Hodson 2017).

Our analysis, summarized in Table 3, highlights instances of actual or anticipated contextual integrity conflicts with the shift of governance of PHI data from in-house patient care within Royal Free Trust to a partnership with a for-profit IT entity, DMH. Existing governance forms, particularly regulatory governance forms, failed to anticipate the rise of PHI contextual integrity conflicts.

[insert table 3 about here]

The resulting public and regulatory scrutiny of these conflicts led to the development of new governance forms. The objective here was not to prohibit the flow of PHI data, as the potential societal benefit of healthcare AI is substantial, but rather to reinforce the existing norms for privacy for a new use context. For example, the Institutional Review Board established by DMH is a governance form intended to build public trust in DMH's PHI data-governance and to ensure regulatory compliance. DMH's Verifiable Data Audit is a socio-technical governance form intended to increase transparency in how DMH uses PHI data. The Royal Free Trust's online patient opt-out feature provides another socio-technical governance form to give patients' voice in how their PHI data are used.

Despite adjustments to governance forms, some major concerns about the contextual integrity of PHI have yet to be fully resolved, particularly those related to the efficacy of informed consent practices to give patients' voice. Expectations for the potential societal benefits to be realized from aggregated PHI complicate the question of how much voice citizens should have in these decisions. Here prioritizing large scale, open-ended data sharing for the purpose of health system improvement and innovation, while desirable values per se conflict with respect for patients' voice. This was evident in the initial failures of system features for patients to either opt-in or opt-out of the data sharing arrangement. Sharon (2016) notes that many emerging forms of big data research pose challenges to our existing understanding of informed consent, where all possibilities of data use, and the risks it poses, are not known at the time of data collection. The lack of transparency in the corporate data-sharing arrangements and the opacity of AI/machine learning systems will necessarily challenge governance and sociotechnical approaches to informed consent. Therefore new models of "'open," "broad," and "portable'" consent have emerged (Sharon 2016).

Our analysis of this case highlights the governance challenges of harnessing PHI big data resources to enhance society, organizational effectiveness, and individual lives, while also respecting the rights and interests of diverse stakeholders. Realizing value from PHI data is not a zero-sum game, as health data can be used and reused to support multiple forms of value creation (Barrett, Oborn, and Orlikowski 2016; Tempini 2017). Neither is it necessarily a win-win situation, as the value of PHI data is tied to the social and economic purposes and outcomes of data-inspired policies and practices, where conflicts in values, norms, and interests among stakeholders abound. This information is often intensely personal (Independent Review Panel 2017), and its (mis)use may hold unintended or unforeseen negative consequences for individuals.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

In this case, the two dominant organizational actors – DMH and Royal Free Trust – agreed to share PHI data so as to promote health services innovation through advanced ICTs (including AI approaches) for the benefit of patients and the public. This is an immensely exciting prospect. One could argue that the provision of specific services to patients (such as the Streams application) warrants DMH's reuse of the PHI data for applications in other domains, and ultimately monetization of these applications. In this case, unbounded possibilities of PHI data reuse under DMH's governance form engendered contextual violations (cf. Denham 2017) requiring mitigation through new governance forms such as advisory committees (e.g., the Independent Review Panel) and DMH's Verifiable Data Audit (DeepMind 2017f; Hern 2017) to maintain some balance.

Regulations are important tools for governing organizational behavior, but they also pose the risk of stifling innovation. Ultimately, data governance goals and structures need to optimize the outcomes across the interests of a diverse array of stakeholders. The failure of existing laws and policies to fully protect patients' PHI in this case led to the Independent Review Panel to recommend new regulatory forms amid growing concern that existing regulatory mechanisms may not be adequate to govern big data analytics and AI. For example, it is possible that an algorithm may learn to be more precise in identifying health concerns for particular populations while being less precise for others, thereby reinforcing or exacerbating existing health disparities. This algorithmic discrimination may be deliberate or an unexpected outcome. In other domains, scholars have begun to develop algorithmic audits, field experiments that detect discrimination (Sandvig, Hamilton, Karahalios, and Langbort 2014). But such efforts are blunted by the opacity with regard to how the algorithms work, given their corporate ownership and control. However, increased public pressure may lead to the furtherance of such an audits, or other technical means of providing oversight. This call is echoed by the Royal Statistical Society (2016), which suggests an inquiry about "methods that the public can use to hold algorithms to account" (3) and expert public testimony (cf. House of Lords Select Committee on Artificial Intelligence 2017).

The tensions regarding PHI use are likely to increase in the future. The UK's Data Protection Act was replaced with the European Union General Data Protection Regulation (GDPR) on May 25, 2018. Kuner et al. (2017) have observed that Article 22(1) of the GDPR relates to "personal data used for automated decisions" and notes that data should only be gathered for "specified, explicit, and legitimate purposes, and subsequent processing that is incompatible with those purposes is not permitted" (1). Since machine learning often relies on live data streams, "it may be difficult to reconcile such dynamic processes with purposes that are specified narrowly in advance" (1). Given such concerns in their review of the DMH-Royal Free Trust agreement, the Independent Review Panel concluded that a new model of regulation should be explored (Independent Review Panel 2017, 11).

The novel use of AI raises several questions about whose values are encoded into algorithms and thus translated into machine judgments. As many scholars have argued (e.g., Friedman and Nissenbaum 1996), technical systems are not value-free – they carry the biases of the social and cultural context in which they arise. Rakhal Gaitonde, a patient

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

advocate, noted:

> One of the basic steps in the design of an AI system is the setting of a goal for an agent … Playing Atari, the goal is to maximise your points with least effort … What would be an equivalent goal in patient care? A sense of wellbeing? The normalising of a set of biochemical parameters? In a situation where we are all only too aware of the way in which corporate interest has conspired to influence the definition of disease, how would one set the goal within the sustainable business model DeepMind have discussed? (quoted in Armstrong 2016, 182)

In cases where such technologies or practices can be shown to more effectively support relevant societal values, there exists moral justification to enable new data flows (Nissenbaum 2010) and novel governance forms. However, explicitly addressing the underlying value that is sought through use and reuse of PHI data, and, relatedly, the values and interests of all stakeholders, is critical to arriving at such judgments.

Conclusions

The vast and growing stockpiles of PHI hold the promise of substantive improvements in the precision of personalized healthcare, in the quality and reliability of health service delivery, and in addressing growing costs of health services. Such improvements will depend in large part on technology-enabled innovations via big data analytics to develop new forms of value form PHI. To more effectively govern the complex, emergent, and networked nature of PHI storage and utilization, balancing between individual rights and the potential public good that could be realized and also the interests of innovative firms hoping to profit from such activities, requires that we examine, within each context where data are generated, used, and reused, who are the stakeholders and what their relevant interests and values are.

Going beyond high-level assessments of cyber-security and general health data privacy protection (Independent Review Panel 2017; Powles and Hodson 2017), additional analytical tools and methods could help ensure that PHI governance goals are articulated, debated, and negotiated before conflicts develop, and ultimately these goals are realized in practice. As one approach, this study drew on Nissenbaum's notion of contextual integrity and our earlier research on forms of data governance to examine the controversial public-private partnership between the Royal Free Trust and DMH. It shows how an analytical framework employing governance dimensions can be used to assess how actual or anticipated violations of contextual integrity may develop as there is a shift from traditional governance forms, such as those centered on clinical entities that are expected to adhere to health privacy regulations, to novel governance forms, such as those in which for-profit innovation firms are powerful stakeholders with substantively different norms, values, and interests in PHI data.

This case study is limited to a single emerging case. It relied on publicly available documentary sources, and interviews with experts published in newspapers and trade journals. Therefore, there may be other stakeholders or views not captured in this study.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Future research is required to address a number of key issues in greater detail, such as: What governance forms can help individuals effectively find voice in how their data are governed and (re)used? How can we foster transparency in big data analytics with regards to AI/machine learning and intellectual property constraints? What other governance forms might emerge that will help balance these interests?

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

References

Agency for Healthcare Research and Quality. 2017. All-payer claims databases. Retrieve on December 29, 2017, from: https://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/apcd/index.html

Arndt, R. Z. 2018. Apple is officially in the EHR business: Now what? *Modern Healthcare*, January 26. Retrieved on March 16, 2018, from: http://www.modernhealthcare.com/article/20180126/NEWS/180129910

Acquisti, A., Brandimarte, L., and G. Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347(6221):509-514.

Anderson, C. L., and R. Agarwal. 2011. The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information. *Information Systems Research* 22(3):469-490.

Armstrong, S. 2016. The computer will assess you now. *The British Medical Journal, 355*(5680):1-3. doi.org/10.1136/bmj.i5680

Barrett, M., Oborn, E., and W. Orlikowski. 2016. Creating value in online communities: The sociomaterial configuring of strategy, platform, and stakeholder engagement. *Information Systems Research* 27:704-723. doi:10.1287/isre.2016.0648

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., and G. Escobar. 2014. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs* 33(7): 1123-1131.

Belanger, F., and H. Xu. 2015. Editorial: The role of information systems research in shaping the future of information privacy. *Information Systems Journal* 25:573–778.

Blumenthal, D. 2010. Launching HITECH. *New England Journal of Medicine* 2010(62): 382-385.

British Academy and the Royal Society. 2017. Data management and use: Governance in the 21st century. Retrieved on March 1, 2018, from: https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf

Cortez, A., Hsii, P., Mitchell, E., Riehl, V., and P. Smith. 2018. Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024 (White Paper). Washington, DC: Office of the National Coordinator for Health Information Technology. Retrieved on March 16, 2018, from: https://www.healthit.gov/sites/default/files/pghd_white_paper_final_formatted_by_onc_508_11-29-17.pdf

Data Governance Institute. nd. About us. Retrieved on January 20, 2018, from: http://www.datagovernance.com/

Davenport, T. H., Barth, P., and R. Bean. 2012. How 'big data' is different. *MIT Sloan*

Management Review 54(1):22-24.

Davidson, E. J., Østerlund, C. S., and M. J. Flaherty. 2015. Drift and shift in the organizing vision career for personal health records: An investigation of innovation discourse dynamics. *Information and Organization* 25(4):191-221.

DeepMind. 2017a. About DeepMind Health. Retrieved on October 15, 2017 from: https://deepmind.com/applied/deepmind-health/deepmind-health-faqs/

DeepMind. 2017b. DeepMind Ethics & Society. Retrieved on March 8, 2017 from: https://deepmind.com/applied/deepmind-ethics-society/fellows/

DeepMind. 2017c. DeepMind Health and personally identifiable data. Retrieved on October 15, 2017 from: https://deepmind.com/applied/deepmind-health/data-security/personally-identifiable- data/

DeepMind. 2017d. DeepMind Health's Independent Review Panel. Retrieved on October 15, 2017 from: https://deepmind.com/applied/deepmind-health/transparency-independent-reviewers/independent-reviewers/

DeepMind. 2017e. Streams in NHS hospitals. Retrieved on October 17, 2017 from: https://deepmind.com/applied/deepmind-health/working-nhs/how-were-helping- today/

DeepMind. 2017f. Trust, confidence and verifiable data audit. Retrieved on November 20, 2017 from: https://deepmind.com/blog/trust-confidence-verifiable-data-audit/

DeepMind Health. 2017. DeepMind Health 2017 patient involvement and engagement events. Retrieved on December 3, 2017 from: https://deepmind.com/documents/122/2.%20DeepMind%20Health%202017%20Patient%20Involvement%20events%20report%20VF%20LARGE%20TEXT%20VERSION.pdf

Deloitte. 2016. 2016 global health care outlook: Battling costs while improving care. Retrieved on March 6, 2018 from: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-2016-health-care-outlook.pdf

Deering, M. J. 2013. Issue Brief: Patient-generated health data and health IT. Washington, DC: Office of the National Coordinator for Health Information Technology.

Denham, E. 2017. Four lessons NHS trusts can learn from the Royal Free case. Retrieved on October 15, 2017 from: https://iconewsblog.wordpress.com/2017/07/03/four-lessons-nhs- trusts-can-learn-from-the- royal-free-case/

Diamond, C. C., Mostashari, F., and C. Shirky. 2009. Collecting and sharing data for population health: A new paradigm. *Health Affairs* 28(2): 454-466.

Eden, K. B., et al. 2016. Barriers and facilitators to exchanging health information: A systematic review. *International Journal of Medical Informatics* (88):44-51. doi:10.1016/j.ijmedinf.2016.01.004

Elliott, T. E., Holmes, J. H., Davidson, A. J., La Chance, P. A., Nelson, A. F., and J. F.

Steiner. 2013. Data warehouse governance programs in healthcare settings: A literature review and a call to action. *eGEMs* 1(1):15-21.

Eyal, N. 2013. Denial of treatment to obese patients: The wrong policy on personal responsibility for health. *International Journal of Health Policy and Management* 1(2): 107-110.

Friedman, B., and H. Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on Information Systems* 14(3):330–347.

Hawkes, N. 2016. NHS data sharing deal with Google prompts concern. *The British Medical Journal 353*: 2573. Retrieved on October 15, 2017 from: http://www.bmj.com/content/353/bmj.i2573

Hern, A. 2017. Google's DeepMind plans Bitcoin-style health record tracking for hospitals. *The Guardian*, March 9, n.p. Retrieved on December 1, 2017 from: https://www.theguardian.com/technology/2017/mar/09/google-deepmind-health-records-tracking-blockchain-nhs-hospitals

Hodson, H. 2016. Revealed: Google AI has access to huge haul of NHS patient data. *New Scientist*, April 29, n.p. Retrieved on September 30, 2017 from: https://www.newscientist.com/article/2086454-revealed-google- ai-has-access-to-huge-haul-of-nhs-patient-data/

Holmes, J. H., Elliott, T. E., Brown, J. S., Raebel, M. A., Davidson, A., Nelson, A. F., and J. F. Steiner. 2014. Clinical research data warehouse governance for distributed research networks in the USA: A systematic review of the literature. *Journal of the American Medical Informatics Association* 21 (4):730-736.

House of Lords Select Committee on Artificial Intelligence. 2017. Artificial intelligence: Is it good for our health? Retrieved on November 22, 2017 from: https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/healthcare-artificial-intelligence-evidence-session/

Hripcsak, G., et al. 2014. Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 health policy meeting. *Journal of the American Medical Informatics Association* 21 (2):204– 211.

Independent Review Panel. 2017. DeepMind Health Independent Review Panel annual report. Retrieved on September 30, 2017 from: https://deepmind.com/documents/85/DeepMind%20Health%20Independent%20Review%20Annual%20Report%202017.pdf

Information Commissioner's Office. 2017. Royal Free: Google DeepMind trial failed to comply with data protection law. Retrieved on September 30, 2017 from: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/

"Information sharing agreement." 2015. Retrieved on September 30, 2017 from: https://storage.googleapis.com/deepmind-data/assets/health/Royal%20Free%20-%20DSA%20-%20redacted.pdf

Khatri, V., and C. V. Brown. 2010. Designing data governance. *Communications of the ACM* 53

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

(1):148-152.

Kuner, C., Svantesson, D. J. B., Cate, F. H., Lynskey, O., and C. Millard. 2017. Machine learning with personal data: Is data protection law smart enough to meet the challenge? *International Data Privacy Law* 7(1):1-2.

Ladley, J. 2012. *Data governance: How to design, deploy, and sustain an effective data governance program*. Waltham, MA: Morgan Kaufmann.

Libert, T. 2015. Privacy implications of health information seeking on the Web. *Communications of the ACM* 58(3):68-77.

Lomas, N. 2017. DeepMind now has an AI ethics research unit: We have a few questions for it. *TechCrunch,* October 4. Retrieved on March 8, 2018 from: https://beta.techcrunch.com/2017/10/04/deepmind-now-has-an-ai-ethics-research-unit-we-have-a-few-questions-for-it/?ncid=rss

Lycett, M. 2013 "Datafication": Making sense of (big) data in a complex world. *European Journal of Information Systems* 22(4):381-386.

Lydall, R. 2017. New mother receives pioneering kidney treatment after app detects life-threatening illness. *The Standard*, February 27. Retrieved on September 30, 2017 from: http://www.standard.co.uk/news/health/new-mother- receives-pioneering-kidney-treatment-after-app-detects-lifethreatening-illness-a3476936.html

Mai, J. E. 2016. Big data privacy: The datafication of personal information. *The Information Society* 32(3):192-199.

Montgomery, K., Chester, J., and K. Kopp. 2018. Health wearables: Ensuring fairness, preventing discrimination, and promoting equity in an emerging Internet-of-Things environment. *Journal of Information Policy* 8: 34-77. doi:10.5325/jinfopoli.8.2018.0034

Murdoch, T. B., and A. S. Detsky. 2013. The inevitable application of big data to health care. *Journal of the American Medical Association* 309(13):1351-1352.

Nissenbaum, H. 2010. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

OECD. 2017. Recommendation of the OECD Council on health data governance. Retrieved on March 5, 2018 from: https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf

Powles, J., and H. Hodson. 2017. Google DeepMind and healthcare in an age of algorithms. *Health Technology* 7(4):351-367. doi.org/10.1007/s12553-017-0179-1

Ram, A., and R. Waters. 2018. Alphabet AI unit urged to clarify its business model. *Financial Times*, June 14. Retrieved on July 31, 2018 from: https://www.ft.com/content/215062da-6fe3-11e8-852d-d8b934ff5ffa

Rodwin, M. 2009. The case for public ownership of patient data. *Journal of American Medical Association* 302(1):86-88. doi:10.1001/jama.2009.965

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Rosenbaum, S. 2010. Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research* 45(5p2):1442-1455.

Ross, T., Ng, D., Brown, J. S., Pardee, R., Hornbrook, M. C., Hart, G., and J. F. Steiner. 2014. The HMO Research Network virtual data warehouse: A public data model to support collaboration. *eGEMs* 2(1):1049-1058.

Royal Free Hospital. 2017. New app helping to improve patient care. Retrieved on September 30, 2017 from: https://www.royalfree.nhs.uk/news-media/news/new-app-helping-to- improve-patient-care/New app helping to improve patient care

Royal Free London NHS Foundation Trust. 2016. Memorandum of understanding with DeepMind Technologies. January 28. Retrieved on September 30, 2017 from: https://drive.google.com/file/d/0BwQ4esYYFC04anR4VHM3aXZpMTQ/view

Royal Statistical Society. 2016. Evidence to the Royal Society and British Academy on data governance. Retrieved on March 4, 2018 from: http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS-evidence-to-Royal-Society-and-British-Academy-on-Data-Governance-Nov-2016.pdf

Sandvig, C., Hamilton, K., Karahalios, K., and C. Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. Paper presented at Data and Discrimination: Converting Critical Concerns into Productive Inquiry (preconference), International Communication Association Annual Conference, Seattle, WA, May.

Sankin, A. 2017. Your medical data is for sale, and there's nothing you can do about it. *Reveal*, January 20. Retrieved on December 1, 2017 from: https://www.revealnews.org/blog/your-medical-data-is-for-sale-and-theres-nothing-you-can-do-about-it/

Saran, C. 2017. NHS data not fit for AI, Lords select committee told. *ComputerWorld*, November 23. Retrieved on December 1, 2017 from: http://www.computerweekly.com/news/450430604/NHS-data-not-fit-for-AI-Lords-Select-committee-told

Scott, W. R. 2001. *Institutions and organizations* (2nd Edition). Thousands Oaks, CA: Sage.

Services agreement between DeepMind Technologies Limited and Royal Free London NHS Foundation Trust. 2016. Retrieved on September 30, 2017 from: https://storage.googleapis.com/dmhir-documents/DeepMind%20RFL%20Services%20Agreement.pdf

Sharon, T. 2016. The Googlization of health research: From disruptive innovation to disruptive ethics. *Personalized Medicine* 13(6):563-574.

Solon, O. 2014. A simple guide to Care.data. *Wired,* February 7. Retrieved on July 6, 2018 from: https://www.wired.co.uk/article/a-simple-guide-to-care-data

Solove, D. 2010. *Understanding privacy*. Cambridge, MA: Harvard University Press.

Tanner, A. 2016. This little-known firm is getting rich off your medical data. *Fortune*,

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

February 9. Retrieved on June 12, 2017 from: http://fortune.com/2016/02/09/ims-health-privacy-medical-data/

Tempini, N. 2017. Till data do us part: Understanding data-based value creation in data-intensive infrastructures. *Information & Organization* 27(4):191-210.

Triggle, N. 2014. Giant NHS database rollout delayed. BBCNews.com, February 18. Retrieved on July 14, 2018 from: https://www.bbc.co.uk/news/health-26239532

U.S. Department of Health and Human Services. 2016. *Examining oversight of the privacy & security of health data collected by entities not regulated by HIPAA*. Washington, DC: U.S. Department of Health and Human Services.

U.S. Department of Health and Human Services. 2018. *Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024*. Washington, DC: U.S. Department of Health and Human Services.

Wagner, K. 2016. Risk stratification for better population health management. Healthcare Financial Management Administration. *Leadership+,* July 28. Retrieved on March 15, 2018 from: http://www.hfma.org/Leadership/Archives/2016/Summer/Risk_Stratification_for_Better_Population_Health_Management/

Wakefield, J. 2017. Google NHS deal rebuked again by DeepMind panel. BBC News, July 5. Retrieved on September 30, 2017 from: http://www.bbc.com/news/technology-40497020

Winter, J. S. 2014. Surveillance in Ubiquitous Network Societies: Normative conflicts related to the consumer in-store supermarket experience in the context of the Internet of Things. *Ethics and Information Technology* 16(1): 27-41.

Winter, J. S., and E. Davidson. 2017. Investigating values in personal health data governance models. Paper presented at the 23rd Americas Conference on Information System, Boston, MA, August.

Zuboff, S. 2015. Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1):75-89.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Data Domain	Stakeholders	Value/application	Governance Goal	Governance Form
Organizationally-collected data, e.g., Individual's health history EMR (emergency medical responders) clinical encounter data Prescription /pharmacy data Lab data Imaging data **Personally-generated data**, e.g., Activity data (e.g., diet, exercise) Clinical data (e.g., glucose level) **Digital trace data**, e.g., Behavioral data from digital sources Online shopping	**Direct**, e.g., Individual Family members Health care provider 3rd party payers Employers **Indirect**, e.g., Government policy makers Heath researchers "The public" or "communities" **Health system**, e.g., Health IT firms Pharmaceutical firms Medical equipment manufacturers	**Improvements via data analytics,** e.g., Individual's health Organizational performance Health system's efficiency and effectiveness Evidence-based health services Community or population health Monetization of data	**Assuring and maintaining**, e.g., Trust in governance Privacy Data security Regulatory compliance **Facilitating**, e.g., Data access Data analytics Innovations **Protecting IP**	**Policies**, e.g., Data privacy **Regulations**, e.g., EU's GDPR (General Data Protection Regulation) **Organizational**, e.g., Data access committee Neutral third-party data organization **Technology**, e.g., Algorithms Cyber-security tools **Standards**, e.g., Data harmonization via codes Interoperability protocols

Winter, J. S., & Davidson, E. (2019). “Big data governance of personal health information and challenges to contextual integrity.” *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Web searches				

Table 1. Personal health information (PHI) data governance dimensions and examples.

Winter, J. S., & Davidson, E. (2019). “Big data governance of personal health information and challenges to contextual integrity.” *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

2015 September 29	2016 February	2016 April 29	2016 May	2016 November 10	2017 July 1	2017 July 3	2017 November 23
• Service Agreement between Royal Free Trust and DeepMind Health, which included PHI data sharing. Intitially, this was not a public document.	• DeepMind Health Independent Review Panel formed to examine DeepMind Health's current and planned work with NHS.	• *New Scientist* publishes the 9/29/15 Service Agreement it received via a FOI request.	• Information Commissioner's Office(ICO) begins investigation into the PHI data sharing arrangement (noting at least one complaint from the public).	• Revised Service Agreement between Royal Free Trust and DeepMind Health.	• DeepMind Health releases first Independent Review Panel report. Panel calls for DeepMind Health to address ICO'sconcerns about regulatory violations.	• ICO rules that the agreement failed to comply with the Data Protection Act.	• House of Lords Select Committee on AI expresses concern that data access required by AI may be conflict with upcoming EU's General Data Protection Regulation (GDPR).

Table 2. Timeline: Google DMH-Royal Free Trust partnership.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Tensions due to PHI data flow to new context	Adjustments in governance forms	Unresolved tensions
PHI Data Domains		
• Five years of health data on all patients provided to DMH, a for-profit IT firm. • Potential to combine patient health data with other data domains, such as parent firm's archives of consumer trace data.	• Public assurances by DMH that PHI would not be combined with data from other sources. • "Verifiable Data Audit" promised for data access and use, consistent with patient consent.	• DMH's intention to develop AI applications.
Stakeholders		
• Public and policy makers' reactions to data sharing arrangement with a for-profit firm. • Data sharing with DMH AI researchers, who have no direct involvement in health services.	• Public assurances that Alphabet will not have access to the data. • Making DMH data sharing agreements publicly accessible. • DMH's patient engagement workshops to increase transparency and patient buy-in related to health data and AI.	• New stakeholders in the future through open-ended data sharing arrangement for development of AI applications. • Questions about DMH's eventual business model.
PHI Data Value		
• Value of PHI beyond individual patient care and improvement of NHS for development of corporate	• Revised Services Agreement acknowledging acquisition of PHI for AI development with regulatory	• Standardizing and centralizing NHS clinical data to enhance its value for AI applications.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

intellectual property, applications and services, and algorithms.	approvals; DMH's retention of IP rights.	• Negotiating sharing with NHS of commercial value derived from PHI.
Governance Goals		
• Emphasizing PHI reuse for health service innovation and IP development over established governance goals, e.g. privacy, transparency in PHI use, and regulatory compliance.	• DeepMind Ethics & Society Fellows unit established to assess privacy, transparency, and fairness. • DeepMind policy to make future data sharing agreements public. • Regulatory requirement to comply with data protection regulations in future trials.	• Conflict of interest in DeepMind auditing its own complaince. • Standardization of NHS PHI data to accommodate AI development in ways that reduce value of PHI for clinical care.
Governance Forms		
• Identifiable PHI data on *all* patients seen at Royal Free Trust hospitals migrated without notification to individual patients. • Possible attempts to avoid patient consent regulations to enable efficient access to a broad spectrum of PHI. • First data sharing agreement differed from standard NHS third-party agreements.	• ICO oversight committee for all uses of the PHI created. • DMH Independent Review Panel assembled. • Revised data sharing agreement in November 2016. • "Verifiable Data Audit" to provide a real-time audit and verification of data access and use.	• Oversight committee panel members are not public servants and therefore not subject to NHS oversight and accountability. • ICO ruling that revised data sharing agreement failed to comply with the Data Protection Act and finding of possible conflict with GDPR. • Weak execution of opt out option via the Royal Free Trust's website.

Table 3. Summary of contextual integrity conflicts and adjustments to PHI governance forms.

Winter, J. S., & Davidson, E. (2019). "Big data governance of personal health information and challenges to contextual integrity." *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

Notes

[1] The acronym PHI may apply to personal health information broadly, or to data protected by regulation. We use the term PHI in the broader sense.

[2] Our preliminary taxonomy includes PHI governance forms such as organizational (e.g., a hospital), inter-organization (e.g., a health-information exchange entity), community (e.g., a social network of patients sharing clinical data), personal (e.g., individuals maintaining their own personal health records), marketplace (e.g., commercial PHI data aggregator), and public good (e.g., entities sharing PHI for academic or policy research). A full description is beyond the scope of this paper. Please see Winter and Davidson (2017) for additional details.

[3] Existing third-party data sharing arrangements included the 2014 Care.data program that made patient data available to organizations both inside and outside of the NHS, including health charities pharmaceutical companies, universities, hospital trusts, and other private companies, subject to approval (Solon 2014). This Care.data database sought to centralize PHI and has been a contentious process, with patient advocates claiming that patients have not been properly informed and given meaningful options to opt out (Triggle 2014). The program was halted in 2016.

[4] Interestingly, the report recommended the Independent Reviewers receive an honorarium, suggesting that the group's allegiance may be to DMH in the future.

Winter, J. S., & Davidson, E. (2019). “Big data governance of personal health information and challenges to contextual integrity.” *The Information Society, 35* (1), 36-51. doi:10.1080/01972243.2018.1542648

---

[5] See https://www.royalfree.nhs.uk/deepmind-patient-opt-out-form/ for the “DeepMind patient opt out form”.