

Investigating Values in Personal Health Data Governance Models

Full Paper

Jenifer Sunrise Winter
University of Hawai'i Manoa
School of Communications
jwinter@hawaii.edu

Elizabeth Davidson
University of Hawai'i Manoa
Shidler College of Business
edavidso@hawaii.edu

Abstract

In this paper, we seek to deepen discourse on health data governance beyond the important issues of privacy and data security to consider what types of value are potentially afforded by personal health information (PHI) data and, importantly, whose values and interests shape governance structures and goals toward realizing value. We conducted a discourse analysis of texts addressing PHI data use and governance. Through analysis of a broad array of documents using qualitative analysis and guided text mining, we identified six overlapping, but distinctive, models for PHI governance. Each model presents an array of stakeholders, value to be realized from analysis, assumed stewardship roles, and governance structures and goals. This analysis extends consideration of widely shared governance goals, highlighting possible issues and conflicts among actors' values and interests, particularly when data "slip" between governing models. We consider policy implications and areas of future research from this analysis.

Keywords

Personal health data, privacy, data governance, discourse analysis.

Introduction

As more business 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 are renewing management attention to long-acknowledged opportunities in and challenges to information and data governance (Cohasset Associates 2014). Generally, the business goals of governance are to manage and then to "mine" these vast data stores to improve organizational performance (Davenport et al 2012) or for entrepreneurial, innovative, and competitive opportunities (Lycett 2013).

Within the healthcare industry, digitized health data hold the promise to improve efficiencies in the healthcare industry (Blumenthal 2010) and to "greatly expand the capacity to generate new knowledge" (Murdoch and Detsky 2013, p. 1351) to improve delivery and precision of clinical medicine. Enhanced storage, processing, and analytics give new views/perspectives into healthcare data down to the personal level as well as across populations defined by disease states, geographies, ethnic groups, and so on. Data integrity, privacy, and security of personal health information (PHI) are subject to government regulation due in large part to the potential harm to individuals or groups, depending on how these data are used (e.g., to discriminate based on health status). Thus, along with the opportunities for new value to be mined from accumulating health data stores, there are growing concerns about how to protect privacy and security of these valuable data resources (Rosenbaum, 2010; Belanger 2015).

This has inspired much discourse about responsibilities for the security of health data, protection of privacy from unauthorized uses (without necessarily addressing what constitutes "authorized" access), and managing the integrity of the data by ensuring it is accurate and complete. This, in turn, has revitalized discourse on data governance and the related concept of data stewardship (e.g., Rosenbaum 2010; Hripcsak et al 2014). These discourses range from a plethora of practical advice (i.e., how to make data governance work) to less frequent discussions of philosophical and societal implications of data

stewardship. Joining in this latter approach, we argue that in addition to “how to” questions, we need to consider the “why” and “for whom” questions that are the essence of data stewardship and governance.

In this paper, we seek to broaden and deepen the discourse on health data governance and stewardship beyond important issues of privacy and data security to consider what types of value are potentially afforded by PHI data and, importantly, whose values and interests shape governance structures and goals. To do so, we conducted a discourse analysis of texts that address PHI data use and governance. By reading and analyzing a broad array of documents and using qualitative analysis and guided text mining, we identified six overlapping but distinctive models for personal health data governance. Each model presents an array of stakeholders, value to be realized from analyzing data, assumed stewardship roles, and governance structures and goals. This analysis extends consideration of widely shared governance goals (privacy and security) to highlight possible issues and conflicts among actors’ values and interests, particularly when PHI data “slip” beyond one governing model to other models.

Background and motivation

As data are increasingly understood as a strategic resource, organizations have implemented data governance programs to create business value from data assets (Khatri and Brown 2010). Data governance relates to the management of enterprise data, including issues of data aggregation and integration, data integrity, security, and privacy. Complying with regulatory requirements (such as the Sarbanes-Oxley Act of 2002) place constraints on practices of data governance. The information systems field has long acknowledged these opportunities and challenges of managing data stores in order to reuse data for a wider scope of business analysis (i.e., “business intelligence”). Data warehousing developed as a technical and organizational approach to aggregate and manage data generated within discrete transactional systems and diverse business functions and units (e.g., Inmon 1996; Kimball and Ross 2013). Inter-organizational data management and exchange among supply chain network partners have focused on technology standards (such as EDI, or electronic data exchange) and organizational arrangements (e.g., Elgarah et al 2005). Now, with an exponential increase in the volume and variety of “big data”, researchers are again focusing on information and data governance issues (George et al 2014).

Digitization of health-related data is progressing rapidly, generating highly coveted data sources and robust discussions of data governance (Diamond et al 2009; Kanaan and Carr, 2009; Rodwin, 2009; Rosenbaum, 2010; Elliott et al 2013; Hripcsak et al 2014; Ross et al 2014). *Data governance* entails the processes and institutional structures for managing a data store (or data domain), including policies for granting accessing (to who, how) and appropriately using data, whereas *data stewardship* implements governance goals by acquiring, storing, and granting access to data; stewardship also conveying a fiduciary (trust) relationship between the steward and the individuals or entities whose data are governed (Rosenbaum et al 2010, pg. 1444). Government policymakers’ attempts to regulate health data exchange and governance among industry actors (e.g., healthcare providers) highlights pervasive issues with access rights, ownership, data standardization, privacy, and security of health data (Rosenbaum, 2010).

High expectations for the societal and economic value that could be realized by using health data for health system improvements and innovation add organizational energy and resources, as well as societal pressure, on healthcare sector stakeholders to resolve governance issues in order to realize these values from healthcare data (Cohasset Associates 2014; Hripcsak et al 2014). Nonetheless, realizing effective governance of healthcare data remains an elusive goal. Rosenbaum (2010) observes that, “*An intense struggle over health information access has been a hallmark of the health care system for decades ... Opponents raise a host of concerns, citing patient privacy, the confidential nature of the patient/professional relationship, and health information security. Naturally, opponents also are focused on their own interests, given the potentially deleterious impact of uncontrolled data access on their liability under a host of civil and criminal laws, as well as on their competitive market position*” (p. 1443). This latter point highlights two related issues, i.e., what are *types of value* that might be derived from access to and use of a health data store, and which *stakeholder values and interests* guide the design of governance and stewardship policies, structures, and goals?

Critiquing the aggregation of digital Internet trace data, Zuboff (2015) posits that “surveillance capitalism” creates economic value for technology firms and their clients (such as advertisers) in the form of profits, market share, and competitive advantage, which may sacrifice individual privacy and autonomy

of the individuals subject to invisible surveillance. On the other hand, Jin and Feenberg (2015) argue that ICTs provide affordances for value creation by and among community members themselves that can alter such disparities. These questions about value afforded by data and stakeholder values and interests are relevant to health data governance. Beyond clinical and organizational data, many new sources and types of personal health data are now collected and aggregated (e.g., consumer-generated data from fitness trackers and de-identified prescription records collected by data aggregators). In the U.S., these latter types of personal health data are not subject to government regulations such as the Health Insurance Portability and Accountability Act (HIPAA). These data may also be used in ways that are invisible to individuals who are the subjects of those data and who do not benefit (and might be harmed) by their use (Office of the National Coordinator for Health Information Technology 2015a; Tanner 2016).

To investigate possible issues related to value afforded by PHI data and stakeholder values and interests in data governance requires that we systematically examine existing health data governance models as well as models emerging as new sources of PHI data become available. As a first step, in this paper we report our analysis of six overlapping, but distinctive, models and the dimensions that characterize each.

Methods

We conducted a discourse analysis of texts related to PHI data governance using an iterative, grounded approach involving qualitative analysis and coding along with software-assisted text mining. For initial coding, we used WordStat 7.13 text mining software to identify patterns in the documents and then conducted manual coding using QDA Miner 4.133 software to refine and extend coding categories.

To create the corpus for analysis, we selected documents related to data governance and stewardship, particularly those related to health data and PHI data specifically. We sought documents that demonstrated perspectives of a variety of relevant stakeholders, i.e., academics, government agencies (federal and state), citizens' groups, corporations, journalists, not-for-profit lobbyists, and health-care industry consultants. Documents included published journal articles available in indexed databases, as well as white papers, policy documents, websites, and blogs available on the Internet. Search and retrieval was an iterative process, guided by the ongoing analysis. It was not our intention to amass all relevant works or a statistically representative sample of documents, but to ensure that the corpus was rich and represented the various viewpoints of key stakeholders (see Davidson et al 2015 for a similar approach.) Because the field is rapidly changing due to technology advances and regulatory action, we primarily focused on documents published in the past five years (87.5% of the corpus was published in 2011 or later, and the remainder were seminal documents). In total, we selected 160 documents for inclusion.

We prepared the corpus for analysis by cleaning the data (i.e., removing any unnecessary information and designating stop words). This was accomplished in WordStat 7.13. Categories were then defined by the researchers using WordStat's dictionary tool. Once the data had been properly prepared in WordStat, we used the keyword in context (KWIC) feature to perform topic extraction, identify codes by machine-guided review of the text and import them into QDA Miner 4.133 for further analysis. Documents were then manually coded (in QDA Miner 4.133) to refine and extend coding categories from the text mining steps. By examining data dimensions resulting from these steps (data domain, actors, stakeholders, value afforded by data to stakeholders, governance structures, governance goals) we identified the six (6) models of data governance for PHI data discussed below. Our analysis highlights the relationships and potential conflicts among types of value from PHI data and the values shaping these governance models.

Findings

We present findings by first explaining the dimensions that characterize and differentiate governance models and then elaborating the six (6) models we identified.

Data Domains: PHI data with personally-identifying (or potentially identifying) characteristics include: medical history data; clinical data collected in EHR systems of hospitals, physicians, and laboratories; pharmacy prescription data; patient-generated health data from medical devices (such as a glucose monitor) or general purpose activity tracker; and medical expense claims data. Other PHI data domains are created by analyzing data not directly connected to health status, but health status can be inferred analytically. These data can be characterized as consumer-generated health data, for instance

trace data from Internet search activities, credit card purchases and online shopping, or geospatial/location data. Such data are used to generate personal health profiles and, if these profiles correlate with a commercially desirable outcome, may lead to more or less favorable treatment for those with certain health conditions, even when data are de-identified from specific individuals (Libert 2015).

Value afforded by PHI data: Health data value arises in how these data can be used to create direct sources of value to stakeholders. The value afforded is not limited to initial use, as data can be reused for multiple other purposes. Moreover, how data are governed, and the stewardship practices for aggregating and permitting access to data, can afford new forms of value for different stakeholders or inhibit realization of value by others. For instance, healthcare providers that adopt EHR use the digitized clinical data for day-to-day practices and, importantly, to justify and submit payment claims to a health insurer (in the U.S.) or other funders. These data, collected for operational and revenue-generating purposes, could be aggregated and mined to identify system-wide opportunities to improve organizational efficiencies to benefit various payers (e.g., patients, insurers). Realizing these latter forms of value may compromise the self-interest of the provider organization, if inefficiencies have been compensated (e.g., payment for duplicate services). Table 1 illustrates key value types identified through the analysis.

Code	Description	Stakeholders primarily affected	Example
Individual health enhancement	Using PHI data to improve one's own health and wellness	Individual	"These wearable devices have been adopted by individuals seeking to enhance their personal fitness through increased personal health surveillance and social connections with others using the devices" (Chiauzzi et al 2015, p. 2)
Organizational improvement	Using PHI data to improve efficiency, effectiveness, marketability, etc.	The organization that generates the data source	"The data translate into billions of dollars for many types of enterprises, in the form of new markets, new products, and cost-savings" (Sarasohn-Kahn 2014, p. 3)
Evidence-based healthcare	Using PHI data to assess and specify best clinical treatments	Patients; Health service providers	"So there would seem to be increased opportunities for analyzing large volumes of composite data to improve quality, comparative effectiveness, and public health" (Diamond et al 2009, p. 455)
Population health management	Using PHI to manage delivery of health services to identified groups to reach improvement goals	Patients in the population group; Health service providers (with value-based reimbursement)	"It is clear that the context of DDD [digital disease detection] differs in significant ways from other types of big data activity concerned with health. DDD has a public health function, aiming ultimately to improve health at the population level" (Vayena et al 2015, p. 3)
Health system improvement	Using broadly aggregated PHI to reduce costs and to improve quality and safety of health services	Society at large (improved health reduced tax burdens); Health system funders (e.g. government)	"Use of these health data beyond direct clinical care of individuals from whom the data were collected is essential to enhance healthcare experiences for individuals, improve healthcare quality and outcomes for individuals and population groups, and reduce system-wide healthcare costs" (Hripcsak et al 2014, p. 206)
Evidence-based pricing	Using PHI data to justify prices of health care services, medications, etc.	Pharmaceutical firms; Health system providers	"Consumers' health scores could be useful for providers and payers as they move to value-based payment. A McKinsey & Company analysis points to a trillion-dollar savings over the next decade by tying payments more aggressively to outcomes" (Sarasohn-Kahn 2014, p.8)
Monetization	Aggregating PHI data for resale to other organizations	3 rd party aggregators; Clients or partner of aggregators	"In turn, IMS sells insights from its more than half a billion patient dossiers mainly to drug companies. So-called health care data mining is a growing market and one largely dominated by IMS" (Tanner 2016, ¶ 1)

Table 1. Value Types Afforded by PHI Data

Stakeholders: Stakeholders are individuals or collective actors such as organizations that have some “stake” in the potential value that PHI data represents, and who are affected in some manner by data access and use policies and practices. Some stakeholders are central actors in a governance model, such as an organization that generates the digitalized PHI data. Other stakeholders may not actively generate or use the PHI data but are influenced by data use, for instance, individuals whose prescription data are aggregated and sold to pharmaceutical firms. Among the many stakeholders we identified are individual subjects (e.g., as patients or consumers), the organizations that create data (e.g., hospitals, physician practices, and third-party payers such as insurers), and IT vendors that supply health IT (HIT) systems or support PHI applications or platforms (such as Apple Health). Regional health industry organizations, government policymakers, medical researchers, and community health organizations are also key stakeholders, as are third-party health data aggregator firms and their clients.

Governance structures: The governance structure category includes sub-categories such as policies, regulations, organizational units, practices, and technologies that reflect and carry out governance goals. For instance, HIPAA regulations are policy structures 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). A Data Access Committee (organizational structure) may be established to evaluate and authorize legitimate requests for PHI for research purposes. Algorithms for personal de-identification (technology structures) can be used to remove PHI data from HIPAA oversight. An important governance structure is the steward role and responsibilities for implementing governance goals relative to a data domain or a specific data store. One commonly advocated steward role is the “neutral third party” organization that is charged with aggregating data from entities that create specified digitalized PHI, possibly harmonizing and assuring data quality and integrity, and enforcing policies for access to the data. Within organizations this role may be played by a data warehouse support unit in the IT or clinical informatics department, whereas for inter-organizations data exchange the steward role may be carried out by a neutral third-party entity, such as a health information exchange (HIE).

Governance goals: This category includes sub-categories representing the values-based objectives and intentions for governing a data domain using an array of governance structures. Protecting the privacy of individual health data is one of the most frequently discussed goals of PHI data governance. A related goal is to secure PHI data from unauthorized access, whether through a data breach or by stakeholders who are not authorized to view PHI (e.g., in the U.S., employers cannot view their employees’ health data.) Table 2 highlights some of the most commonly articulated goals we identified. It is important to realize that some governance goals are aimed at maximizing the value afforded by PHI data for specific stakeholders. Thus, the Rosenbaum (2010) quote cited earlier highlighted potential conflicts among stakeholders with an interest in PHI. For instance, a data aggregator may desire opacity on how they process and use putatively de-identified pharmacy data whereas third-party payers may desire transparency to understand how pharmaceutical firms use data to increase profits (Tanner 2016).

By comparing across these dimensions, we identified the following prototypical models of PHI data governance evident in various stakeholders’ discussions of PHI data uses and management.

Organizational PHI data governance: This model is well known in the IS field, as many businesses generate data through their transactional systems and aggregate data for enterprise-wide analysis (Khatri and Brown 2010; Kimball and Ross 2013). In healthcare settings, the governance structures typically include centralizing data stores via a data warehouse or data lake with the organization performing data stewardship functions (Elliott et al 2013). HIPAA regulation and legal rights to PHI data stores are externally sourced governance structures. The key stakeholders in this model are health providers, insurance companies and other third party payers that generate transactional PHI-related data. Primary values types from PHI data include organizational improvement and evidence-based pricing, whereas data governance goals include demonstrating compliance and maintaining trust, providing efficient data access, and data mining for innovation and quality patient care.

Inter-organizational PHI data governance: Healthcare services often require coordination across organizations, and hence the exchange of PHI data housed within organizational systems. Inter-organizational data governance models are aimed at facilitating PHI data exchange between organizations to realize the data’s value to enable health system improvements (e.g., reducing medical errors, eliminating redundancies in patient care). A key example is a health information exchange (HIE)

organization designed to enable secure PHI data transfer and sharing among healthcare service providers. As a (neutral) trusted third party, the HIE entity serves as the data steward to ensure data quality, integrity, security, and access. Other relevant governance structures include HIPAA regulations and data harmonization, aggregation, storage, and access technologies. Governance goals include efficient access (for authorized individuals), regulatory compliance, and a trusted mechanism for data exchange.

Code	Sub-codes	Description	Example
Compliance	Quality reporting, Data retention	Supporting activities and processes that enable regulatory compliance related to PHI	<i>“Besides cost imperatives, compliance issues related to legislation, such as HIPAA, SOX and Basel II, determine how organizations must deal with the lifecycle of data, its retention and archival”</i> (Khatri and Brown 2010, p. 151)
Efficiency	Cost reduction/ control, Systems integration, Risk reduction, standardization	Making PHI data available for various uses with less effort to compile data.	<i>“Information management and control is an undeniable healthcare imperative”</i> (Cohasset Associates 2014, p.1)
Innovation	Data mining, forecasting health issues	Providing access to PHR data to enable new insights or approaches.	<i>“Also, with the massive interest in big data across insurance, financial services and other markets, capital is flooding into this segment of the software business, which is leading to a relentless stream of innovation and pseudo-innovation”</i> (Liebman 2015, p. 15)
Data-driven engagement	Patient engagement, Giving voice, Sharing data analysis, Global justice	Providing stakeholders access to and control over their PHI data	<i>“It’s time to give the patients who are the sources of the data a voice in this discussion”</i> (Tanner 2016, ¶ 6)
Quality patient care	Improved patient care (outcomes), Data quality	Enabling analysis of healthcare outcomes with PHI access	<i>“Improve the quality and safety of patient care confirmed as a key driver by 95% of survey respondents”</i> (Cohasset Associates 2014, p.8)
Trust	Privacy, Security, Transparent data management, Opacity, Access, Access restrictions	Engender trust among stakeholders that PHI data are managed according to policies and stated goals.	<i>“Mechanisms to ensure trust are critical for addressing the privacy of individuals whose records are being accessed and individuals or organizations that share health data. Ensuring public trust is critical in systems that share sensitive information”</i> (Diamond et al 2009, p. 456)

Table 2. PHI Data Governance Goals

Public Good PHI data governance: This model promotes access to aggregated, (often) de-identified patient data to researchers to innovate or identify efficiencies in health systems delivery. This perspective acknowledges scientific advancements and innovations may occur through analysis of PHI data governed explicitly to do so (e.g. Hripcsak et al 2014; Rosenbaum 2010). For example, Ross et al (2014) and Holmes et al (2014) review the HMO Research Network’s (HMORN) virtual data warehouse (VDW) uniting 17 health care systems across the United States and providing data for thousands of research projects to date. Key stakeholders in this model include the public, medical and public health researchers (who mine the data), and healthcare service providers. In this model, the stewardship role varies among trusted third parties. Several governance structures identified in the analysis include technologies for data aggregation, organizational structures such as data access committees, PHI de-identification practices and techniques, and HIPAA regulations. Governance goals include efficient access for researchers and data mining for innovation in health services. However, the public interest in governing PHI data to realize value (e.g., health system improvement) can conflict with the interests of private organizations in profiting (or profiteering) from aggregated PHI (Rodwin 2009; Tanner 2016).

Community Health PHI data governance: This model focuses on grassroots efforts promoting community health. The value of data is population health management (and improvement) within a particular community (geospatial or online). For example, PatientsLikeMe is an online network of

patients with various medical conditions. Members connect with others with similar conditions to get advice and support (Bradley et al 2016). The platform also serves to conduct real-time research via clinical reporting tools that allow aggregate data to be analyzed to better improve health outcomes. Community members produce self-reported medical data, and the organization's research team uses these data for scientific and commercial medical research (Tempini 2014). The platform enables sharing of personal health data to improve one's own medical condition, help others with similar conditions, and transform medicine, enriching the good of society at large (Eichler et al 2016). As with the Inter-organizational and Public Good models, the stewardship role varies among trusted third parties, and governance structures may also vary greatly. For instance, PHI may or may not be de-identified, and community members may voluntarily cede HIPAA protection (at least for their own PHI data). Governance goals typically center on data-driven engagement, for instance addressing health disparities, engaging and giving voice (to patients), forecasting community health issues, and sharing data analysis.

Personal PHI data governance: This model is based on the belief that individuals will become more engaged in their own health when they have access to personal health data (Davidson et al 2015) and has gained momentum with growing prevalence of wearable devices and mobile software apps that monitor specific health conditions (e.g., glucose monitors) and gather data such as body temperature, heart rate, activity level, and user-input information about diet (sometimes termed the “quantified self”). With personal health record models, individuals are granted some steward roles over their own PHI housed in providers' clinical systems (Davidson et al 2015), whereas data generated by the individual through direct data entry or use of wearables are typically housed on an IT vendor's cloud-based infrastructure (e.g., Fitbit or Apple Health). The key stakeholders for this model are individuals and IT vendors, and they share stewardship along with any affiliates either party shares the data with. The key values underlying the data are personal health enhancement (for individuals) (Chiauzzi et al 2015) and monetizing PHI data (for the IT vendor). Governance structures include basic security, privacy policies and, in some cases, de-identification. Dolan (2014) also highlights the role of software and physical devices as a governance structure – the software, once installed on phone and connected to devices, can operate independently, or under the control of the user (consumer). Governance goals typically focus on efficient data access, data-focused engagement and trust in data stewardship practices (as stated in privacy policies).

Aggregator PHI data governance: In this model, data are either gathered from trace data activities performed on the Internet or via mobile phones, or purchased from healthcare sector or other organizations. Typically, these data are not personally identifiable or have been de-identified prior to sale, and thus are not covered by HIPAA. The types of value afforded by these data vary considerably. For instance, de-identified PHI data can be aggregated, combined with other data sources, and used to inform medical research (Sarasohn-Kahn 2014). Health data can also be purchased, aggregated and analyzed, and sold to other organizations, generally in the healthcare industry (Tanner 2016). Under this model, the aggregator firm and its clients are primary stakeholders, with others such as individuals indirectly influenced by the actions taken based on aggregated data (such as denying health insurance coverage). The stewardship role is managed by the aggregator firm and any associates it shares data with. The values realized through the data are monetization and evidence-based pricing. Governance structures include use of de-identified PHI, data aggregation technologies, and intellectual property claims. Governance goals include data mining, forecasting health issues, innovation, and intellectual property protection.

Discussion

Through a process of discourse analysis, we identified six prototypical models of PHI data governance evident in publicly available discussions of health data, their value and potential applications within the healthcare industry, and policy issues related to health data governance and stewardship. Figure 1 depicts the six models discussed above, and Tables 1 and 2 summarize key aspects of our findings on the types of value afforded by PHI and PHI governance goals. This analysis deepens our understanding of healthcare governance issues beyond discussions of privacy and security of personal health data and extends consideration of health data value beyond general claims about improvements to, or even transformation of, the healthcare sector (Blumenthal 2010). We suggest that the values and interests that underlie health data governance policies and structures are often depicted as unproblematic and widely shared. While it is true that many stakeholders in the healthcare industry value health data for their potential to inform health system improvements, population health, and evidence-based medicine, there are also potential

conflicts between the values and interests that underlie various models for PHI data governance, and therefore in the value that can be realized or derived from governing PHI data sources.

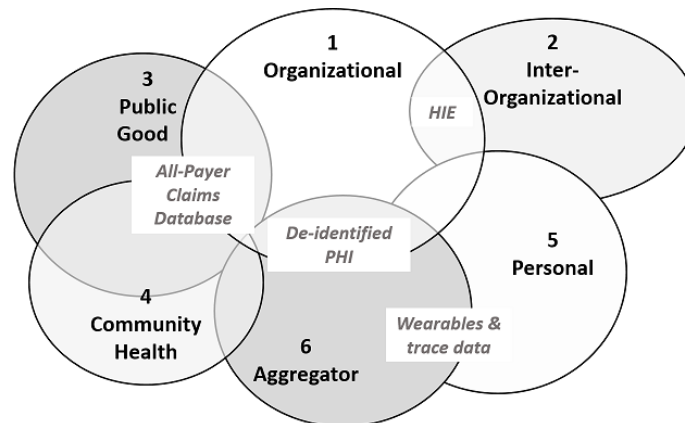


Figure 1. Data governance models in healthcare

For instance, both the Public Good and the Aggregator models are based on aggregating data across multiple sources to enhance the value realized from PHI through data mining and analysis. However, in the Public Good model, underlying values and interests that influence how PHI data are governed and which entities assume stewardship relate to realizing health system improvement through PHI data. Improvements entail promoting efficacy and cost-effectiveness of treatments and reduction of excess healthcare costs. In contrast, under the Aggregator model, profitability (or profiteering) by monetizing health data entails private, proprietary control over data and opacity in how data are actually governed. For example, the private firm IMS purchases pharmacy data from pharmacy chains, insurance companies, and healthcare providers, aggregates the data, and resells the data to pharmaceutical firms (Tanner 2016), allowing all parties to monetize health data they govern. Pharmaceutical firms may then use the data to target physicians' prescribing practices and justify patented drug prices with putative evidenced-based pricing – actions that may counter the public good of a more affordable health care system.

Even when there is some alignment in values, stakeholder interests may differ, and hence generate conflicting approaches to data governance. Such conflicts are evident when the Organizational model overlaps with other PHI governance models, and data “leaks” between governance structures (See Figure 1). Organizations generally seek to improve their own operational efficiencies and competitive positioning through PHI (in addition to providing better healthcare services). For health service providers, sharing information with competitors may not be in the organization's economic interests. These conflicts are evident with the U.S. health information exchange (HIE) approach to data governance. There has been little sustained success among HIEs, despite investment and normative pressure from the U.S. Federal government, and recent complaints about “information blocking” practices have been lodged (Office of the National Coordinator for Health Information Technology 2015b; Eden et al 2016). By comparison, intra-organizational data sharing within health provider networks have flourished (Downing et al 2017). Similarly, the failure of institutional arrangements for clinical data sharing between health providers and personal health record firms contributed to the failure of the Google Health personal health record, whereas organizationally controlled patient portals, which allows individuals to view their PHI within the provider's EHR system, have been more successful (Davidson et al, 2015). Thus, the governance goals of data-driven engagement are limited to the patient's engagement with the EHR owner/provider. Finally, conflict between the Organizational and other governance models is evident between insurance companies, which mine their medical claims data for organizational performance improvement and the Public Good or Community Health model. In the U.S., private insurers have resisted recent federally mandated efforts to construct regional “all-payers claims” databases of de-identified health claims records for medical, public health, and public policy research based on claimed governance structures of intellectual property and data ownership (Freedman et al 2016).

An emerging conflict between PHI data governance models is the use of personally generated health data from wearable devices as well as trace data generated by everyday activities that are IT-mediated, such as Internet browsing and mobile phone use. Advances in data analytics have allowed data scientists to

attribute health conditions to individuals based on these trace data. For instance, Microsoft researchers were able to use Internet search queries to identify individuals with pancreatic cancer prior to any disease diagnosis (Markoff 2016). Important governance questions are, for what purposes would such data be used and what governance structures and goals would determine these uses? Does governance of such data and the PHI profiles derived from them fall under the Aggregator model, and hence the stewardship of IT vendors or other private firms to monetize data as they see fit? Or, should data fall under the Personal model, so that individuals decide who can see their data and for what purposes? Today, because the individual is not aware what data have been collected or sold, or how big data algorithms will attribute health status, there is a lack of transparency and individuals have little voice in these decisions.

Concluding remarks

Few would question the value of PHI data to promote individual, population, and societal health or improvements in healthcare systems. Given the economic impact of the healthcare sector in the U.S. and globally, few would question the monetary value of these data either. Thus, there are many questions about how health data should be governed and for what goals. Here, we have argued that to better address questions of “how to” and “who” should govern health data stores, we need to consider which of the many potential values that PHI data represent and whose values and interests will be prioritized, or suppressed, by any specific governance model. Realizing value from governing PHI data is not a zero-sum game, as health data can be used and reused to support multiple forms of value creation. Neither is it necessarily a win-win situation, as the value(s) realized lie not in the data per se, but in the social and economic purposes and outcomes of data-inspired policies and practices. Also important are the resources committed to “extracting value” or mining PHI data. Will the “best and brightest” data scientists be dedicated to trolling “big health data” to improve the health and well being or to monetize PHI data?

To better understand and thus to more effectively manage (or regulate) the complex, emergent, and networked nature of personal health data requires that we understand who are the stakeholders in relation to personal health data and to identify their relevant interests (Kaanen and Carr 2009; Rosenbaum 2010). Our paper contributes to answering these questions by delving beneath important, high-level questions about health data privacy and security, and serving the public good through health data research, to consider how different arrangements of stakeholders, data value, governance structures and goals align around different models for PHI data governance. We identified and characterized six models and considered how underlying conflicts in stakeholder values and interests might disrupt socially desirable data governance and stewardship.

This analysis is a starting point to facilitate further discussion. One limitation of our discourse analysis is that the six models reflect the mixed private-public funding and service delivery model in the U.S. Examining PHI governance models in single-payer sectors, as in Canada, and unified-payer/-provider sectors, as in the U.K. and elsewhere, could reveal other models and other tensions, conflicts, and potential governance structures. A next step in our ongoing research project will be to solicit feedback from stakeholders involved in these various models, such as medical researchers and private insurers among others. Case studies will also be useful to further detail governance models and to understand the processes through which specific instantiations of a model are initiated and evolve. Finally, our focus has been on the policy, social, and economic factors that influence data governance. Without question, health data are complex and health IT systems lack standardization, which presents significant challenges to data stewardship in terms of harmonizing and aggregating disparate sources of PHI. Blending a technical focus on the “how to” questions with an appreciation of social and organizational issues will better inform policy makers on how to promote PHI data governance structures and stewardship entities.

REFERENCES

- Belanger, F. and Xu, H. 2005. “Editorial: The Role of Information Systems Research in Shaping the Future of Information Privacy,” *Information Systems Journal* (25), pp. 573-778.
- Blumenthal, D. 2010. “Launching HITECH,” *New England Journal of Medicine* (2010:62), pp. 382-385.
- Bradley, M., et al. 2016. “Patients’ Motivations and Interest in Research: Characteristics of Volunteers for Patient-led Projects on PatientsLikeMe,” *Research Involvement and Engagement* (2:1), pp. 33.
- Chiauzzi, E. 2015. “Patient-centered Activity Monitoring in the Self-management of Chronic Health Conditions,” *BMC Medicine* (13:77), pp. 1-6.

- Cohasset Associates. 2014. "A Call to Adopt Information Governance Practices" AHIMA 2014 Information Governance in Healthcare: Benchmarking White Paper (available at <https://goo.gl/Dsypho>).
- Davenport, T. H., et al. 2012. "How 'Big Data' is Different," *MIT Sloan Management Review* (54:1), pp. 43-46.
- Davidson, E., et al. 2015. "Drift and Shift in the Organizing Vision Career for Personal Health Records: An Investigation of Innovation Discourse Dynamics," *Information and Organization* (25), pp. 191-221.
- Diamond, C. C., et al. 2009. "Collecting and Sharing Data for Population Health: A New Paradigm," *Health Affairs* (28:2), pp.454-466.
- Dolan, B. 2014. "In-Depth: What Apple's Health App Tracks and What it Forgot to Include," *MobiHealthNews* (available at <https://goo.gl/KGMGKz>).
- Downing, N. L., et al. 2017. "Health Information Exchange Policies of 11 Diverse Health Systems and the Associated Impact on Volume of Exchange," *Journal of the American Medical Informatics Association* (24:1), pp. 113-122.
- Eden, K. B., et al. 2016. "Barriers and Facilitators to Exchanging Health Information: A Systematic Review," *International journal of medical informatics* (88), pp. 44-51.
- Eichler, G., et al. 2016. "Exploring Concordance of Patient-reported Information on PatientsLikeMe and Medical Claims Data at the Patient Level," *Journal of Medical Internet Research* (18:5), doi: 10.2196/jmir.5130.
- Elgarah, W., et al. 2005. "Data Exchange in Interorganizational Relationships: Review through Multiple Conceptual Lenses," *ACM SIGMIS Database* (36:1), pp. 8-29.
- Elliott, T., et al. 2013. "Data Warehouse Governance Programs in Healthcare Settings: A Literature Review and a Call to Action," *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)* (1:1), p. 15-21.
- Freedman, J. D., et al. 2016. "All-Payer Claims Databases—Uses and Expanded Prospects after Gobeille," *New England Journal of Medicine* (375:23), pp. 2215-2217.
- George, G., et al. 2014. "Big Data and Management," *Academy of Management Journal* (57:2), pp. 321-326.
- Holmes, J., et al. 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), pp. 730-736.
- Hripcsak, G., et al. 2014. "Health Data Use, Stewardship, and Governance: Ongoing Gaps and Challenges: A Report from AMIA's 2012 Health Policy Meeting," *JAMIA* (21: 2), pp. 204–211.
- Inmon, W. H. 1996. "The Data Warehouse and Data Mining," *Communications of the ACM* (39:11), pp. 49-51.
- Jin, D. Y. and Feenberg, A. 2015. "Commodity and Community in Social Networking: Marx and the Monetization of User-generated Content," *The Information Society* (31:1), pp. 52-60.
- Kanaan, S. and Carr, J. 2009. "Health Data Stewardship: What, Why, Who, How: An NCVHS Primer" (available at <https://goo.gl/oP5u5T>).
- Khatri, V. and Brown, C. V. 2010. "Designing Data Governance," *Communications of the ACM* (53:1), pp. 148-152.
- Kimball, R. and Ross, M. 2013. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Hoboken, NJ: John Wiley & Sons.
- Libert, T. 2015. "Privacy Implications of Health Information Seeking on the Web," *Communications of the ACM* (58:3), pp. 68-77.
- Liebman, L. 2015. "Turning Data into Dollars," *Insurance Networking News* (Feb.-Mar.), pp. 10-15.
- Lycett, M. 2013 "Datafication": Making Sense of (Big) Data in a Complex World," *European Journal of Information Systems* (22:4), pp. 381-386.
- Markoff, J. 2016. "Microsoft Finds Cancer Clues in Search Queries," *New York Times* (available at <https://goo.gl/2D6njA>).
- Murdoch, T. B. and Detsky, A. S. 2013. "The Inevitable Application of Big Data to Health Care," *JAMA* (309:13), pp. 1351-1352.
- Office of the National Coordinator for Health Information Technology. 2015a. "Health IT Policy Committee, Privacy & Security Workgroup. Final Transcript, January 12, 2015" (available at <https://goo.gl/oX9Vtl>).
- Office of the National Coordinator for Health Information Technology. 2015b. "Report on Information Blocking" (Available at <https://goo.gl/43YsP2>).
- Rodwin, M. 2009. "The Case for Public Ownership of Patient Data", *JAMA* (302:1), pp. 86-88.
- Rosenbaum, S. 2010. "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access", *Health Services Research* (45:5p2), pp. 1442-1455.
- Ross, T., et al. 2014. "The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration," *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)* (2:1).
- Sarasohn-Kahn, J. 2014. "Here's Looking at You: How Personal Health Information is being Tracked and Used", California Healthcare Foundation (available at <https://goo.gl/YCaEFM>).
- Tanner, A. 2016. "This Little-Known Firm Is Getting Rich Off Your Medical Data," *Fortune* (available at <https://goo.gl/B15YiJ>).
- Tempini, N. 2015. "Governing PatientsLikeMe: Information Production and Research through an Open, Distributed, and Data-based Social Media Network," *The Information Society* (31:2), pp. 193-211.
- Vayena, E., et al. 2015. "Ethical Challenges of Big Data in Public Health," *PLoS Computational Biology* (11:2), pp. 1-7.
- Zuboff, S. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology* (30:1), pp. 75-89.