11 Obstacles on the Road to Corporate Data Responsibility

M. Lynne Markus

No issue is more closely identified with big data in public awareness than privacy (Ohm and Peppet, this volume). Many proposals for improving privacy protection focus on national security and law enforcement agencies, or on data brokers and search firms (e.g., Google; Burdon and Andrejevic, this volume). To the extent, however, that every organization today collects, stores, manages, and processes information, every organization has responsibilities for privacy and data protection. The question addressed in this chapter is whether and to what degree corporations (e.g., banks and insurance companies) and other organizations (e.g., not-for-profit health care providers and educational institutions) are able to behave responsibly with data. The chapter’s conclusion is that corporate data responsibility is likely to remain elusive for reasons suggested by the title of this volume. The nonmonolithic context of organizational data use, data governance, and data management creates obstacles on the path to corporate data responsibility.

The observations in this chapter come from two sources. The first was a workshop funded by the National Science Foundation to create an agenda for research on the social, economic, and workforce implications of big data and analytics (Markus and Topi 2015). The second is an ongoing study of data management practices in large business, nonprofit, and government organizations conducted by MIT’s Center for Information Systems Research with Barbara H. Wixom as the principal investigator. The chapter first briefly describes organizations other than national security agendas and data brokers as information businesses with data protection and management responsibilities. It then describes the multifarious context of organizational data use, protection, and management, and shows how this multiplicity creates obstacles for the effective discharge of corporate data responsibility.

Every Organization Is in the Information Business

Big data is not solely the province of law enforcement agencies, technology and data companies, and search firms. Organizations of all kinds collect, store, manage, and use data about their operations, employees, and customers.
Retail banks offer a concrete example. These organizations maintain personal data on employees (e.g., addresses and identification numbers, but also data about compensation and benefits, families, health, and work performance). They collect and manage personal data about customers’ use of various banking products, including checking and savings accounts, secured loans for homes and cars, unsecured loans like credit cards, investment advice, investment products, and insurance. Retail banks buy personal data about prospective customers along with customer sentiment toward banking institutions and services. Banks may also sell customers’ personal data to data aggregators and brokers.

Financial organizations look to the analysis of personal information to help them combat fraud, cross-sell financial products to existing customers, assess the creditworthiness of potential new customers, design new financial products and marketing campaigns, and provide banking services across multiple technology platforms, including branches, kiosks, personal computers, and mobile devices. They also look to big data analytics for guidance on future hiring decisions and lowering employee health insurance costs.

A challenge these organizations face is how to gain the many promised benefits of big data use without compromising personal information privacy. Big data introduces new threats to personal information privacy through the process of matching data elements found in disparate sources, allowing for the reidentification of people in “anonymized” data sets (Ohm 2010) and the virtual “manufacturing” (Pasquale 2015) of personal data from unlikely material. An illustration of this process is the use of facial recognition technology to identify unnamed people in photographs posted on social media sites. It is not far-fetched to conclude that big data allows any item of personal data to reveal all data about a person (Ohm and Peppet, this volume).

The enhanced risks of big data use demand enhanced organizational data protection practices. Lawyers, public policy experts, and authorities on technology, business, and social responsibility have called for changes in organizational data protection behaviors. Some proposals focus on legal and regulatory changes. Examples include legislation about employers’ and employees’ rights of ownership in employees’ personal digital devices and online accounts (Mehta 2014; Park 2014), due process for individuals harmed by big data (Crawford and Schultz 2014), experimentation on the users of online services (Grimmelmann 2015), the need for “chain-link confidentiality” when personal information is disclosed to third parties (Hartzog 2012), and regulatory reporting (Culnan 2011). Other proposals focus on self-regulatory organizational practices. For instance, organizations are urged to reduce their reliance on customer notification and choice, and increase their emphasis on data stewardship and risk assessment (Cate, this volume). They are advised to conduct value alignment exercises (Davis 2012), and develop or adapt codes of ethical conduct about data use (Bennett and Mulligan 2012). Organizations are encouraged to set up data-sharing agreements as well as review boards for data-sharing requests, analytics projects, and experiments on users (Allen et al. 2014; Calo 2013; Grimmelmann 2015; Lin et al. 2014). They are exhorted to train
developers in methods for designing privacy into websites and apps (Friedman, Kahn, and Borning 2006; Langheinrich 2001).

The proliferation of proposals for changes in the data-handling practices of all organizations raises a number of challenging questions. Does effective data protection require organizations to implement all the recommended practices? How likely are organizations to do so, and even if they do, how likely are they to succeed in protecting data effectively? Answering such questions requires a sound understanding of the organizational context in which data are collected, used, and protected. This chapter, accordingly, examines three dimensions of the organizational data use context: differentiation inside organizations and across the organizations that comprise data “supply chains”; multiplicity, conflicts, and gaps in data protection laws, contracts, frameworks, standards, occupational specialties, and formally mandated protection roles; and diversity in organizations’ data management priorities and practices. It also shows how the nonmonolithic character of each dimension creates challenges for effective organizational data protection.

The Organizational Context of Data Use Is Not a Monolith

To understand how—and how effectively—personal data are protected, it is first necessary to understand that organizational data use involves differentiated, not monolithic, actors. Organizational data use involves many individuals, groups, and even distinct legal entities with different goals, authorities, accountabilities, responsibilities, and values. As a rule, these units do not undertake data protection in a coordinated or consistent manner. This can undermine effective data protection, which requires coordination of all parties toward a common goal. Below I consider four types of differentiation in the organizational context of data use that have important implications for data protection: differences among the organizations in data supply chains, differences among subunits inside organizations, differences between organizational leaders and employees, and differences between clients and data scientists.

The Organizations in Data Supply Chains

The use of big data often involves two or more organizations. The simplest cases involve only two organizations. For instance, an organization like a bank or insurance company might purchase data about prospective customers from a data aggregator for marketing purposes. In practice, data supply chains (Martin 2015; Washington 2014) frequently involve more than two organizations. An insurance company, for instance, that sells its policies through independent insurance brokerages may, in order to help brokers improve their sales performance, provide brokers with access to personal data about prospective customers that the insurance company purchased from a data aggregator. Here, the data supply chain involves the insurance company, the data aggregator, and numerous insurance brokerages, each of which must do its part for data protection to be effective.
Data supply chains can be even more complicated, because organizations may contract with data brokers as well as technology products and services firms (e.g., analytics service providers), which in turn contract with other technology products and services firms for support services, such as data storage, data backup and disaster recovery services, data security services, and so forth. Any of the companies in these chains can create breaches in data protection. For example, organizations have lost access to personal data about their customers and employees when a technology company they contracted with acquired the data storage services of a third company that was temporarily or permanently shut down by a legal action or bankruptcy (Armbrust et al. 2010). The organizations that lost data might not even have known that the vulnerable data storage provider was involved in the supply chain (Sullivan, Heiser, and McMillan 2015).

The companies that touch the personal data that organizations buy, collect, analyze, or sell are highly diverse. They include data brokers (e.g., Acxiom), social media companies (e.g., Facebook), "cloud computing" service providers (e.g., Amazon Web Services), enterprise applications software providers (e.g., SAP), "business process outsourcing" companies (e.g., Accenture), consulting and market research firms, and accounting and audit firms. The data protection policies and practices of these companies can differ widely from each other and from the organizations that do business with them. First, as autonomous legal entities, each company can set its own policies within regulatory constraints. Second, companies incorporated in different countries and those that operate in different economic sectors may be subject to different rules. Third, some companies make money from the sale of data and data-related services; they design their terms and conditions of use with their best advantage in mind. For instance, the contracts of data storage providers do not always afford strong performance guarantees to their organizational customers (Kemp Little LLP 2011; Venter and Whitley 2012), offer redress for customers in the event of data loss (Kemp Little LLP 2011), or affirm the customers' ownership of stored data (Trappler 2012). The result is a patchwork of data protection agreements that do not cover the entire data supply chain in a coherent and coordinated manner.

Any of the partners in an organization's data supply chains can be weak links in personal data protection. It is not surprising, therefore, that experts identify external business partners and data services providers as key sources of privacy risk for organizations (Casper 2014). In addition, when lapses in data protection occur, "it can be challenging for organizations, and regulators, to determine which entities are responsible for legal compliance, and what the scope is of their individual responsibilities" (Carey 2015, 307).

**Subunits Inside Organizations**

Every medium- and large-sized organization has internal structural divisions that can lead to differences in data protection practices and gaps in regulatory compliance. Organizations like banks with multiple product lines (e.g., checking and savings accounts, credit cards, and mortgages) and/or markets (e.g., North America versus Europe, or different states or
provinces within countries) are internally complex. They are typically organized into business units or groups of business units, each responsible for a particular combination of products and/or customer segments. These units operate somewhat independently of each other, and may have considerable autonomy to set their own policies and procedures in certain areas of activity (Markus 2014).

One implication of this internal differentiation is that an organization like a bank or insurance company does not always manage and protect data as an integrated whole. Several business units in a large financial services company, for example, might each contract separately with the same external data or services provider, and each of these contracts may specify different data protection terms.

It is critical to the success of data protection to embed decision making about data protection throughout business unit structures (Bamberger and Mulligan 2011a; Johnson and Goetz 2007). Unfortunately, however, business unit leaders do not always share with executives at headquarters a high level of concern with or commitment to data privacy and security (Bamberger and Mulligan 2011a; Johnson and Goetz 2007). Thus, it can be highly consequential for data protection whether organizations choose to centralize data analytics activities at corporate headquarters or decentralize them in business units. A common practice today is to locate analytics activities within multiple business and functional units (e.g., marketing) rather than creating a “center of excellence” at headquarters (where presumably the use and protection of data would be easier to control) (Pearson and Wegener 2013). Even when organizations have analytics units at headquarters, these central units may not have the organizational authority or clout to set and enforce compliance by business units. Given this, data-related policies may vary in nature and quality of enforcement across the units of a complex enterprise.

In addition, organizations that operate in multiple countries are even less likely to protect data in a consistent fashion across all units. The major organizational units of multinational enterprises are legally incorporated in various countries, where they are subject to different laws and regulations. Given cross-country differences in data protection laws (discussed more fully below), it would not be surprising to find considerable variation and gaps in data protection across the units of some multinational enterprises.

Organizational Managers and Employees
The most obvious example of the nonmonolithic character of the organizational data use context lies in the distinction between organizational policy makers and the rank-and-file employees. Executives can set policies, and business units can embrace them, but the compliance of many employees is necessary for data protection success (Hu et al. 2012). Accounts of major security and privacy breaches frequently highlight employee failure to conform to good organizational policies. For instance, employees may fail to install newer software versions containing security patches, share passwords or write them down in places where they can be copied, or leave devices containing sensitive data unprotected in homes or cars. Much
of the literature on corporate security and privacy protection focuses on the development of employee-oriented policies, such as acceptable (technology and data) use policies (Huckabee and Kolb 2014) and the challenges of motivating employees to comply with them (Hu et al. 2012; Puhakainen and Siponen 2010).

To improve individual compliance, security experts design education and training programs, monitor employee behavior, conduct diagnostic phishing exercises, and engage in ethical hacking to gauge employee susceptibility to malicious hackers’ social engineering ploys. Data loss protection technologies are installed to ensure that employees do not download or e-mail protected data. But most experts agree that technology alone cannot protect sensitive data (Hill 2009), just as good organizational data protection policies cannot succeed on their own. Effective data protection requires organizational leaders to consistently enforce employee compliance with well-designed data protection policies and technological controls.

Managerial Clients and Data Scientists
As noted above, data protection effectiveness within organizations depends on such factors as where in organizational structures data protection policies are set and analytics activities are located; the nature of the relationships between policy makers in headquarters and business units heads; and the quality of managerial enforcement of employee compliance with rules and regulations. Another important factor is the values of organizational members along with the ethical choices they make about the use and protection of data. Here I focus particularly on the values and ethical orientations of organizational data scientists—specialists who look for useful patterns by matching and analyzing multiple data sets—and their managerial clients—executive problem owners who commission and oversee data science projects.

Legal scholars (Bamberger and Mulligan 2011b) and business and technology experts (Culnan and Williams 2009; Davis 2012) argue that data protection in an era of big data is not just matter of legal compliance; organizations should articulate their ethical principles and instill appropriate values throughout their organizational cultures. This prescription can be challenging to fill because of the many influences on individuals’ values and beliefs, including national culture, educational preparation, occupational specialty, and specific organizational location and role.

Europeans, for example, are believed to value privacy differently than North Americans (Seddon and Currie 2013), and people in high-technology occupations and organizations are supposedly more prone than others to the pro-innovation bias that is sometimes called “solutionism” (Morozov 2013). A neutral definition of solutionism is “the belief that every problem has a solution based in technology” (Maxwell 2014). A more critical one is the urge to fix with technology “problems that are not problems at all” (Morozov 2013, 6). In general, solutionists minimize the potential of technological innovations to create negative side effects and label as “neo-Luddites” people who raise concerns (Atkinson 2015; Trader 2014;
"What Is a Neo-Luddite?" n.d.). By contrast, other organizational members, including some executives, have more cautious views about the benefits and risks of big data analytics. This may result from uncertainty about benefits, fear of breaking the law or risking customer backlash, or discomfort from lack of personal expertise.

Education and professional socialization play an important role in shaping the values and ethical stances of data scientists and their clients. So it is instructive to learn what these groups know about data protection and the ethical dilemmas of big data use. The occupational specialty of data scientist is still emerging (Booz Allen Hamilton 2015). Well-publicized job opportunities for data scientists have encouraged the emergence and growth of educational programs (West and Portenoy, this volume). Several academic disciplines and professional associations have laid claim to the territory of data scientist education. Although some programs are interdisciplinary, most data science programs are associated with computer science or engineering departments or schools; departments of mathematics and statistics (often located in arts and sciences schools); or departments of information systems and operations research/management science (usually, in the United States, located in business schools) (West and Portenoy, this volume). Naturally, the curricula of data science programs differ depending on disciplinary location. Yet the majority of programs today focus almost exclusively on either the building of data science tools and infrastructure or the use of data science techniques (e.g., machine learning) to solve research or practical problems. A few programs offer limited opportunities for students to learn about data protection laws, ethical dilemmas in big data use, and how to confront ethical concerns constructively. In the sections below, I consider curricula and aspects of professional practice, such as codes of conduct, for each of the three main disciplines in which data scientists are prepared.

Computer Science The field of computer science has a long-standing commitment to ethics research, practice, and education. The ethics of computing is an established research area within computer science (Bynum 2015). A number of computer and information science programs offer courses on ethical theory and its application (Fleischmann and Wallace 2005), or approaches for building positive values like privacy into the design of systems and devices (Friedman and Hendry 2012). The ACM Code of Ethics and Professional Conduct (1992), by a leading computer science association, exhorts members to contribute to society (section 1.1) and human well-being and avoid doing harm (section 1.2)—an injunction against solutionism. Despite these efforts, educators find "an 'ethics gap'" in engineering education, noting that professionals are rarely prepared to deal well with the many ethical challenges in contemporary practice (McGinn 2015). Among those challenges are pressures from clients and superiors to do things that a professional believes is wrong.

Mathematics and Statistics Statisticians are well aware of ethical challenges surrounding the analysis of data. For example, the American Statistical Association enjoin its members,
through its “Ethical Guidelines for Statistical Practice” (Committee on Professional Ethics 1999, sec. I.C), to avoid working toward a predetermined result, which members’ clients might pressure them to do. The guidelines also contain a detailed discussion about the protection of human subjects of research “including census or survey respondents and persons and organizations supplying data from administrative records, as well as subjects of physically or psychologically invasive research” (ibid., sec. II.D); this section explicitly mentions privacy and data confidentiality. The American Statistical Association’s (2014, 13) “Curriculum Guidelines for Undergraduate Programs in Statistical Science” urges educators “to integrate training in professional conduct and ethics,” and some guidance is provided in a companion white paper (Cohen 2014) on how that might be done. For instance, educators should counsel students not to “imply protection of privacy and confidentiality for legal processes of discovery unless explicitly authorized to do so” (ibid., 3).

Similarly, the Data Science Association’s (n.d.) “Code of Conduct” offers a detailed definition of confidential information that clearly includes the personal data of employees and customers. Among other things, the code enjoins professionals to “do no harm”—that is, to avoid negative side effects—a warning against solutionism. The Data Science Association’s code also states that the professional is to abide by the client’s objectives, and acknowledges the dilemmas that doing so may entail. It nevertheless stops short of clarifying the data scientists’ obligations in cases where the client wants the data scientist to do something that is legal, but ethically questionable:

A data scientist shall not counsel a client to engage, or assist a client, in conduct that the data scientist knows is criminal or fraudulent, but a data scientist may discuss the consequences of any proposed course of conduct with a client and may counsel or assist a client to make a good faith effort to determine the validity, scope, meaning or application of the data science provided. (Data Science Association, n.d., rule 3b; emphasis added)

This clause implies that data scientists are not obliged to discuss with their clients the ethics of questionable data uses.

The sources cited above show that mathematicians and statisticians are generally sensitive to the ethical dilemmas and professional responsibilities related to data use and protection. At the same time, one wonders how substantively these issues are treated in data science education. A recently published data science text (Provost and Fawcett 2013) suggests that data scientists in training may be on their own when it comes to dealing with issues like privacy. The authors (ibid.) refer to Daniel Solove’s (2006) eighty-page article “A Taxonomy of Privacy” and Helen Nissenbaum’s (2010) three-hundred-page book Privacy in Context. They conclude:

We bring this up to emphasize that privacy concerns are not some easy-to-understand or easy-to-deal-with issues that can be quickly dispatched, or even written about well as a section or chapter of a data science book. If you are either a data scientist or a business stakeholder in data science projects, you should care about privacy concerns, and you will need to invest serious time in thinking carefully about them. (Provost and Fawcett 2013, chap. 14; emphasis added)
This comment raises the question: If data scientists and business stakeholders do not get the
time to think carefully about these issues during their educations, will they get that time
on the job?

**Business Education** Business schools also educate data scientists in departments of manage-
ment science and information systems. They educate many future clients of data science
projects as well, and their education is relevant, too, since big data analytics has increased
the demand not only for data scientists but also for "data-savvy managers and analysts who
have the skills to be effective consumers of big data insights—i.e., capable of posing the
right questions for analysis, interpreting and challenging the results, and making appropriate
decisions" (Manyika et al. 2011, 87).

Business ethics is a well-established field of research and education (Christensen et al.
2007; Tsalikis and Fritzsché 1989). Many, but by no means all (Floyd et al. 2013), business
schools offer courses in business ethics. Yet the content of business ethics courses does not
usually cover specialized topics like the ethical dilemmas of data use and protection. Fur-
thermore, ethics educators in business schools point out that managers in training, both future
data scientists and their clients, generally lack the training and skills needed to speak up
when they believe something is wrong (Gentile 2010).

The future data scientists prepared by business school analytics programs may get no
additional education in data protection or the ethics of data use. Program descriptions
typically make little or no mention of data privacy and security (Pearson and Wegener 2013),
and few programs dedicate a full course to these topics. Of nontechnical subjects, data
science programs in business schools emphasize "consulting skills" to understand business
requirements.

Attention to ethics and data protection is also limited in the certification programs offered
by industry groups. INFORMS, a professional society with origins in management science,
offers a certification program (CAP) for analytics professionals. The CAP "Code of Ethics/
Conduct" (INFORMS, n.d.) admonishes professionals to behave professionally, follow all
applicable laws, resist pressures to produce analyses biased toward a particular result, and
avoid the unauthorized or illegal use of intellectual property. The single mention of privacy
occurs in the context of human subjects protection, which is familiar in academic and health
care research contexts, but much less known in the general corporate world. Privacy gets a few
mentions in the materials provided to help people prepare for the CAP certification exam,
but the sample test questions do not cover data protection laws or ethical dilemmas related
to big data.

In short, whether they are solutionists or neo-Luddites, neither budding data scientists
nor clients in training appear well prepared to address the ethical gray areas of data use along
with the challenges of personal data protection. In addition, data scientists may have gaps
in their preparation, regardless of the academic discipline or professional association that
organized it.
Summary
No single individual, group, or organization (or technology) can alone protect personal data well. Effective data protection requires the concerted action and compliance of many different parties. Unfortunately, the many divisions within and across the organizations in data supply chains create considerable challenges for effective data protection. Effective data protection is hindered by wide differences—in goals, incentives, values, education, knowledge, and informal norms of behavior—among the many individuals, groups, and organizations involved in organizational data use.

The Governance of Data Protection Is Not a Monolith
As noted above, all organizations use personal data and therefore must protect it. Organizations’ data protection activities are governed by laws, regulations, and contracts, and guided by numerous standards, frameworks of voluntary guidance, and specialized bodies of knowledge and expertise. Regrettably, the governance of data protection is far from monolithic; it contains numerous overlaps, conflicts, and gaps. This diverse and complex governance regime, intended to support data protection, also works to undermine it.

Data Protection Laws, Frameworks, Standards, and Professional Specialties
The European Union has a comprehensive data protection act that spells out the obligations of data controllers and the rights of data subjects (Carey 2015). The United States does not have a comprehensive approach to data protection. Instead, the privacy and security obligations of organizations and the rights of individuals in the United States are spelled out in laws and regulations specific to particular sectors, such as public (listed) companies, health care providers, financial services firms, and organizations that market to children. Examples of the US regulations addressing aspects of data security and/or privacy include Sarbanes-Oxley (applicable to public companies and concerned with financial reporting and information security), HIPAA (health care), Graham-Leach-Bliley (financial services), and COPPA (children’s online privacy). Some organizations are also covered by rules about experiments on people (Grimmelmann 2015). Because of the fragmentation of US data protection laws, most organizations doing business in the United States have to satisfy the data protection demands of multiple regulators. Of course, multinational enterprises have to accommodate several different data protection regimes.

The multiplicity of data protection laws is supported, and possibly complicated, by a variety of security and privacy standards as well as voluntary guidance frameworks. Illustrations include the ISO 17002 security standard, the National Institute of Standards and Technology’s cybersecurity framework, generally accepted privacy principles, and fair information practice principles. Various professional societies provide related education and certifications, including the Information Systems Audit and Control Association and the International Association of Privacy Professionals.
Data Management Regulations
The legal and practical complexity of data protection laws is only part of the story, however, because a large number of additional rules and regulations apply to how organizations manage the data they hold, including but not limited to the personal data of employees and customers. In financial services, for example, data-related regulations have been enacted to prevent, detect, and punish crimes such as insider trading, market and rate manipulation, money laundering, and discrimination on the basis of race or gender in lending and insurance. In addition, regulations aimed at ensuring financial stability, such as Basel and Dodd-Frank, require organizations to build data analytic models and then report the results of these analyses on a regular basis. (Banks may find it necessary to purchase costly data from external data providers such as Bloomberg to perform these analyses well.) Banks and other financial services firms are subject to prosecution and large fines when they cannot provide regulators with accurate data in the form of reports and analyses such as liquidity “stress tests” (Effinger 2015; Rexrode 2015). Public companies are also subject to rules about accounting for and reporting on their financial management; these rules cover data security, too. Laws and corporate concerns about potential litigation dictate practices of record retention and destruction. Finally, financial services firms may voluntarily participate in data standardization initiatives designed to improve regulatory compliance and the efficiency of transactions between organizations. Examples include Mortgage Industry Standards Maintenance Organization data and process standards in mortgage lending (Markus et al. 2006), and legal entity identifier naming conventions in financial reporting.

Thus, organizations face many demands and constraints on their data-handling practices in addition to those intended to protect the security and privacy of personal data. In a recent description of “How We Do Business,” JP Morgan Chase and Co. (2014, 77) reported having “more than 250 banking, securities, and commodities regulators overseeing our business globally.” Given the large number of data-related regulations to which individual organizations may be subject, it is not surprising that corporate executives view “regulation changes and heightened regulatory scrutiny” as their major current business risk (North Carolina State University’s ERM Initiative and Protiviti 2015, 5). As one would expect, financial industry compliance professionals are now in high demand (Effinger 2015).

Too Many Rules
Each individual data management and protection rule can be costly to implement; implementing many rules separately can create a crushing compliance burden. Efficiency considerations suggest the value of harmonizing the rules and adopting an integrated approach to compliance (Proctor, Wheeler, and Pratap 2015). Not surprisingly, an area of professional practice and associated body of knowledge, known as governance risk and compliance (GRC), has emerged to address this challenge (Hill 2009). Certification of GRC professionals is available, and technology vendors offer software and data products designed to support GRC programs (ibid.).
Implementing a GRC program can generate significant savings for organizations. Consulting reports describe savings on the order of 30 percent for companies that centralize, harmonize, and offshore their compliance and control functions (EYGM Limited 2014). Fiserv, a global provider of information management and electronic commerce systems for the financial services industry, explained its benefits from adopting GRC technology:

The company estimates that to produce the type of detailed risk profile it gets from the software over a three-month period now, it would previously have taken about six months using Fiserv’s old manual process. The older method would also have required seven to 10 more staff members and would have cost Fiserv an additional half-million dollars. (Violino 2012, para. 30)

As useful as they are, however, GRC programs and technologies cannot completely eliminate the challenges of compliance with data protection and management regulations. Some regulatory conflicts cannot easily be harmonized. European data protection regulations require organizations to minimize the collection and analysis of personal information in order to preserve privacy (Carey 2015), whereas a new financial regulation, BCBS 239, requires them to actively manage and analyze data in order to reduce risk. Among other things, BCBS 239 requires companies to keep complete as well as accurate records (Deloitte, n.d.). How financial organizations can reconcile these conflicting demands is not entirely clear.

GRC programs also cannot cope with several noteworthy gaps in data protection regulations. In the United States, for instance, health data in government and providers’ medical records are strictly protected, but health data available from other sources are not. Data scientists have been able to mine information from pharmacy, social media, and data broker sources to identify potential candidates for clinical trials, thereby circumventing the onerous controls on searches of medical records (Walker 2013). Another interesting legal gap is created by autonomous systems that operate without “continuous human intervention” (Dahiyat 2007, 387). Examples include self-driving cars and algorithms that automatically determine credit worthiness and mortgage eligibility (Markus et al. 2008). Increasingly, autonomous systems are being designed to “learn” and evolve on the basis of their “experience” (see DeDeo, this volume). Such algorithms raise vexing questions of patent law (see Abbott, this volume), contract law (Dahiyat 2007), liability law (Dahiyat 2010; Elish and Hwang 2015), and fairness.

Too Many Chiefs
Data protection laws and regulations require organizations to designate executives to be held accountable for compliance. This has led to a proliferation of high-ranking organizational positions, each of which has a role in managing and protecting data. The chief privacy officer (Bamberger and Mulligan 2011a) and chief information security officer (Johnson and Mulvey 1995) are obvious instances. Other data-oriented chief roles include, on the compliance side, the chief financial officer, chief legal officer (general counsel), chief compliance officer, and
chief risk officer, and on the technical side, chief information (or technical) officer, chief data officer, chief digital officer, chief information security officer, and chief analytics officer. Many organizations have several such chiefs.

With all these chiefs, an obvious question is: How well do they collaborate in data protection? Articles in the business press suggest that organizations are uncertain about where to locate these positions on the organizational chart, what their reporting relationships should be (should one chief report to another?), and how to divide and coordinate their responsibilities for various data management and protection activities. The importance of good working relationships among the various chiefs has been mentioned (Wheatman and Akdshay 2015), and a few articles hint at antagonisms, turf conflicts, and the possibility of gaps in data management and protection resulting from unclear divisions of labor. When Bank of America, for example, was recently taken to task by regulators for deficiencies in its stress testing, a legally mandated financial risk analysis and reporting activity that is heavily data intensive, executives apparently disagreed over responsibility (Rexrode 2015, para. 10).

In short, the large and growing numbers of chiefs having a prescribed or plausible role in data management and protection make it challenging “to determine [who is] responsible for legal compliance, and what is the scope of their individual responsibilities” (Carey 2015, 307) inside organizations as well as among the organizations in data supply chains.

Data Contracts
Data protection and management are governed not only by laws and regulations but also by private data agreements. Familiar examples of organization-to-individual data contracts are the published privacy policies of health care and financial services providers and the “terms and conditions” of websites and apps (see Cate, this volume). Contracts containing data-related “terms of use” provisions also govern the relationships between organizations and their data brokers, cloud services providers, technology vendors, and business process outsourcers. Public information on the provisions of these agreements is hard to come by because they are typically regarded as confidential. Nevertheless, it seems reasonable to assume that they vary in the quality and enforceability of their data protection provisions. For instance, contracts may not “bind third parties in the onward transfer of personal information” (Hartzog 2012, 689).

Furthermore, the nature of data contract provisions may make it difficult for organizations to follow the recommended practices of responsible data use. Kate Crawford and Jason Schultz (2014) advocate “procedural data due process” as a way to address the harms of algorithmic decision making, such as discrimination in mortgage lending or insurance. Under procedural data due process, companies that make algorithmic decisions would be obliged to give hearings to consumers with grievances and “correct the record,” if needed, including “examining the evidence used, including both the data input and the algorithmic logic applied” (ibid., 127). But a frequent input to the algorithmic decision making of
financial organizations is data purchased from brokers, and data brokers' contracts typically limit buyers' access to the sources of the personal data in the purchased data sets (Tanner 2013). As a result, neither the consumers nor the financial institutions in a procedural due process hearing might be able to get the information they need to correct erroneous data at their source.

Finally, even when the data protection provisions of contracts between organizations are sound and appropriate, all kinds of contracts among business partners are variably enforced (Johnson and Sohi 2016).

Summary

In short, the legal and regulatory environment of data protection is fragmented, creating a patchwork of rules with overlaps, conflicts, and gaps. Standards, frameworks, and technologies for dealing with this complexity abound, spawning bodies of specialized knowledge and certification programs for an army of data protection and regulatory compliance specialists (Bamberger and Mulligan 2011b). This situation has a number of important implications for the effectiveness of organizational data protection.

First, there is no easy or inexpensive way to ensure end-to-end governance across complex data supply chains (Hartzog 2012; Markus and Jacobson 2015). Second, there may be a lack of clarity inside organizations about who is responsible for what. The lack of such clarity often creates internal conflict and gaps in policies along with their enforcement.

Third, people who are not specialists in data protection cannot be expected to be knowledgeable about the issues and rules. For example, a corporate recruiter may not know that US laws governing employer access to employee personal online information vary from state to state (Park 2014). US data scientists and their clients may not know that electric meter data is protected personal data under EU law. Unfortunately, awareness of the issues and rules does not figure prominently in the educational preparation of data scientists and their clients. This increases the pressure on the chiefs to work well together, design good rules, communicate the rules effectively, monitor employee compliance continuously (possibly invading their privacy in the process), and diligently enforce the rules.

Lastly, what is legal and possible for an organization to do with big data may be unethical or socially unacceptable. Organizations are told that they cannot just rely on following the law; in addition, they must articulate organizational values and ethical principles (Culnan and Williams 2009; Davis 2012), adopt industry ethical codes (Bennett and Mulligan 2012) or develop their own, and take steps to instill these codes in corporate culture (Bamberger and Mulligan 2011b; Hu et al. 2012). These meritorious proposals are challenged by differences in beliefs and values about the benefits and risks of big data as well as the need for data protection. Differences in beliefs and values may lead to poor individual compliance, even when organizational policies are sound.
Data Management Capability Is Not a Monolith

Let's assume that governance is a precondition for the success of endeavors like data protection that involve the activities of many different actors. Governance alone cannot guarantee success, however. The goals and policies set by governing bodies must be translated into organizational practices routinely enacted by skilled employees. That takes resources—time, money, and political capital. Organizations can create and staff data protection and management roles like chief privacy officer and chief compliance officer without allocating directly to them the resources necessary to carry out their regulatory mandates, particularly when these mandates require technological change. In many organizations, the likely source of the technical resources needed for data protection is the organizational unit or units involved in information technology and data management (Wheatman and Akdshay 2015). Within information technology units, data protection is one of several, possibly competing priorities (Proctor, Wheeler, and Pratap 2015).

Why Should Organizations Protect Data?

It is useful to review organizational motivations to protect personal data. In brief, there are three motivations: data protection is the right thing to do because it benefits society along with the persons about whom data are collected and analyzed; data protection is required, by law or custom; and data protection is a good thing to do because organizations can benefit from doing so. These motivations do not necessarily conflict. Yet they are more compelling when they reinforce each other.

Values differ, and not everyone believes that protecting personal data is a strong moral imperative for organizations (Davis 2012). It goes without saying that data protection is not the first priority of most organizations; indeed, it is always ancillary to the core mission of organizations in finance, health care, and retailing. So if data protection (or its resource demands) is seen as conflicting with an organization’s core mission, data protection is likely to suffer.

The law mandates certain data protection activities, and noncompliance can be heavily punished. In addition, data breaches and cyberattacks are the focus of much current public and regulatory attention. As a result, data protection has become a major—if not the major—current priority of many organizational executives and boards of directors (North Carolina State University’s ERM Initiative and Protiviti 2015). People in organizations, though, differ in how much they are motivated by fear of doing the wrong thing (Gentile 2010). For some, the desire to avoid losses from poor data protection is outweighed by the desire to gain benefits from pursuing the organization’s core mission through innovative applications of big data analytics. Regulatory mandate alone therefore may not be sufficiently compelling to attract enough resources for effective data protection.

The argument can also be made that organizations gain positive benefits from effective data protection. An organization may be able to enhance its reputation with the public by
becoming known for its data protection practices. (At the same time, actively marketing data protection excellence could increase liability in the case of a breach.) Data protection has the potential to generate financial benefits for organizations, too, in at least two ways. First, devoting attention and resources to data protection can reveal opportunities for improvements in operational efficiency (e.g., GRC programs that substantially reduce compliance costs), or innovative new products and services (e.g., banks’ identity theft protection services). Second, devoting attention and resources to data protection can encourage employees and organizational units to actively pursue big data projects, which may then yield financial benefits.

Ironically, despite all the hype, organizations have not yet deployed big data analytics as pervasively as expected. Among the reasons is an unwillingness to share data within organizations. A business unit in charge of checking and savings accounts, for instance, may be unwilling to share customer data for cross-marketing purposes with a business unit in charge of mortgage lending. An unwillingness to share data may reflect organizational turf battles. But some information technology experts have observed that data sharing is hindered by too-tight internal controls (Tallon, Ramirez, and Short 2014) or excessive concern about the legality or acceptability of a proposed data reuse. By putting in place a process to vet proposed data uses in light of regulations and company values, organizations can encourage employees to share and reuse data. Intel called its version of such a program “protect to enable” (Tallon, Short, and Harkins 2013, 189).

Organizations can have compelling motivations to pursue data protection. But the motivation to protect data is often linked to the motivation to use data. It should not be surprising, then, that a likely source of the resources required for the implementation of policies set by regulatory chiefs (e.g., financial, privacy, risk, and compliance) is the technical chiefs (e.g., information, information security, data, and analytics) who are collectively responsible for supporting intensive data use.

The Varieties of Data Management
In the realm of the technical chiefs, data protection is one of several data management disciplines and priorities (Khatri and Brown 2010). DAMA International, a leading professional association, has developed a framework and body of knowledge on data management, comprising the following disciplines:

- Data architecture management
- Data development
- Data operations management
- Data security
- Reference and master data management
- Data warehousing and business intelligence management
- Document and content management
• Meta data management
• Data management

This framework does not include privacy as a top-level topic (privacy is included under security), nor does the framework adequately deal with the unique challenges of big data and analytics (such as algorithm version control and updating). More important for data protection, each of the disciplines in the DAMA framework is a major organizational data-related activity and priority in its own right.

Consider data operations management. This essential discipline ensures that data are available when needed and provides backup and recovery from crashes. Data architecture management is a particular challenge for retail banks, which are often organized into product-focused units or have grown through acquisition. These banks can have multiple stores of customer data, each with different naming and retrieval conventions. In order to use big data and analytics for activities like cross-selling financial products to their customers, banks may have to engage in a costly program of systems or data integration. Even banks with well-designed data architectures may suffer from data quality problems. Some banking data are key entered by people (including customers) who are not directly affected by erroneous or missing data, resulting in data that are insufficient to support analytics projects. Thus, organizations need to create data quality metrics, assign responsibility for routine monitoring of the indicators, and take corrective action. Although some people have argued that data quality is unimportant in the presence of massive big data reservoirs (Mayer-Schonberger and Cukier 2013), data scientists frequently mention the challenges involved in accurately matching data (Wigan and Clarke 2013).

In short, getting value from big data and analytics involves a number of supportive data management disciplines, of which data protection is only one. Depending on an organization's level of data management "maturity," data protection may not be high on the list of the data-related priorities of the people charged with data management (Friedman, White, and Judah 2015). Furthermore, significant organizational events, such as mergers or acquisitions, strategic redirections, and corporate reorganizations, can require data rearchitecting, thereby de-emphasizing data protection on the list of an organization's data priorities.

Summary
All things considered, data protection is an organizational challenge that competes with many other issues for management time, attention, and resources. Although external events (high-visibility data breaches in other organizations) and changing regulations (cybersecurity breach reporting rules, and the increased liability of directors and officers) can refocus priorities in the short run, continuing shortfall in the implementation of better data protection policies, practices, and technologies is probably a fact of life.

Within the technical units charged with supporting data and technology, data protection is just one of several crucial data management disciplines. As a result, proposals for
investment to improve data protection may compete with proposals for investment to support intensive data use.

Concluding Remarks

Every organization in every data supply chain needs to do its part for effective data protection. Within organizations, all employees must do their parts as well. But organizations and their employees face significant hurdles in the discharge of data responsibility.

The first challenge lies in the nonmonolithic character of data use within and across organizations. Data protection requires the coordinated and concerted efforts of many individuals and groups that differ widely in organizational role, occupational specialty, knowledge, incentives, culture, and values. Divisions among these individuals and groups create variable commitments to data protection.

Second, the regime of data governance now in place to support data protection also undermines it. There are too many conflicting and incomplete rules. Specialized experts have emerged to navigate the rules, but they are too numerous, contentious, and thinly placed to address the ethical issues that continue to arise as technology evolves. Perhaps more important, the growth of data protection as an area of specialization (several, actually) seems to run counter to the ideal of getting all organizational members to embrace data protection (Bamberger and Mulligan 2011a; Hu et al. 2012). Particularly worrisome are the ethics gaps (McGinn 2015) among data scientists, clients, and specialists in data protection.

Data management regulations impose large and growing costs on organizations. Unless or until organizations can find ways to benefit from better data protection (e.g., through cost reduction, competitive advantage, or innovation), their compliance will be formulaic and grudging.

Third, data protection is only one of several technical data management disciplines and priorities. Data protection is costly, and organizations are told that they should not protect all data equally but rather in proportion to their value to the enterprise. One might prefer that advice to refer to the information's value to the individual whose personal data are involved. But the fact remains that in order to afford the cost of protection, organizations must have benefits from using data. Governments and data businesses clearly have those benefits, but for many organizations, the big benefits of big data and analytics still lie in the future. This means that proposed investments for data protection may compete with proposed investments to support data use.

The good news is that organizations can have strong motivations to protect personal data. Not only does data protection benefit employees, customers, and all citizens, and not only is data protection a regulatory mandate, but better data protection and management can also benefit organizations in their reputations and bottom lines.

These observations have important implications for policy proposals intended to improve data protection, whether the proposals emphasize government regulation or organizational
self-governance. For one thing, getting all individuals and groups lined up and moving in the same direction is a daunting challenge at best, and there will be no quick technical fix. For another, it is hard to get data protection “just right.” Too little data protection is bad for organizations, not just employees and customers, because it leaves organizations vulnerable to risks from both inside and out, and because it discourages employees from sharing data and benefiting from the use of analytics (Tallon, Short, and Harkins 2013). But too much protection is also bad. Not only is it costly to organizations; it is bad for people because it can paradoxically encourage organizational employees to ignore or circumvent the rules. Thus, merely adding more rules to the existing mix is likely to be counterproductive. A clean-sheet redesign of today’s data protection regime may not be in the cards. Yet for every new rule added, two older rules should be stricken from the books.

Acknowledgments

This work was supported in part by the National Science Foundation under award #1348929, M. Lynne Markus, principal investigator. The opinions expressed herein are those of the author, not of the National Science Foundation.