Yale Journal of Law & Technology
Volume 28, Symposium Issue

Information About Data

Mihailis E. Diamantis[*], Chen Sun[**], Rishab Nithyanand[***]

Deterrence-based approaches to privacy enforcement rely on an overlooked and often false premise—that firms know what their own data practices are. There is good reason for skepticism because operational information tends to become siloed within firm subunits. Information about data management is no different. Firms may neglect to memorialize relevant information in reports for internal distribution. And even if such reports are generated, they may not be presented in a manner that is intelligible across firm constituencies.

This paper looks outside of privacy law for a solution. Recent scholarship on securities disclosures has highlighted the variety of goals that disclosures serve. While the traditional purpose of financial disclosures is to inform outside investors, the process of preparing disclosures has beneficial internal effects too. It forces firms to study their own financial health and ensures that relevant corporate units are apprised of the results.

Mandatory disclosures about corporate data practices could have similarly beneficial effects. While some states already require firms to publish generic information about data practices to consumers, these disclosures lack basic attributes that make financial disclosures effective—they lack detail, no human signs them, and they are not filed with any state authority. Securities-style disclosures hold more promise. By carefully tailoring the content, format, and required signatories of data practice disclosures, authorities could force firms to generate, translate, and internally propagate important information about data. Firms that actually know what they are doing with data are more susceptible to efforts aimed to deter data misuse.

---

[*] Ben V. Willie Professor in Excellence, University of Iowa College of Law.
[**] PhD Candidate, University of Iowa, Department of Computer Science.
[***] Assistant Professor and Emeriti-Faculty Scholar, University of Iowa, Department of Computer Science.

Article Contents

Non-financial institutions often have no idea what they're doing with personal data.

-- R.J. Assaly, Chief Product Officer at Reflixivity[1]

We have no idea where we keep all your personal data.

-- Facebook Engineers[2]

Introduction

Last year, U.S. legislators learned that General Motors was sharing customers' individualized driving data and called on the Federal Trade Commission to investigate.[3] The story that emerged from various actors within GM was a confused jumble. GM officials and later a GM spokesperson separately reported that GM only collected "de-identified location data" to help "enhance[e] city infrastructure and road safety." At the same time, GM was marketing the program to customers as a way to lower their personal (and hence very identifiable) insurance rates.[4]

A few years earlier, legislators summoned Facebook CEO Mark Zuckerberg to testify about an incident in which his company paid teens for their detailed internet data.[5]

---

[1] Interview with Authors, via Zoom, Oct. 3, 2024.

[2] Sam Biddle, *Facebook Engineers: We Have No Idea Where We Keep All Your Personal Data*, Intercept (Sept. 7, 2022), https://theintercept.com/2022/09/07/facebook-personal-data-no-accountability/ [https://perma.cc/GR39-QYJL].

[3] Katie Balevic, *Your New GMC Truck May Be Tracking Your Every Movement and Sharing It with Data Brokers*, BUS. INSIDER (July 28, 2024), https://www.businessinsider.com/gmc-gm-hyundai-honda-location-data-collection-ftc-investigation-2024-7?utm_source=chatgpt.com [https://perma.cc/EW5V-ZMBV].

[4] In their contribution to the symposium, Joel Reardon, Ken Bamberger, and Serge Egelman uncover a broad trend of firms making such inconsistent representations about their data practices. *Anonymity, Consent, and Other Noble Lies: An Empirical Study of the Data Economy*, 27 YALE J.L. & TECH. __ (forthcoming 2025).

[5] Devin Coldewey, *Zuckerberg Unconvincingly Feigns Ignorance of Data-Sucking VPN Scandal*, TECHCRUNCH (July 29, 2020), https://techcrunch.com/2020/07/29/zuckerberg-unconvincingly-feigns-ignorance-of-data-sucking-vpn-scandal/ [https://perma.cc/8WGG-9ETL].

Zuckerberg professed ignorance about the project and about whether it was ongoing. A couple of years earlier, surrounding the Cambridge Analytica scandal, Zuckerberg told Congress he "didn't know if there were any conversations at Facebook" about informing affected users.[6] Whether Zuckerberg truly was "ignorant or [was] deliberately misleading" legislators about what he knew,[7] the effect was the same—Facebook was dodging accountability.

There is a lot people don't know about how corporations collect, use, and transfer personal data. A quarter century ago, scholars observed that consumers especially are in the dark.[8] And now, despite decades of advocacy and statutory intervention, studies continue to show that "[h]igh percentages of Americans don't know . . . anything about [company's] basic [data] practices and policies."[9] This is problematic. Consumer ignorance undermines traditional ethical and legal foundations for data use—like informed consent or contextual integrity—and disempowers unwary individuals vis-à-vis the businesses who watch them.[10] Lay ignorance about how corporations

---

[6] Sam Biddle, *Mark Zuckerberg Is Either Ignorant or Deliberately Misleading Congress*, INTERCEPT_ (Apr. 18, 2020), https://theintercept.com/2018/04/11/mark-zuckerberg-is-either-ignorant-deliberately-misleading-congress-or-both/ [https://perma.cc/ZR7E-X68P].

[7] *Id.*

[8] Jeff Sovern, *Opting in, Opting Out, or No Options at All: The Fight for Control of Personal Information*, 74 WASH. L. REV. 1033, 1084 (1999) ("[M]any consumers apparently do not know how businesses use their information."); Dominique-Chantale Alepin, *Social Media, Right to Privacy and the California Consumer Privacy Act*, 29 COMPETITION: J. ANTI., UCL & PRIVACY SEC. CAL. L. ASSOC. 96, 98 (2019) ("we do not know what companies are doing with our data." (quoting Jennifer Lynch, Surveillance Litigation Director with the Electronic Frontier Foundation)).

[9] Joseph Turow et al., AMERICANS CAN'T CONSENT TO COMPANIES' USE OF THEIR DATA 2 (2023), *available at* https://www.asc.upenn.edu/sites/default/files/2023-02/Americans_Can%27t_Consent.pdf [https://perma.cc/TGD3-WSUA].

[10] Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and A New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1854 (2011) ("Yet the lack of transparency regarding practices of data collection and tracking creates an asymmetry of knowledge about existing information collection practices between consumers and the organizations that collect information about them."); Turow et al., *supra* note 9

collect, use, and share personal data is a well-documented phenomenon.

The GM and Facebook examples illustrate a less familiar and potentially more pernicious possibility: that firms can be ignorant of *their own* data practices. David Choffnes et al. observed two years ago that "[d]evelopers often do not fully understand the information flow in their own systems, due to their complexity, 'time to market' pressure, and use of third-party software or hardware with their own opaque data practices." [11] If true, this sort of corporate ignorance would undercut a foundational and unstated premise of any present effort to deter corporate misuse of data. Enforcement against an entity that is unaware of its own behavior is futile.[12] Such an entity can neither compare its own behavior to benchmarks nor credibly correct problems as they arise. "[E]ven when companies want to comply with privacy laws, [they may find it] challenging to do so."[13]

This paper investigates whether firms know what personal data they're collecting, know what they do with it, and know what internal and external commitments they've made regarding data use. We start by offering a model of corporate knowledge and how firms may come to lack it about even seemingly obvious operational facts (Part II). On this model, a corporation knows how it handles data only when two conditions are satisfied: 1) its data practices and policies have been represented in human-comprehensible form, and 2) that representation has been internally distributed to all relevant functional units. Various organizational dysfunctions, some generic and some specific to technical enterprises, can cause failures of either conditions.

We then apply this model in an empirical investigation of whether firms are generally aware of their practices and

("Genuine opt out and opt in consent requires that people have knowledge about commercial data-extraction practices.").

[11] David Choffnes et al., *A Scientific Approach to Tech Accountability*, 37 HARV J.L. &. TECH. 1201, 1210 (2023).

[12] Indeed, ignorance of one's own conduct is a very old defense to criminal liability. Rex v. Arnold, Y.B. 10 Geo. 1 [1724] (defining insanity as being "totally deprived of [one's] understanding and memory, so as not to know what [one] is doing, no more than an infant, a brute, or a wild beast").

[13] Choffnes et al., *supra* note 11, at 1210.

policies for collecting, sharing, and using data (Part III). Through a combination of interviews and surveys of employees at corporations that handle large amounts of personal data, we uncover a nuanced picture. Our results show that firm self-awareness is not inevitable. Firms are most likely to know their own practices for handling sensitive personal information when they implement robust documentation and oversight procedures, and then reinforce those procedures through strong cultural norms about inter-unit communication.

Finally, we offer a path forward (Part IV). Ignorance can be advantageous to corporations. In both courts of law and courts of public opinion, plausibly denying awareness of internal abuses is a well-trod line of defense. But the law needn't sit back and let these incentives dictate outcomes. The law has various information forcing mechanisms at its disposal. Drawing on our prior work,[14] we propose implementing a securities-style disclosure regime for corporate data practices and policies. As firms complete the required disclosure forms, they would come to satisfy the first condition of corporate knowledge, i.e. representing practices and policies in human comprehensible form. If the disclosure then requires signoff from representatives in relevant corporate functional units, firms would come to satisfy the second condition of corporate knowledge, i.e. distribution. The act of filing the form with authorities would ensure the integrity of the process and frustrate any subsequent corporate efforts to deny awareness of how personal data is handled.

I. Modeling Corporate Knowledge and Ignorance

Before we can investigate whether corporations know how they handle personal data, we need an understanding of what such corporate knowledge would amount to. While there is an intuitive sense in which humans can be knowledgeable or ignorant, the concept of corporate knowledge continues to vex lawyers, judges, and philosophers.[15] The law does have its

---

[14] Mihailis E. Diamantis et al., *Forms of Disclosure: The Path to Automating Closed Book Privacy Audits*, 37 HARV. J.L. & TECH. 1265 (2024).

[15] Mihailis E. Diamantis, *Functional Corporate Knowledge*, 61 WM. & MARY L. REV. 319 (2019).

definitions and redefinitions, but every attempt defies common sense and sound policy. For example, the most widespread legal test for corporate knowledge is *respondeat superior*, which, in relevant part, states that a corporation knows whatever at least one of its employees knows.[16] The doctrine has been widely criticized for holding even very large corporations accountable for the secret knowledge of isolated rogues within their ranks.[17] Variants like state law's inner circle approach[18] or the First Circuit's collective knowledge doctrine[19] only make matters worse.[20]

The problem with existing legal definitions of corporate knowledge is that they divorce corporate knowledge from any meaningful understanding of corporate action. Knowledge and action are conceptually intertwined.[21] People act on the basis of what they know, or at least what they think they know. If an injured person drives himself to the hospital, it would be implausible to claim he didn't know the way. If he instead drove around in circles, it would be implausible to claim he did. *Respondeat superior* and its variants allow for cases where a corporation counts as knowing something (legally speaking) but doesn't at all act like it. A rogue employee may know a program is doomed to failure (perhaps he's the one who sabotaged it!) even as the rest of the corporation charges enthusiastically ahead. Since our concern in this essay is the kind of ignorance that would inhibit a corporation from implementing responsible data practices, the conceptual

---

[16] 18B AM. JUR. 2d *Corporations* § 1822 (2016).

[17] Preet Bharara, *Corporations Cry Uncle and Their Employees Cry Foul: Rethinking Prosecutorial Pressure on Corporate Defendants*, 44 AM. CRIM. L. REV. 53, 65 (2007) ("[A] multinational corporation may theoretically be indicted, convicted, and perhaps put out of business based on the alleged criminal activity of a single, low-level, rogue employee who was acting without the knowledge of any executive or director, in violation of well-publicized procedures, practices, and instructions of the company.").

[18] MODEL PENAL CODE § 2.07(1)(c) (AM. LAW INST. 1985).

[19] United States v. Bank of New England, 821 F.2d 844 (1st Cir. 1987).

[20] Mihailis E. Diamantis, *Corporate Criminal Minds*, 91 NOTRE DAME L. REV. 2049, 2068-71 (2016).

[21] BERTRAND F, MALLE, HOW THE MIND EXPLAINS BEHAVIOUR: FOLK EXPLANATION, MEANING, AND SOCIAL INTERACTION (2004); Mihailis E. Diamantis, *How To Read a Corporation's Mind*, *in* THE CULPABLE CORPORATE MIND (Elise Bant ed., 2023).

connection between knowledge and behavior must be maintained.

For present purposes, we eschew the legal definition of corporate knowledge and borrow one that's more familiar to cognitive scientists and philosophers. "Access consciousness" refers to the information processing function of awareness.[22] Roughly speaking, a system is access conscious of an internal state if the state's "content—what is represented by [it]—is processed via that information processing function, that is, if its content gets to the executive system, whereby it can be used to control reasoning and behavior."[23] In other words, there are two conditions for access consciousness. First, the system must prepare a representation of the state that is suitable for information processing. Second, that representation must be transmitted to the executive system for planning and action.

The concept of access consciousness was developed to model human awareness of perceptual states, but it translates easily into a criterion for an organization's knowledge of its data practices and policies. First, the organization must prepare representations of its data practices and policies that can be distributed internally. To avoid confusion, we use "data" in this essay to refer exclusively to consumer data that a corporation collects, processes, or shares. "Information about data" refers to representations of facts about the firm's data collection, processing, and sharing practices, or relevant policies thereto. The sort of information about data that a firm must generate to satisfy the first condition includes not only the high-level (and frequently ambiguous) information typically displayed in consumer-facing privacy policies—e.g. what data is collected and with whom shared—but also more granular information about the mechanics of internal data management—e.g. how long internal use permissions last and what metadata is preserved on transfer.

Once a firm has generated information about its data handling practices and policies, the second thing it must do is

---

[22] D. Schacter et al., *Access to Consciousness: Dissociations Between Implicit and Explicit Knowledge in Neuropsychological Syndromes*, *in* THOUGHT WITHOUT ACTION (L. Weiskrantz ed., 1988).

[23] Ned Block, *On a Confusion About a Function of Consciousness*, 18 BEHAV. & BRAIN SCI. 227, 228 (1995).

distribute that information for use in organizational planning and behavior. In the case of human access consciousness, that means that the brain must transmit its representation to the frontal lobe and prefrontal cortex, where executive reasoning takes place. Unlike humans, organizations have distributed executive systems. As reflected in their organization charts, firms use nested hierarchies to carry out function specific planning and operations. We identify four firm units whose proper function involves acting on information about data:

- Research and Development: This unit often has the most extensive contact with data. Its personnel directly handle data to carry out the firm's data management practices and to improve the products/services it offers. They will often be the predominant source of information about data handling practices and the most important recipient of it.
- Marketing: Marketing personnel use both data and information about data. When drafting consumer-facing materials, they make statements about how their firm handles data and what its internal data policies are. Marketing may also use data to refine and target ad campaigns.
- In-House Counsel: A firm's attorneys will need information about data practices to confirm that they comport with the state and federal requirements, as well as any other legal commitments (e.g. contractual) the firm has assumed. They also provide information about data to business counterparties and government regulators.
- C-Suite: The highest-level executives need information about data to inform strategic decision making and business planning.

Figure 1 shows the sort of interaction among functional units that would amount to a firm knowing its own data practices:

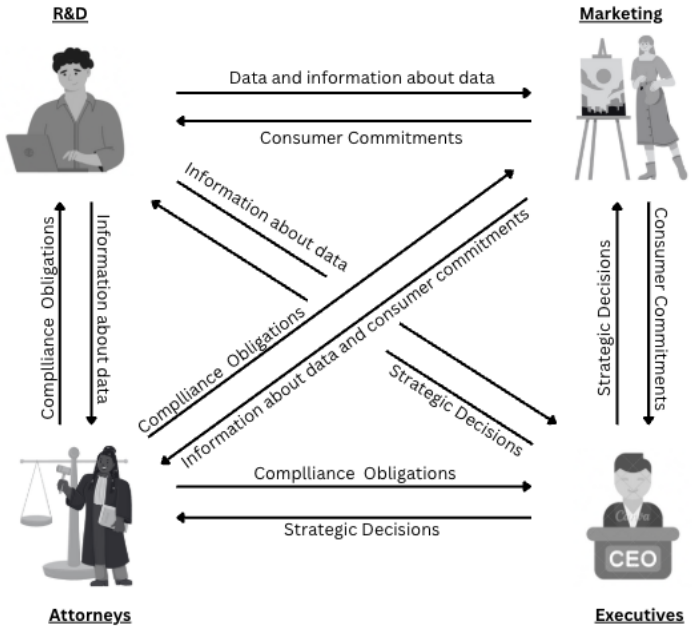

Figure 1: Representation of a firm knowing its data practices.

With a model of firm knowledge in place, now we can anticipate ways in which a firm could fall short, whether by failing one of the two conditions directly or by failing to satisfy them in a way that defeats their anticipated interaction. For example, the firm may fail to generate representations of information about data. This may be because no one within the research and development team has been charged with crafting reports of what those practices are. Even if reports are generated, they may not be in a format that other units can readily use to direct planning and action. Marketing personnel, attorneys, and executives may not be able to understand information about data if it is represented in overly technical language.

Alternatively, a firm may generate representations of information about data but fail to distribute them. Knowledge about data practices and policies requires regular and easy communication between functional units. This may not be an issue in micro-firms, where all employees know each other and work in close proximity. However, "larger, more complex

organizations often suffer from information silos, which occur when departments or divisions within a large organization are isolated from other parts of the organization."[24] There is a natural "propensity of departments or divisions within a large organization to become isolated, with a resulting failure to communicate and pursue common goals."[25] Indeed, one study found that eighty-three percent of large companies report having compromised performance because of internal silos.[26] Silos can even exist within individual units of a firm.[27]

Information silos "result in difficulty communicating properly throughout the organization."[28] They can arise for simple reasons, like lack of physical proximity between functional units, entrenchment within departments, individuals with obstructive personalities,[29] and incompatible software tools.[30] They can also arise for more complex reasons of organizational structure, like lack of process, missing infrastructure, or failure to assign oversight responsibilities.[31]

---

[24] Veronica Root Martinez, *Complex Compliance Investigations*, 120 COLUM. L. REV. 249, 256 (2020).

[25] Richard E. Levy & Robert L. Glickman, *Agency-Specific Precedents*, 89 TEX. L. REV. 499, 510 (2011); *see* Fabio Bento et al., *Organizational Silos: A Scoping Review Informed by a Behavioral Perspective on Systems and Networks*, 10 SOCIETIES 56 (2020) (discussing the persistent problem of organizational silos).

[26] F. Stone, *Deconstructing Silos and Supporting Collaboration*, 31 EMPLOY. RELAT. TODAY 11 (2004)

[27] Juergen Rilling et al., *Beyond Information Silos—An Omnipresent Approach to Software Evolution*, 2 INT'L J. SEMANTIC COMPUTING 341 (2008).

[28] *Id.* at 257.

[29] Aishat Jeleel-Ojuade, *The Role of Information Silos: An Analysis of How the Categorization of Information Silos Within Financial Institutions, Hindering Effective Communication and Collaboration* (July 10, 2024), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4881342 [https://perma.cc/CV8P-RCRL].

[30] Juergen Rilling et al., *supra* note 27.

[31] Nicola Faith Sharpe, *Process over Structure: An Organizational Behavior Approach to Improving Corporate Boards*, 82 S. CAL. L. REV. 261, 291 (2012) ("[S]tructure delimits organizational responsibilities and communication channels."); Martinez, *supra* note 24, at 263 ("[P]rocess refers to the actions, practices, and routines firms employ to communicate and analyze information necessary for creating an effective ethics and compliance program.").

Information silos are especially common concerning technical topics like information about data. Technologists seem to have particular difficulty communicating with other firm units. This may be because "today's information systems are so opaque"[32] and translating them for nontechnical audiences takes skill and care.[33] Given that the majority of S&P500 companies have no technology expertise on their board,[34] communication barriers between technical functions and corporate leadership are bound to arise. Horizontal barriers are common too,[35] for example between technology and marketing functions within a firm.[36]

Firms' awareness of their own data handling practices is a critical ingredient for state efforts to induce responsible data collection and use. Most basically, knowledge about data misuse is often a condition of legal liability.[37] When

---

[32] Choffnes et al., *supra* note 11, at 1209.

[33] *See also* Orla Cox & Hetal Kanji, *Building Effective Cybersecurity Governance*, Harvard Law School Forum on Corporate Governance (Nov. 10, 2022), https://corpgov.law.harvard.edu/2022/11/10/building-effective-cybersecurity-governance/ [https://perma.cc/6RMJ-QP3H] ("[To ensure robust cyber security practices,] the CISO [Chief Information Security Officer] will need to communicate cyber risk and metrics in terms that resonate with the board.").

[34] Sarah Kuranda Vallone & Edna Twumwaa Frimpong, *State of Cyber Awareness in the Boardroom*, Harvard Law School Forum on Corporate Governance (Oct. 10, 2023), https://corpgov.law.harvard.edu/2023/10/10/state-of-cyber-awareness-in-the-boardroom// [https://perma.cc/C573-KQDM]. A similar problem used to arise with respect to financial literacy on the board too. Robert Charles Clark, *Corporate Governance Changes in the Wake of the Sarbanes-Oxley Act: A Morality Tale for Policymakers Too*, 22 GA. ST. U. L. REV. 251, 272 (2005) ("[N]ew standards require all members of audit committees to be 'financially literate.' . . . The assumption is that having financial literacy and expertise on the committee makes it less likely that questionable, confusing, or misleading accounting policies and judgments will go undetected and uncorrected.").

[35] Martinez, *supra* note 24, at 266-67 ("[T]here is often a failure to share information across organizational units, leading to the creation of information silos.").

[36] Stefan Sleep et al., *Removing Silos to Enable Data-Driven Decisions: The Importance of Marketing and IT Knowledge, Cooperation, and Information Quality*, 156 J. BUS. RES. 113471 (2023).

[37] *See, e.g.*, Schwartz & Solove, *supra* note 10, at 1864 ("COPPA only regulates a website or online service if it is 'directed to children,' or if the

corporations can claim ignorance, whether genuine or plausibly feigned,[38] the law's deterrent apparatus falls apart. Even more importantly, corporations themselves are often the first, best, and only line of defense when it comes to responsible data management. "An underlying assumption of all modern compliance efforts is that organizations are in the best position to monitor and police the behavior of their members."[39] Enforcement authorities can influence corporate incentives, but their impact on data practices is almost always mediated by the corporation itself.[40] Silos make internal monitoring and rational policy integration impossible, creating disorder as functional units operate under different understandings of how the firm handles data.[41] In such circumstances, even well-intentioned efforts at responsible data management are destined to fail.[42]

II. Measuring What Corporations Know

The goal of our study was to uncover organizational norms and barriers that might prevent firms from knowing

---

operator of the website 'has actual knowledge that it is collecting personal information from a child.' It is relatively easy for website operators to avoid acquiring actual knowledge that they are collecting information from a child.").

[38] Epic Games, *Epic FTC Settlement and Moving Beyond Long-Standing Industry Practices*, https://www.epicgames.com/site/en-US/news/epic-ftc-settlement-and-moving-beyond-long-standing-industry-practices [https://perma.cc/BJW6-BAF9] (Dec. 19, 2022) ("While game developers may be familiar with COPPA, they may not be aware of its global application.").

[39] Martinez, *supra* note 24, at 253.

[40] *See also* Mihailis E. Diamantis, *Clockwork Corporations: A Character Theory of Corporate Punishment*, 103 IOWA L. REV. 507, 526 (2018) ("[D]eterrence theorists hope corporations will choose is to reduce the probability of committing crime by investing in compliance programs. . . . However, what deterrence theorists often neglect is that corporations also have another option—reduce the probability of detection by investing in concealment.").

[41] Martinez, *supra* note 24, at 305 ("If either the data is poor or if important information is missing then an aggregation principle will not add value to the firm's compliance efforts.").

[42] *Id.* at 266-67 ("These silos damage compliance efforts because they impede a firm's effort to prevent, detect, and . . . fully investigate the nature and scope of misconduct within the organization.").

information about their own data handling practices. We employed a two-phase design that integrated surveys and interviews. In the first phase, we conducted interviews with well-qualified subjects to gain insights into the variety of governance structures, norms, and challenges present in different organizations. In the second phase, we used the insights gathered from our interviews to inform the design of a survey to assess organizational data governance knowledge at scale.

The interviews allowed us to gain an in-depth understanding of employees' and experts' experiences related to the governance of sensitive data, while the survey provided a mechanism to gain broad insights about communication norms and perceived barriers across a large variety of firms and employee roles. This mixed methods approach was well suited for our study because generating and propagating information about data involves both observable structures (e.g., policies) and subjective experiences (e.g., how implementation challenges are handled) that cannot be easily captured by a single method. Further, it allowed us to use qualitative findings (from interviews) to contextualize our quantitative data (from surveys).

Together, these data were used to answer two research questions: (*RQ1) How do organizational norms and processes influence employees' awareness of organizational data governance policies and processes; (RQ2) How do organizational norms and processes influence employees' perception of integration in organizational data governance?* We explore these questions because employee awareness of and perceived integration in data governance policies and processes is crucial to firm knowledge. Afterall, awareness and integration will shape employee behavior, guide compliance efforts, and ensure that practices align with both internal policies and external regulatory requirements. By investigating how specific organizational norms and processes affect employee awareness and trust, our analysis will highlight where interventions could facilitate internally distributed knowledge of firm-specific data practices.

We describe our interview and survey procedures below, followed by an integrated discussion of findings as they pertain to the research questions.

A. Interview Procedures

The study began with a series of semi-structured interviews conducted with nine participants who were selected to ensure functional diversity across organizational roles related to data governance. We recruited through professional networks and referrals. Our participants included data engineers, compliance officers, legal counsel, privacy engineers, and product leads with experience handling, directing, or overseeing sensitive personal data. Their employers included small- to large-sized firms of varying maturity levels. This diversity in participants allowed us to capture a range of perspectives on both operational and governance processes. Each interview lasted between 30-60 minutes and was conducted via Zoom. Before beginning, participants were provided with an overview of the study's purpose and assured that their responses would be anonymized unless permission was explicitly granted.

The interview protocol focused on three core topics:

- *How firms handle sensitive data.* Participants were asked to describe how their organizations collect, store, access, and process different types of sensitive data. We explored the technical and procedural infrastructures in place to support these activities, including internal access control systems, logging mechanisms, and data retention protocols.
- *Sensitive data sharing and communication across functional units*. This topic focused on how data-related decisions and responsibilities were distributed across legal, technical, marketing, and executive functions. Participants were asked how information about sensitive data flowed across organizational boundaries. We followed up by probing for descriptions of cross-functional meetings, approval processes, common causes of miscommunication and errors, and escalation pathways.
- *How firms create and communicate data governance policies and processes.* Participants were asked to discuss the internal rules, procedures, and formal

standards that shape their organization's handling of sensitive data. We inquired about how these policies are created, communicated, and enforced. We also asked about employee onboarding, training, and compliance monitoring.

Throughout the interview, participants were encouraged to provide examples from their experience, describe common challenges, recount problem-solving approaches, share known industry norms, and reflect on barriers they had encountered.

Interviews were analyzed using a thematic coding approach. A preliminary coding framework was developed based on the interview protocol and iteratively refined as transcripts and notes were reviewed. Codes were organized around key concepts such as awareness gaps, formal procedures for data sharing, internal communication norms, and barriers that impede effective implementation of governance strategies. These were used to inform the development of scales and items for our survey.

B. Survey Methodology

Our survey was designed to examine how employees across different organizations and operating in a variety of roles perceive and participate in the governance of sensitive data. The primary goal of the survey was to identify cross-unit communication norms, implementation challenges, and barriers that impede effective governance.

1. Recruitment and Participants

This study received IRB approval in March 2025. Following approval, we conducted a two-phase deployment of our survey instrument: a pilot phase to validate our scales, and a full-scale deployment to collect representative data.

a. Pilot Testing

In March 2025, we recruited 31 eligible participants via targeted outreach on social media platforms (e.g., Reddit and LinkedIn) to serve as pilot testers. Participants were required to be currently employed at U.S.-based companies that handle sensitive data. Of the 31 responses received, 9 were excluded due to poor data quality (e.g., straight-lining or completion times more than two standard deviations below the mean). The

remaining 22 high-quality responses were used to assess the validity of our survey scales and inform refinements to item wording and structure.

b. Full-scale deployment

Following successful pilot validation, we partnered with a market research firm in April 2025 to recruit a broader and more diverse sample of respondents. Eligibility criteria were consistent with those used in the pilot phase: participants had to be currently employed at organizations headquartered in the United States and handling sensitive personal data. All participants provided informed consent prior to participation and were assured that the survey collected no identifying information. A total of 197 valid and complete responses were obtained and used for the final analysis. Participants represented a broad cross-section of organizational contexts in terms of firm size and functional responsibility. Respondents were distributed across firm types with 72 (36.5%) employed at large firms with more than 1,000 employees, 56 (28.4%) at mid-sized firms with 250-999 employees, and the remaining (35.1%) at small firms with fewer than 250 employees. Functionally, participants were distributed across technical (38.1%), marketing (34%), and legal/compliance (26.9%) roles. Figure 2 illustrates the breakdown of participants by firm size and functional responsibility.

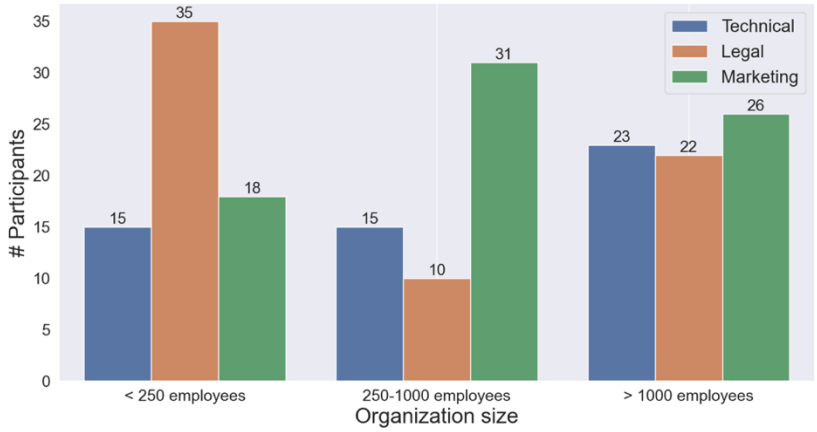

Figure 2: Breakdown of participants by organization size and functional unit.

2. Survey Instrument

To encourage candid responses, the survey did not collect personal identifiers or organization names, and all questions were optional. The instrument comprised three components: (1) background questions to contextualize participant and organizational characteristics; (2) scales assessing perceptions of organizational data governance norms; and (3) scales identifying perceived barriers to effective governance. Table 1 summarizes the scales designed and validated for this study. Throughout the survey, participants were provided with the following definition of sensitive data: *"Sensitive data is any data that could be used to cause harm or embarrassment to an individual within or outside your organization. Sensitive data includes biographic data, location data, financial records, legal records, biometric data, voice or video data, app interaction data, and web browsing data, even if it cannot be directly attributed to an individual."*

Scale	*Description*
Awareness (ODG-A)	A reflective measure of an individual's awareness of their organization's data governance policies and associated processes.
Robustness (ODG-R)	A formative measure of the robustness of an organization's formal process for sharing data across functional units.
Integration (ODG-I)	A reflective measure of how an individual perceives their functional unit integration into organizational data governance.
Support (ODG-S)	A reflective measure of perceived organizational support for compliance with data governance policies.
Environment (ODG-E)	A formative measure of the quality of the environment created by an organization's data governance.
Trust (ODG-T)	A reflective measure of an individual's trust in an organization's

	data governance policies and processes.

Table 1: A summary of the scales used in our survey instrument.

a. Participant and organization background

To contextualize employee perspectives, we captured organizational characteristics (e.g., age, size) and employee roles. Awareness was assessed via the six-item ODG-A scale (Table 2), using a five-point Likert scale. Items were designed to reflect both operational (items 1–3) and representational (items 4–6) knowledge of governance. Validation analyses using responses from our pilot deployment supported a two-factor CFA model (CFI = 0.97, RMSEA = 0.07; Cronbach's α = 0.73 and 0.84, respectively).

	Item	*Dimension (Factor)*
1	I understand the types of sensitive data my organization handles.	Operational
2	I understand the technologies that process sensitive data within my organization.	Operational
3	I understand the mechanisms for logging access and use of sensitive data in my organization.	Operational
4	I understand my organization's policies related to the handling of sensitive data.	Representational
5	I understand my organization's processes for ensuring compliance with data privacy regulations.	Representational
6	I understand the public representations related to the handling of sensitive data made by my organization.	Representational

Table 2: The Organizational Data Governance Awareness (ODG-A) reflective scale.

b. Organizational norms and practices

This section evaluated communication and policy implementation norms. Participants first reported the frequency of cross-unit data sharing and whether formal processes existed. If so, they completed the nine-item ODG-R scale (Table 3), covering documentation, oversight, and review. Due to poor CFA model fit (CFI = 0.56, RMSEA = 0.42) and incoherent EFA results, ODG-R was treated as a formative index.

	Item	*Subscale*
1	The process requires documentation of the source of the data.	Documentation
2	The process requires documentation of the purpose for which the data was gathered.	Documentation
3	The process requires documentation of the intended use of the data being shared.	Documentation
4	The process involves reviews from a team lead or manager in the sending unit.	Oversight
5	The process involves reviews from a team lead or manager in the receiving unit.	Oversight
6	The process involves reviews from a member of the legal or compliance unit.	Review
7	The process involves reviews from a member of a technical unit (e.g., engineering, operations management, etc.)	Review
8	The process requires concerns raised during review to be addressed prior to approval.	Review

9	The proposed data sharing plans and reviews are documented and stored for later review.	Documentation

Table 3: The Organizational Data Governance Robustness (ODG-R) formative indicator subscales.

The ODG-I scale assessed perceived integration, with three tailored variants for technical, legal, and marketing roles. The scale (Table 4) showed good reliability and a reasonable two-factor CFA model (CFI = 0.95, RMSEA = 0.16; Cronbach's α = 0.82, 0.73).

	Item	*Dimension (Factor)*
1	The C-suite executives in my organization consult with representatives of the technical units prior to making decisions related to the governance of sensitive data.	Communication
2	The marketing units within my organization consult with representatives of the technical units prior to making statements to external stakeholders about the organization's data handling practices and policies.	Communication
3	The legal and compliance units within my organization consult with representatives of the technical units to understand how the technical units handle sensitive data.	Communication
4	The C-suite executives in my organization are committed to understanding how the technical units handle sensitive data.	Commitment
5	The marketing units in my organization are committed to understanding how the technical units handle sensitive data.	Commitment

6	The legal and compliance units in my organization are committed to understanding how the technical units handle sensitive data.	Commitment

Table 4: The Organizational Data Governance Integration (ODG-I) reflective scale shown to technical units.

Organizational support was measured using the five-item ODG-S scale (Table 5), capturing infrastructural and cultural support. The scale demonstrated strong internal consistency (Cronbach's $\alpha = 0.88$ and 0.70) and two factor CFA model fit (CFI = 0.98, RMSEA = 0.09).

	Item	*Dimension (Factors)*
1	My organization provides training to help employees achieve compliance with their policies for handling sensitive data.	Infrastructure
2	My organization provides tools to help employees achieve compliance with their policies for handling sensitive data.	Infrastructure
3	My organization encourages cross-unit communication to help employees achieve compliance with their policies for handling sensitive data.	Culture
4	My organization encourages record keeping and documentation when handling sensitive data.	Culture
5	My organization routinely verifies compliance with their policies for handling sensitive data.	Culture

Table 5: The Organizational Data Governance Support (ODG-S) reflective scale.

c. Perceived barriers

Barriers were assessed via two scales: ODG-E (Table 6) and ODG-T (Table 7). ODG-E items addressed policy clarity, consistency, stability, and onboarding. While EFA supported a single factor, poor CFA fit (CFI = 0.58, RMSEA = 0.46) led us to treat it as a formative scale.

	Item
1	My organization's existing policies or processes for handling sensitive data are easy to understand.
2	My organization's existing policies or processes for handling sensitive data are consistent.
3	My organization's policies or processes for handling sensitive data change infrequently.
4	My organization's onboarding or training program is adequate for understanding their policies or processes for handling sensitive data.

Table 6: The Organizational Data Governance Environment (ODG-E) formative indicator scale.

The ODG-T scale measured trust in organizational response to data-related concerns. It showed excellent internal consistency (Cronbach's $\alpha = 0.90$), high item-total correlations, and good CFA fit (CFI = 0.93, RMSEA = 0.12), supporting its use as a reflective scale.

	Item
1	I feel like it would be easy to resolve issues about data originating from my own unit.
2	I feel like it would be easy to resolve issues about data originating from a different unit.
3	I feel like I could report concerns about the handling of sensitive data without fear of reprisal.
4	I feel like my organization would take adequate steps to address my concerns about the handling of sensitive data.
5	I feel like my organization would act quickly to address my concerns about the handling of sensitive data.

Table 7: The Organizational Data Governance Trust (ODG-T) reflective scale.

C. Analysis and Results

Table 8 shows the average measures recorded from our study participants.

Category	*Scale*	*Dimension*	*Mean*
Participant background	Awareness (ODG-A)	Operational	4.2
		Representational	4.2
Norms and practices	Robustness (ODG-R)	Documentation	4.1
		Oversight	4.0
		Review	3.8
	Integration (ODG-I)	Communication	3.9
		Commitment	4.1
	Support (ODG-S)	Infrastructure	4.2
		Culture	4.1
Perceived barriers	Environment (ODG-E)	-	4.1
	Trust (ODG-T)	-	4.0

Table 8: Average measures recorded for each scale in our survey instrument.

On average, participants reported being aware of their organization's data governance practices. For process robustness (ODG-R), participants largely agreed that their organizations had defined documentation and oversight procedures, but review mechanisms were less consistently reported. For integration (ODG-I), participants described only moderate levels of cross-unit communication and commitment, with marketing staff in particular perceiving lower commitment from other units. Mid-sized firms were viewed as offering the strongest communication integration, suggesting that organizational size shapes governance dynamics. Support (ODG-S) was rated higher for infrastructure (training, tools) than for cultural reinforcement, while trust (ODG-T) and environment (ODG-E) received moderate ratings overall, with

lower environment scores among technical and marketing staff than legal staff.

Taken together, these descriptive results suggest that formal governance structures exist in most organizations, but cultural buy-in, trust, and integration are uneven. Formal processes may create the appearance of governance maturity, but employees outside of compliance roles often perceive weaker support and less inclusion in governance discussions. In the remainder of this section, we focus on understanding how specific organizational norms and processes shape two key outcomes: employees' awareness of data governance policies and procedures (RQ1), and their perceived functional integration in data governance (RQ2).

1. Methods of Analysis

To identify which factors most strongly influenced awareness (RQ1) and integration (RQ2), we estimated a series of nested regression models. Four outcome variables were examined separately: operational awareness (ODG-A Operational), representational awareness (ODG-A Representational), cross-unit communication (ODG-I Communication), and commitment to cross-unit integration (ODG-I Commitment). We added predictors to a baseline model in sequential blocks to assess both (1) which features of governance are associated with each outcome variable, and (2) how much additional explanatory power each block contributed beyond prior blocks.

1. (M1) Baseline organizational context model. We first included only background characteristics: functional unit and organization size. This baseline established whether outcomes differed systematically across organizational contexts before we considered governance-specific features.

2. (M2) Adding robustness measures to M1. We then introduced the robustness of governance processes measures (ODG-R), which measured whether data sharing required documentation, oversight, and multi-stakeholder review. Comparing M1 with M2 allowed us to test whether more formalized and rigorous processes

explained awareness and perceived integration beyond baseline characteristics.

3. (M3) Adding support measures to M2. Next, we added organizational support measures (ODG-S) as predictors. Comparing M2 and M3 allowed us to test whether tangible resources and cultural reinforcement explained additional variance beyond the presence of robust processes.

4. (M4) Adding environment and trust measures to M3. Finally, we included employees' perceptions of the governance environment (ODG-E) and their trust in governance (ODG-T) as predictors. Comparing M3 and M4 allowed us to test whether employees' confidence in the governance system mattered as much as its design, and whether the lived experience of governance explained additional variance beyond formal rules and supports.

For each of the above models, we report their R^2 values and standardized β coefficients. The R^2 value represents the proportion of variance in the outcome variable that can be explained by the predictors included in the model and the β coefficients indicate the strength of the association between a specific predictor and outcome variable. Models with higher R^2 values are better able to explain the outcome variable, and predictors with higher β coefficients are more influential on the outcome variable.

2. Factors Influencing Awareness of Data Governance Policies and Procedures (RQ1)

Our nested regression analyses showed that employees' awareness of governance policies and processes was primarily explained by robust governance processes and organizational support.

Using operational awareness as the outcome variable, the fit of our regression model improved from an R^2 of 0.06 in the baseline model (M1) to 0.35 when robustness measures were included as predictors (M2). Including organization support

variables as predictors (M3) further improved R^2 to 0.45. The strongest predictors were robustness ($\beta = 0.29^*$) and cultural support ($\beta = 0.22^*$). Trust and environment variables did not explain any additional variance once the robustness and support variables were included in the model.

Similar patterns were observed while examining representational awareness as an outcome variable. Adding robustness measures increased R^2 from 0.06 to 0.34 and adding governance support variables increased it to 0.45. Once again robustness was found to be the strongest predictor ($\beta = 0.29^*$), while infrastructure ($\beta = 0.18^*$) and cultural ($\beta = 0.15^*$) support variables made smaller contributions. Trust and environment variables had no additional effects on our model.

Together, awareness appeared to be a structural outcome. Organizations that implemented robust documentation and oversight procedures, and reinforced them through cultural norms, had employees that better understood day-to-day governance practices and formal governance commitments. Trust and perceptions of the governance environment did not have any meaningful effect on awareness once these structural features were in place.

3. Factors Influencing Perceived Integration in Data Governance (RQ2)

In contrast to awareness outcomes, cross-unit integration outcomes depended less on structural processes and more on trust and, in the case of cross-unit communication, organization size.

Using our cross-unit communication integration as the outcome variable, we found that the model fit increased from an R^2 of 0.08 in the baseline model (M1) to 0.22 when robustness variables were included (M2), remained unchanged when support variables were included (M3), and further increased to 0.30 when environment and trust variables (M4) were added as predictors. Examining the influence of each predictor in our final model (M4), employees in mid-sized firms (250-1,000 employees) reported statistically higher communication integration ($\beta = 0.30^*$). Trust was also found to have a significant association on communication integration ($\beta = 0.25^*$). Process robustness and support measures did not statistically influence our cross-unit communication measure.

Examining our cross-unit commitment measure as an outcome variable, we found that trust in the organization was decisive. Including trust variables as predictors increased R^2 from 0.33 (in model M3) to 0.45 (in model M4). The trust variable was also the strongest predictor of the outcome variable ($\beta = 0.51^*$). No other predictor variables in our final model (M4) had a statistically meaningful influence on the outcome variable.

Together, our analysis shows that perceptions of communication integration were strongest in mid-sized firms and in organizations where employees trusted governance processes. Perceived integration, in contrast, depended almost entirely on trust. Without confidence that concerns would be heard and acted upon, employees did not feel that their unit's perspectives were integrated, regardless of the existence of formal processes.

4. Takeaways

Our analysis reveals a meaningful distinction between awareness and integration. Awareness improves through structure: robust processes, documentation requirements, and cultural reinforcement. Integration, on the other hand, relies on trust: employees' belief that they can raise concerns safely and that organizations would respond appropriately.

This distinction underscores that mandatory documentation and review of data practices can reliably improve distributed awareness of information about data within a firm. However, such requirements are not sufficient to improve integration unless paired with trust-building mechanisms (e.g., anti-retaliation protections, transparent escalation procedures, and visible organizational follow-up). In short, awareness requires rules and integration requires trust.

III. Ensuring Corporations Keep Themselves Informed

Our initial qualitative study indicated that firms do not reliably know their own data practices and policies. As one of our interviewees observed, "A lot of PII leakage [is due to] not knowing what you're doing; [it] doesn't come from malicious

intention, but not thinking through."[43] Our survey study indicated that firms are most likely to "know what they're doing" with sensitive data when they have robust data documentation and oversight procedures, reinforced through cultural norms that encourage cross-unit communication about data practices and expectations. In terms of the model of corporate knowledge sketched previously, firms must document representations of their data practices (the first condition) and then communicate them in suitable format within the firm for planning and behavior (the second condition).

Communication between R&D and other units could break down because there is no reliable channel or because the information R&D prepares is not represented in a form that is readily comprehensible to non-technical units. We propose that appropriately structured mandatory disclosures to state or federal authorities could help firms solve both problems.

A motivating example will help. In 2019, the Federal Trade Commission settled its investigation into Facebook's Cambridge Analytica scandal. As discussed in the introduction, CEO Mark Zuckerberg's testimony painted a picture of dysfunction within Facebook. Important information about data had, according to him, become siloed within firm units. Since it didn't flow to Zuckerberg, he couldn't exercise strategic oversight. Whether this was the truth or a hard-to-disprove lie, the solution the FTC identified was the same: to require disclosures that would provoke the documentation and distribution of information about data within Facebook.

FTC Chairman Joe Simmons explicitly characterized the strategy in terms of "mandating . . . flows of information" throughout the corporate hierarchy.[44] The settlement created a third-party independent assessor with wide ranging access to Facebook documents and personnel for fact-gathering,

---

[43] Assaly, Interview with Authors, via Zoom, Oct. 3, 2024.

[44] Fed. Trade Comm'n, Statement of Chairman Joe Simons and Commissioners Noah Joshua Phillips and Christine N. Wilson, *In re Facebook* (July 24, 2019), https://www.ftc.gov/system/files/documents/public_statements/1536946/092_3184_facebook_majority_statement_7-24-19.pdf [https://perma.cc/BNV8-PBE5].

sampling, and testing.[45] The assessor, an independent privacy committee (composed of independent Facebook directors), and Facebook management were all to meet regularly to discuss and document risks to Facebook's privacy program.[46] Regular privacy training programs would distribute information about program updates and risks to Facebook personnel.[47] The committee would also file quarterly reports to the Zuckerberg and designated compliance officers. "[T]hese multiple information flows . . . [were designed to] provide both a constant reminder of the company's privacy obligations as well as overlapping oversight."[48] This would help to build a privacy-aware corporate culture. To ensure the integrity of the process, Facebook was to file quarterly reports with the FTC, personally certified by Zuckerberg and Facebook's compliance officers. "False certifications would subject [these individuals] to personal liability, including civil and criminal penalties."[49] As James Kohm, then head of the FTC's enforcement unit, said, one goal was to leave "no way that the CEO can bury his head in the sand" by engineering or feigning his own ignorance.[50]

Whether purposely or not, the FTC's settlement with Facebook channeled a recent trend in disclosure theory. Disclosures, like those the SEC requires, are typically seen as vehicles for informing external parties about firm performance. They help investors decide where to send their money, tip off public enforcers to illicit activity, and facilitate activism by private parties.[51] But the process of preparing disclosures also

---

[45] United States v. Facebook, Inc., Stipulated Order for Civil Penalty, Monetary Judgment, and Injunctive Relief, Case No. 19-cv-2184 (D. DC July 24, 2019), https://www.ftc.gov/system/files/documents/cases/182_3109_facebook_order_filed_7-24-19.pdf [https://perma.cc/45R8-NQ6V].

[46] *Id.*

[47] *Id.*

[48] Fed. Trade Comm'n, *supra* note 44.

[49] *Id.*

[50] Aart Shahani & Avie Schneider, *FTC to Hold Facebook CEO Mark Zuckerberg Liable for Any Future Privacy Violations*, NPR (July 24, 2019), https://www.npr.org/2019/07/24/741282397/facebook-to-pay-5-billion-to-settle-ftc-privacy-case [https://perma.cc/2CPH-ZYA6].

[51] Ann M. Lipton, *Not Everything Is About Investors: The Case for Mandatory Stakeholder Disclosure*, 37 YALE J. ON REGUL. 499 (2020)

has benefits inside the firm, particularly for breaking down information silos.[52] It requires the sort of cross-unit collaboration that consultancies say is a preferred method for silo busting.[53] To meet their disclosure obligations, firm leaders must "set up reporting systems to receive information about the issues addressed in the [report] and . . . [assume a greater] role in overseeing and understanding."[54] Enhanced disclosures induce executives to become more engaged in the disclosure topic and more proactive in seeking out information necessary for oversight.[55]

("[T]here are concerns about the practicability of developing a workable structure for mandatory disclosure. Sustainability disclosures must be specific enough to provide investors and capital markets with meaningful and readily comparable information. At the same time, relevant sustainability issues vary substantially by issuer and industry, making a detailed line-item approach less feasible. This problem is not insurmountable. Some sustainability issues are arguably common to all firms. The alternative, a principles-based approach, complicates policing the accuracy of issuer disclosures and risks producing low-quality or boilerplate disclosures.").

[52] Roy Shapira, *A New Caremark Era: Causes and Consequences*, 98 WASH. U.L. REV. 1857, 1893 (2021) ("[C]an combat the problems of information silos and upward flows of information."); Ronald P. O'Hanley, *Long-Term Value Begins at the Board* (Mar. 20, 2017), https://corpgov.law.harvard.edu/2017/03/20/long-term-value-begins-at-the-board/ [https://perma.cc/U6JP-L9ZZ].

[53] William S.H. O'Neil, *Silo-Busting: Overcoming the Greatest Threat to Organizational Performance*, 11 Sustainability 6860 (2019) (recommending that firms "[c]reate cross-organizational unit programs and projects for people to stimulate collaboration" and "[c]reate communities/networks to share knowledge, practice, and experience across organizational units"); Marc S. Gerber et al., *Enhancing Controls and Procedures for Climate-Relater Disclosures*, Harvard Law School Forum on Corporate Governance (Dec. 10, 2024), https://corpgov.law.harvard.edu/2024/12/10/enhancing-controls-and-procedures-for-climate%E2%80%91related-disclosures/ [https://perma.cc/7QAX-LT2Q] ("[To meet sustainability disclosure obligations] companies should consider . . . [a]assembling and regularly communicating with cross-functional teams and external advisors to coordinate a comprehensive and harmonized approach.")

[54] Jill E. Fisch, *Making Sustainability Disclosure Sustainable*, 107 GEO. L.J. 923, 962 (2019).

[55] Erin Massey Everitt, *Sarbanes-Oxley's Officer Certification Requirement—Has Increased Accountability Equaled Increased Liability*, 6 DEPAUL BUS. & COM. L.J. 225, 235-36 (2008); Clark, *supra* note 34, at 264-65 ("The main intended effect of the statutory certification requirement is that it will 'focus the mind' of these officers and make them more diligent

"To ensure effective compliance, [large companies] need to . . . build [internal] reporting systems"[56] like the ones needed to prepare disclosures. The remainder of this Part identifies features of a disclosure regime that would help ensure firms keep themselves informed about how they use data.

Disclosures must be mandatory. Facebook's voluntary disclosures prior to the Cambridge Analytica scandal failed to yield provable, internal awareness of the company's data practices. In general, voluntary disclosure regimes are not reliable.[57] For example, corporations that make disclosures on environmental, social, and governance ("ESG") matters largely do so on a voluntary basis.[58] Voluntary systems rely on corporations' internal incentives, but those incentives often favor missing, incomplete, misleading, or low-quality disclosure.[59] "While voluntary reporting frameworks are better than nothing . . . , they do not provide the consistency, accuracy and completeness that is inherent in [mandatory] securities filings."[60] It is also often unclear what sort of liability, if any, would result from exaggerated, "washed," or otherwise false voluntary disclosures.[61]

Voluntary disclosures neither generate quality representations nor assure their distribution internally. When

---

in their reviews, and more demanding of their subordinates. Anecdotal evidence suggests that it has raised consciousness and made executives more careful.").

[56] Asaf Eckstein & Roy Shapira, *Compliance Gatekeepers*, 41 YALE J. ON REGUL. 469, 479 (2024).

[57] Fisch, *supra* note 54, at 950 ("[V]oluntary sustainability reporting is not reliable" because they "tend[] to be vague, general, or boilerplate.").

[58] Fisch, *supra* note 54, at 944 ("Most sustainability information is disclosed not in issuer financial or securities filings, but in [voluntary] standalone sustainability reports.").

[59] Eckstein & Shapira, *supra* note 56, at 519 (2024) ("Both the service providers and their direct clients have little interest in delivering better compliance gatekeeping, because the costs of the less-than-ideal compliance gatekeeping are largely externalized on public shareholders and society at large.").

[60] Jonas Kron, Senior Vice President, Trillium Asset Mgmt. LLC, Comment Letter on Concept Release on Business and Financial Disclosure Required by Regulation S-K (July 21, 2016), https://www.sec.gov/comments/s7-06-16/s70616-276.pdf [https://perma.cc/9SSL-EWPG].

[61] Fisch, *supra* note 54, at 962.

negative information is included in a report, non-standardized formats enable drafters to hide it using deceptive formatting[62] or to bury it in an excess of information.[63] In other words, voluntary disclosures will not predictably generate the sort of detailed, honest, and comprehensible internal representations that are critical for ensuring distributed firm knowledge. Furthermore, voluntary disclosures have little effect on breaking down information silos. This is because voluntary disclosures are often prepared by public relations or marketing personnel, without review by attorneys, sign-off from leadership, or oversight by auditors.[64] Generally speaking, firms "take less care in preparation of [voluntary] disclosures."[65] Mandatory disclosures improve the quality and internal distribution of information.[66]

Disclosures must require sub-certification. The disclosures Facebook had to provide under the FTC settlement required signatures from various individuals, like the CEO and different compliance officers.[67] "Sub-certifications involve having key personnel in the relevant departments certify the accuracy and completeness of the information they provide."[68] Signatories face personal liability for material misstatements. Ensuring review and accountability from knowledgeable individuals leads to "better disclosures"[69] with "reduc[ed] risk of errors or omissions."[70]

In addition to ensuring higher quality information is generated, sub-certification helps to ensure it is distributed to relevant units: "[T]he need to sign a certification certainly does

---

[62] Diamantis et al., *supra* note 17.
[63] Fisch, *supra* note 54, at 959 ("The absence of standardized disclosure requirements may lead issuers to disclose such a high quantity of information that it results in information overload.").
[64] *Id.* at 950.
[65] *Id.*
[66] Karen K. Nelson & A.C. Pritchard, *Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors*, 13 J. EMPIRICAL LEGAL STUD. 266, 287-95 (2016).
[67] Such sub-certification is not novel or unusual. For example, Section 302 of Sarbanes-Oxley requires both the CEO and the CFO to personally certify financial statements.
[68] Gerber et al., *supra* note 53.
[69] Clark, *supra* note 34, at 284-85.
[70] Gerber et al., *supra* note 56.

focus the mind."[71] Referring back to the model of firm knowledge described in Part II, disclosures should require personal sign off from corporate leadership and top managers in research, legal, and marketing. This will ensure that information about data reaches the units best poised to use the information to direct firm activity.

A state or federal authority must oversee disclosures. The FTC was explicit in its settlement with Facebook that Facebook and its individual leaders would face consequences for disclosure misstatements. Requiring that firms file disclosures with a government authority helps ensure timeliness and integrity because noncompliance becomes noticeable and actionable.[72] Statutes generally provide criminal and civil liability for lying to authorities,[73] and agency-specific statutes and regulations offer additional causes of action.[74]

A federal disclosure requirement would have the broadest reach within the U.S. economy. The SEC, which has authority to require disclosure of "material" financial information, could be a natural choice.[75] In light of the increasing value of data and

---

[71] Clark, *supra* note 34, at 284-85.

[72] Fisch, *supra* note 54, at 962 ("[T]he risk of liability for noncompliance is what gives teeth to the statutory disclosure provisions.").

[73] *See, e.g.*, 18 U.S.C. § 1001 (providing criminal liability for knowingly making false statements to a federal agency); 18 U.S.C. § 1343 (providing criminal liability for using the wires to further any fraudulent scheme).

[74] *See, e.g.*, 15 U.S.C. § 78r (providing civil liability for false statements in SEC filings); 17 CFR § 240.10b-5 ("[I]t shall be unlawful for any person . . . [t]o make any untrue statement of a material fact or to omit to state a material fact . . . in connection with the purchase or sale of any security.").

[75] TCS Industries, Inc. v. Northway, Inc., 426 U.S. 438, 449 (1976). In recent years, the SEC has extended its disclosure authority to require filings on some non-traditional items, like sustainability, Commission Guidance Regarding Disclosure Related to Climate Change, Securities Act Release No. 9106, Exchange Act Release No. 61, 469, 75 Fed. Reg. 6289, 6293-97 (Feb. 8, 2010), and cyber-security, U.S. S.E.C., *SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies* (July 26, 2023), https://www.sec.gov/newsroom/press-releases/2023-139 [https://perma.cc/G5AN-UCJP]. The SEC has also stepped up its focus on AI-related disclosures. Marsha Mogilevich et al., *SEC Comment Letter Trend: AI-Related Disclosures*, Harvard Law School Forum on Corporate Governance (Jan. 16, 2025),

recent high-profile scandals from data misuse, the SEC could plausibly determine that information about data is material to corporate finances.[76] The FTC also probably has the authority to require firms to disclose information about data,[77] and the Commerce Department's Bureau of Industry and Security has shown some appetite to require technology-focused reporting.[78] In the near term, however, federal agencies are unlikely to implement new disclosure requirements in light of the current administration's "deregulatory focus."[79]

State-level intervention is the more likely path forward for now.[80] Because of its pull with the tech industry and its first-in-

---

https://corpgov.law.harvard.edu/2025/01/16/sec-comment-letter-trend-ai-related-disclosures/ [https://perma.cc/KB55-ZJUL].

[76] *See generally* Diane Lourdes Dick & Joseph W. Yockey, *Governing Toxic Data*, 28 YALE J.L. & TECH., SYMPOSIUM ISSUE 2026 (discussing the value and risks to corporations of holding massive amounts of consumer data).

[77] Diamantis et al., *supra* note 14.

[78] Harry Clark et al., *Proposed AI Reporting Requirements: Key Takeaways for Companies*, Harvard Law School Forum on Corporate Governance (Oct. 25, 2025), https://corpgov.law.harvard.edu/2024/10/15/proposed-ai-reporting-requirements-key-takeaways-for-companies/ [https://perma.cc/44QC-EATS].

[79] Amelie Champsaur et al., *A New Regulatory Environment for Climate and Other ESG Reporting Rules*, Harvard Law School Forum on Corporate Governance (Feb. 8, 2025), https://corpgov.law.harvard.edu/2025/02/08/a-new-regulatory-environment-for-climate-and-other-esg-reporting-rules/ [https://perma.cc/LPD9-DHBG]. While cybersecurity and sustainability disclosures were aggressively enforced under Biden, they are being sidelined or abandoned under Trump. Jennifer Lee & Charles Riley, *SEC Priorities Regarding Cybersecurity Enforcement in the Second Trump Administration*, Harvard Law School Forum on Corporate Governance (Mar. 26, 2025), https://corpgov.law.harvard.edu/2025/03/26/sec-priorities-regarding-cybersecurity-enforcement-in-the-second-trump-administration/ [https://perma.cc/9D7W-FSWA]; U.S. S.E.C., *SEC Votes to End Defense of Climate Disclosure Rules* (Mar. 27, 2025), https://www.sec.gov/newsroom/press-releases/2025-58 [https://perma.cc/D4RC-MLVV]; Bloomberg Law, *SEC Abandons ESG Enforcement Group Amid Broader Backlash* (Sept. 12, 2024), https://news.bloomberglaw.com/esg/sec-quietly-dissolves-climate-and-esg-enforcement-task-force [https://perma.cc/54X2-LA7J].

[80] Beth Sasfai et al., *Climate and Sustainability Regulations: 2024 End-of-Year Review*, Harvard Law School Forum on Corporate Governance (Jan. 25, 2025), https://corpgov.law.harvard.edu/2025/01/26/climate-and-sustainability-regulations-2024-end-of-year-review/ [https://perma.cc/M8BC-PZMU] ("companies do not expect to report under the SEC climate rules in the near future. . . . [M]any companies will

nation data privacy standards, California has been a norm leader in tech regulation.[81] The state already requires limited disclosures from some firms to state authorities[82] and sweeping disclosures to consumers from all firms that meet minimum thresholds.[83] Broader disclosure to authorities would be a natural extension of California's existing practice that could spread to other states.

The disclosure instrument must be suited to its function. Any governmental body that implements the disclosure regime we propose will face the challenge of preparing a disclosure mandate that forces internal distribution of information about data while neither provoking too much firm pushback nor inviting gamesmanship. In prior work, we have proposed detailed, formulaic disclosure mandates that create an official representation of firms' data use policies that can be easily audited by external authorities.[84] Such disclosures would not be suitable here. The present goal is different in two ways. First, the goal is internal distribution of information rather than external disclosure for its own sake. Second, the subject of disclosure is more amorphous—how firms actually use data, rather than their official data policies.

Effective disclosures that meet these two goals must consider "who are the users of the report and what decisions are they assumed to be making."[85] "[E]nsur[ing] that key messages are hitting home, particularly if concepts are particularly technical or complicated," requires striking a

---

nonetheless find themselves subject to ongoing stakeholder and regulatory pressure to report on climate and other sustainability topics, including upcoming reporting under [EU and California law]."); Gerber et al., *supra* note 53.

[81] *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 39 (6th ed. 2018) (nothing that other states follow California on privacy legislation); *see also* DAVID VOGEL, TRADING UP: CONSUMER AND ENVIRONMENTAL REGULATION IN A GLOBAL ECONOMY (1995) (describing California's ability to influence policy in other states).

[82] California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.99.82(b)(2)(A) (West 2018).

[83] *Id.* § 1798.140(d).

[84] Diamantis et al., *supra* note 14.

[85] Richard Barker, *Corporate Sustainability Reporting*, 49 J. ACCOUNTING & PUB. POL'Y 1, 16 (2025)

balance.[86] Highly technical and lengthy disclosures may be best for accuracy and completeness, but they come at the cost of comprehension by non-technical audiences.[87] Overly technical representations risk fortifying the information silos that the disclosures are meant to break down. It is important to find a "common language" for technical information to ensure it is "translated . . . across the company."[88]

There are different models available for balancing completeness with accuracy, and uniformity with the idiosyncrasies of different firms' data use. One regime could be modeled off SEC cybersecurity disclosures. Recognizing that firms necessarily vary widely in their approach to cybersecurity, these disclosures eschew rigid templates while still requiring disclosure of cybersecurity risk management and strategy, including processes for identifying, assessing, and managing risks.[89] Actual firm disclosures can get quite detailed. Nearly all disclosures by covered firms "discuss specific components of the company's cybersecurity program, which most prominently include references to identity and access management, logging, monitoring, penetration testing and vulnerability scanning, governance, risk and threat intelligence, employee awareness and training, and security monitoring."[90]

---

[86] Kuranda Vallone & Frimpong, *supra* note 34.

[87] *Id.* Cost is another point to bear in mind when it comes to lengthy disclosures. Fisch, *supra* note 54, at 959-60 ("Increasing the number of issues addressed, requiring issuers to provide hard sustainability data, and formulating line-item disclosure requirements would potentially increase the informational content of sustainability disclosure. That increase would come, however, at a substantial cost both to issuers preparing the information and to investors relying on it."). Increased disclosure can negatively impact corporate innovation due to direct costs of preparing and indirect costs (e.g. competitors getting access to info or impact of regulation on firm investment). Shiu-Yik Au & Hongping, *Too Much of a Good Thing? Mandatory Risk Disclosure and Corporate Innovation*, 50 Tan, J. of Acc. & Pub. Pol'y 107292 (2024).

[88] Cox & Kanji, *supra* note 33 (citing Principles for Responsible Investment, *Engaging on Cyber Security: Results of the PRI Collaborative Engagement* 2017-2019 (Apr. 22, 2020), https://www.unpri.org/cyber-security/engaging-on-cyber-security-results-of-the-pri-collaborative-engagement-2017-2019/5680.article).

[89] 17 CFR § 229.106.

[90] Gosnell Handler et al., *Cybersecurity Disclosure Overview: A Survey of Form 10-K Cybersecurity disclosures by S&P 500 Companies*, Harvard Law

Despite this detail, the average length of a cyber-security disclosure is a modest 980 words, with disclosures over 2000 words being extremely rare.[91]

Another possible model for the disclosures we propose is the management discussion and analysis (MD&A) section of public corporations' annual and quarterly financial statements. "This section . . . allows company management to tell its story in its own words."[92] MD&A disclosures typically involve a narrative discussion of a firm's financial health. As a principles-based approach to disclosure, the MD&A tends to be "less detailed" and more flexible to adapt to a broader range of firm's circumstances. [93] Jill Fisch has proposed them as a template for sustainability disclosures. She would require covered firms "to identify and explain the three sustainability issues most significant to their operations."[94] While most firms will face far more than three sustainability issues at any given time, determining which three are most significant necessarily entails analysis and reflection on the rest.

An analogous approach to disclosing information about data might ask firms to identify the three most significant data practices that important constituents within the firm are most likely to misapprehend. Paired with an expectation that firms identify remedial measures they commit to undertake, such

---

School Forum on Corporate Governance (Jan. 9, 2025), https://corpgov.law.harvard.edu/2025/01/09/cybersecurity-disclosure-overview-a-survey-of-form-10-k-cybersecurity-disclosures-by-sp-100-companies/ [https://perma.cc/AP5P-EBZG].

[91] Gosnell Handler et al., *Cybersecurity Disclosure Overview: A Survey of Form 10-K Cybersecurity disclosures by S&P 500 Companies*, Harvard Law School Forum on Corporate Governance (Jan. 9, 2025), https://corpgov.law.harvard.edu/2025/01/09/cybersecurity-disclosure-overview-a-survey-of-form-10-k-cybersecurity-disclosures-by-sp-100-companies/ [https://perma.cc/NX5L-L29A].

[92] https://www.sec.gov/files/reada10k.pdf [https://perma.cc/ZF87-F447].

[93] Business and Financial Disclosure by Regulation S-K, Securities Act Release No. 10,064, Exchange Act Release No. 77,599, 81 Fed. Reg. 23,916, 23,946-47 (proposed Apr. 22, 2016); Fisch, *supra* note 54, at 955 ("[T]he MD&A . . . disclosures are primarily principles-based. As such, they offer flexibility that both permits tailoring the disclosures to the issuer's particular circumstances and allows the disclosures to evolve in response to changes in issuer and market conditions.").

[94] Fisch, *supra* note 54, at 956.

disclosures would ensure that firms generate and distribute information about data by focusing on higher-priority items first and working incrementally each year toward completion. Borrowing heavily from 17 CFR § 229.303, which sets out the requirements for the 10-K MD&A section, an annual disclosure tailored to information about data might ask covered firms to:

> [P]rovide material information relevant to the three most significant data misuse risks from operations of firm, including an evaluation of the amounts and certainty of the risk of data misuse from operations. The discussion and analysis must focus specifically on material events and uncertainties known to managers within technical, legal, and marketing units of the firm that are reasonably likely to cause actual data use to depart from reported data use policies or known legal requirements.

By requiring input from technical, legal, and marketing units, the process of preparing such a disclosure would necessarily entail cross-unit discussion of data misuse risks. This disclosure would be relatively low-cost and low-risk for firms. It does not ask for a comprehensive catalogue of data misuse risks; just the top three. As for MD&As, responses would be narrative and not subject to traditional audit.[95] This lowers the temperature for disclosing firms.

At the same time, the disclosures would not be toothless. Narrative disclosures require active engagement since they cannot be produced through formulaic, tick-the-box exercises.

---

[95] Felicia H. Kung, *The Regulation of Corporate Bond Offerings: A Comparative Analysis*, 26 U. PA. J. INT'L ECON. L. 409, 445 (2005) ("MD&A, in contrast, consists of 'soft information,' such as management's analysis about the company and its prospects, which is not easily subjected to what is normally viewed as an audit."); KC Goyer, *Nancy Temple's Duty: Professional Responsibility and the Arthur Andersen Verdict*, 18 GEO. J. LEGAL ETHICS 261, 275–76 (2004) ("Management writes the MD&A, and the independent auditor is obligated to read it because it accompanies audited financial statements. However, the auditor does not audit the MD&A, but only reads it for material inconsistencies or material misstatements of fact.").

Gamesmanship is hard to envision. Firms that disclose the same risks year to year would raise red flags by signaling that they have not taken measures to address known risks. Signatories have personal skin in the game too, since they expose themselves to personal liability for knowing misstatements or half-truths.[96]

Conclusion

Deterrence requires self-awareness. However large or certain the sanction, no one can align their conduct with legal requirements if they do not know what their conduct is. Such self-knowledge is not inevitable. For many firms that handle large amounts of personal data, we have argued that ignorance about how they use that data may be business as usual. Information does not flow of its own accord within a firm. All operational information tends to get stuck in firm silos, and information about data is no exception.[97] Ensuring that information about data flows to the units that need it takes deliberate effort. Our survey of technical, marketing, and legal personnel demonstrates that small and large firms are most likely to face challenges. Cross-unit communicative integration requires a culture of trust where employees are confident in their firm's governance structures. Unfortunately, firm culture is notoriously hard to engineer.[98] This will be a disheartening result for privacy advocates who hope that robust enforcement of data laws can provoke more responsible data stewardship.

While authorities may not be able to force cultural change, they do have more direct tools to force information to flow within firms. Mandatory disclosures are a familiar fixture of firm regulation. As traditionally conceived, disclosures compel firms to share operational information (e.g. about finances or environmental impact) with external parties. But disclosures

---

[96] *See* Macquarie Infrastructure Corp. v. Moab Partners, L.P., 601 U.S. __ (2024) (holding that officers are not liable for pure omissions in MD&A filings, but are liable for misleading half-truths).

[97] *Supra* nn. 24-30 and accompanying text.

[98] Michelle M. Harner, *Barriers to Effective Risk Management*, 40 SETON HALL L. REV. 1323, 1357 (2010) ("[C]hange [of corporate culture] is hard and slow.").

can also force information to flow *within* a firm. The process of preparing and validating a disclosure requires coordination and exchange across multiple units. Furthermore, by mandating sub-certification by representatives of different firm units, disclosure regimes can require the information in the disclosure to circulate internally. Suitably constructed mandatory disclosures could ensure that firms have the self-awareness needed for authorities to be able to deter data misuse.