BROWN UNIVERSITY

SCHOOL OF PROFESSIONAL STUDIES

DATA GOVERNANCE AND CYBERSECURITY CROSSROADS

CHAD THIEMANN
FEBRUARY 2019

A Critical Challenge Project

submitted in partial fulfillment of the requirements for an

Executive Masters in Cybersecurity

Reviewed and approved by the following:

Dr. Ravindra Pendse
VP for Information Technology & CIO and Professor of Computer Science
University of Michigan
Adjunct Professor - Brown University School of Professional Studies
CCP Faculty Advisor

Frank Price
Vice President and Chief Information Security Officer (CISO) - CVS Health
CCP Co-Advisor

Abstract

The recent proliferation of technology (i.e., internet usage, mobile devices, or IOT devices) has resulted in an unprecedented accumulation of data. Complete, timely, secure, and accurate data is essential for both public and private organizations across varied industries and disciplines to operate effectively and efficiently in order to achieve their goals. Additionally, organizations increasingly rely on the use of confidential data to facilitate processes which require them to maintain privacy and security over this data, along with meeting a continually expanding list of compliance obligations. The Data Management Association International defines data governance as, "The exercise of authority and control over the management of data assets. It is the planning, supervision, and control over data management and use."[1]

While data governance can include a variety of data management functions, for the purposes of this critical challenge project, I will focus on the functions of data privacy and data security management. The intent of this effort is to provide an overview of this topic, the current state of data governance today, its risks and implications, and to provide key principles, guidelines, and controls to manage the security and privacy of confidential data while meeting operating objectives and compliance obligations. Data governance is also essential to creating a sense of trust and confidence for internal data users who need accurate, timely, and insightful information, as well as amongst one's customers who demand security and the responsible privacy-minded use of their data.

[1] DAMA International, *The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK)*, 2nd Edition, Bradley Beach, New Jersey: Technics Publishing, 2010

TABLE OF CONTENTS

Data Governance Background & Overview

Data governance, as it pertains to security and privacy, comes down to three primary questions:

1) Do you know where all your data is?
2) Is it classified appropriately given its value, sensitivity, and regulatory obligations?
3) Are security and privacy controls implemented in accordance with its classification?

If an organization is unable to answer these three fundamental questions related to their business critical or confidential data, they will be unable to successfully execute efficient and effective data governance over said data. Without data governance, the application of security and privacy controls can become a "shot in the dark" as opposed to a precision risk-based application of finite resources. The primary objectives of a data governance program as it pertains to the management of privacy and security considerations include:

- Protecting an organization's data against internal and external security and privacy threats.
- Ensuring the organizations are compliant with relevant laws, regulations, and contractual commitments.
- Ensuring that data governance processes include an embedded capability to document proof of compliance in anticipation of audits or investigations.
- Identify data stakeholders, establish decision rights and clarify accountability.[2]

Every system, electronic or manual, that generates, collects, stores, transmits, uses, archives, and dispositions data and information must be governed with security and privacy protections to facilitate trust, integrity, and transparency.[3]

---

[2] Helen Sun (May 2011). *Enterprise Information Management: Best Practices in Data Governance – A White Paper on Enterprise Architecture* (Online). Available: https://www.oracle.com/assets/oea-best-practices-data-gov-1357848.pdf

[3] AHIMA (2017). *Information Governance – Principles for Healthcare (IGPHC)™* (Online). Available: http://www.ahima.org/topics/infogovernance

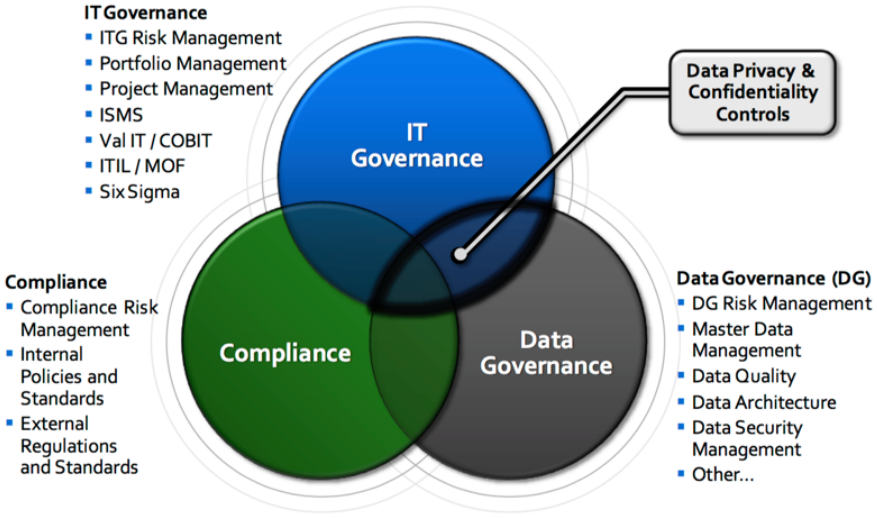
Some common goals of a data governance program, in general, include defining, approving, and communicating data strategies, policies, standards, architecture, procedures, and metrics. Programs will also routinely have capabilities to track and enforce compliance with said governing standards as well as serve to manage and mediate data related issues. Considerations must be given to data governance processes, standards, and controls and the fact they must integrate with those of the information security and privacy teams as well. Data governance cannot exist on an island. While it typically takes a top-down enterprise approach, data governance is woven into many direct, ancillary, and peripheral processes involving both internal and external stakeholders.

Some of the more common data governance standards include the Data Management Association International (DAMA) - Data Management Body of Knowledge (DMBOK); The Open Group Architecture Framework (TOGAF); Object Management Group; ISO 15489-1:2016 Information & Documentation – Records Management; ISO 8000 - Standard for Data Quality and Enterprise Master Data, and COBIT 5. There are also a variety of vendor derived data governance standards (i.e., Microsoft's Trustworthy Computing – Data Governance for Privacy, Confidentiality & Compliance or IBM's InfoSphere), typically aligned to various products or services they offer in this vertical. One of the aims of this project is to highlight the best concepts from various sources and to demonstrate how a hybrid approach that aligns to one's data, architecture, and processes is the most effective path. A hybrid model serves great benefits when one considers that data governance often involves applying these concepts to existing systems and processes that were not created with the related objectives, goals, and considerations in mind. Start-up companies have the benefit of creating a data governance program from the ground up, in advance of intaking data itself. Large, complex companies that have been around

for years and have witnessed numerous mergers and acquisitions have a much more complex road to navigate. For them, the cost and level of effort necessary to reach the higher echelons of data governance maturity will be considerably higher and take more time than a young or nascent start-up organization.

Ultimately, this Critical Challenge Project looks to highlight the considerations, implications, and alternatives when designing and deploying a data governance program and its interface with data protection (i.e., privacy and security) practices. The graphic below, from Microsoft's Trustworthy Computing Framework, pictorially represents the intersection between data governance, compliance, and IT governance – highlighting the critical security and privacy segment discussed here.

Intersection of Data Privacy & Security Controls



The Intersection of Data Privacy, Security & Governance [4]

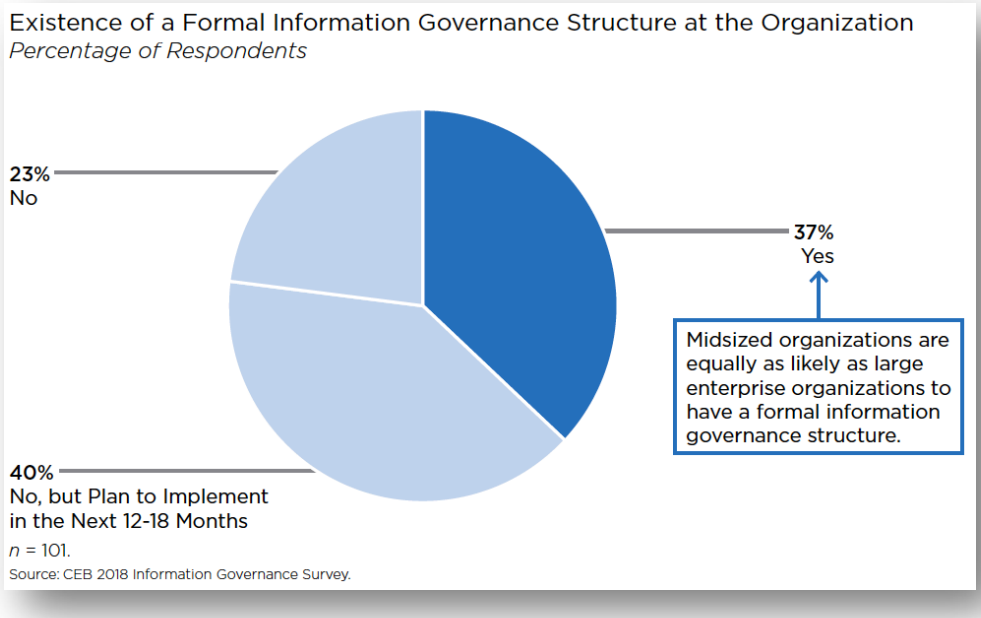
To better understand the implications of data governance in today's data-rich organizations, one must delve into the current state of data governance deployments, maturity,

[4] Javier Salido & Patrick Voon (January 2010). *A Guide to Data Governance for Privacy, Confidentiality, and Compliance – Part 1 The case for Data Governance* (Online). Available: https://iapp.org/media/pdf/knowledge_center/Guide_to_Data_Governance_Part1_The_Case_for_Data_Governance_whitepaper.pdf

and influencers. The following section will provide insights into the current state of data governance through an analysis of recent surveys, benchmarking efforts, and narratives.

The Current State of Data Governance

Through a corporate partnership with CEB Gartner, I was able to serve as a contributor and participant in their 2018 Information Governance Survey and Benchmarking Exercise – a project they execute biennially – enabling direct coordination with my Critical Challenge Project (CCP) objectives. This enabled direct access to the current state of data governance efforts today. For the purposes of this CCP, I will refer to (and reference) data and information governance interchangeably. The most resounding observation from the CEB Gartner exercise was the fact that while organizations are in the middle of digital transformations and they are amassing data at rates and quantities unseen before, only 37% of survey organizations claimed to have a formal data governance program in place.



Formal Data Governance Program Existence Metrics [5]

[5] CEB Gartner (2018). *Information Governance Benchmarking – 11 Key Findings* (Online). Available: https://www.cebglobal.com/member/data-privacy/

While many key findings of the aforementioned CEB Gartner Information Governance Survey (along other third-party benchmarks) are included throughout the following analysis, additional insights can be found in the appendices at the end of the report.

Another major lesson learned from this initiative is the fact that ownership of data governance is fairly fragmented with no clear or consistent ownership model. Information Technology (IT), Information Security, and Legal typically own most of data governance activities with Privacy and Compliance serving as the next greatest contributors. This fragmentation is demonstrated in the table below.

Owners of and Participants in Information Governance Activities

Percentage of Respondents

	Activities	Owners	Participants
Understand Information Assets	Assessing and Classifying the Organization's Data	IT (22%)[1]	Information Security (31%), **Legal (34%)**, **Privacy (34%)**[2]
	Inventorying the Organization's Data	IT (31%)	Information Security (22%), **Legal (19%)**, **Privacy (19%)**
	Creating and Maintaining an Accurate Data Map	IT (44%)	Information Security (34%), **Legal (22%)**, Business (19%)
Incident Management	Creating and Maintaining the Data Breach Response Plan	Information Security (41%)	**Legal (56%)**, IT (50%), **Privacy (22%)**, **Compliance (22%)**
	Responding to Potential Incidents	Information Security (47%)	**Legal (66%)**, IT (44%), **Privacy (34%)**
Business Support	Training the Business on Acceptable Collection and Use	**Privacy (22%)**	**Legal (41%)**, **Compliance (28%)**, Information Security (25%)
	Translating Standards Into Procedures	**Privacy (16%)**, IT (16%)	**Legal (31%)**, **Compliance (28%)** and Information Security (28%)
Ongoing Monitoring	Monitoring Information Collection, Use and Access	IT (31%)	**Legal (28%)**, **Privacy (25%)**, Information Security (25%)
	Monitoring Third-Party Information Access	Information Security (28%)	**Legal (31%)**, IT (19%), **Compliance (16%)**, **Privacy (16%)**

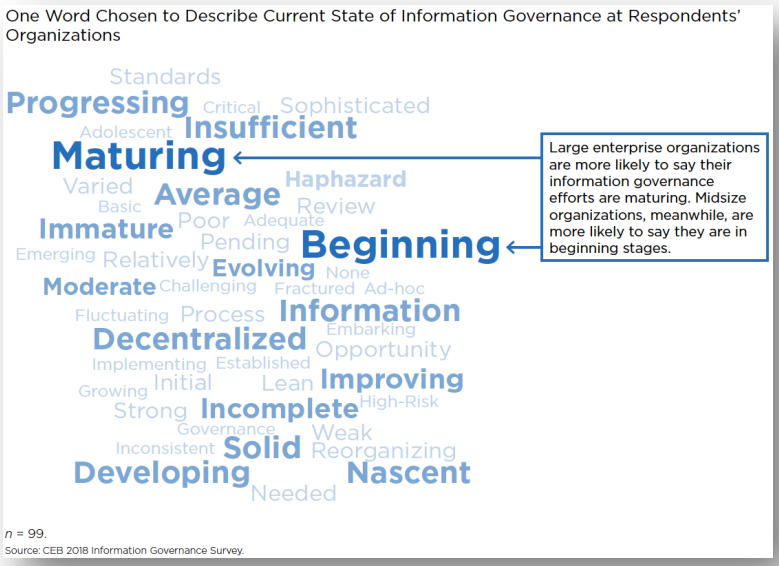
n = 32.

Source: CEB 2018 Information Governance Survey.

Data Governance Ownership & Participant Survey Metrics [6]

[6] CEB Gartner (2018). *Information Governance Benchmarking – 11 Key Findings* (Online). Available: https://www.cebglobal.com/member/data-privacy/

Survey participants agreed that significant investment in digitization is resulting in massive growth in the amount of data organizations collect, use, and retain.[7] Other influencing factors such as the changing regulatory environment, customer scrutiny, and demand for transparency are driving organizations to invest in data governance now. A majority of legal, compliance, and privacy executives surveyed, and their Boards, view improving data governance as a critical priority. Executives also shared the common desire to have data governance efforts drive agreement around a set of standards and processes to more effectively govern the collection, use, sharing, and retention of data. The CEB Gartner benchmarking survey measured the overall satisfaction score for data governance efforts is 4.06 out of 7, meaning organizations are only moderately happy with their current frameworks and efforts.[8] The word map below is a pictorial representation of words used to describe the current state of data governance efforts.



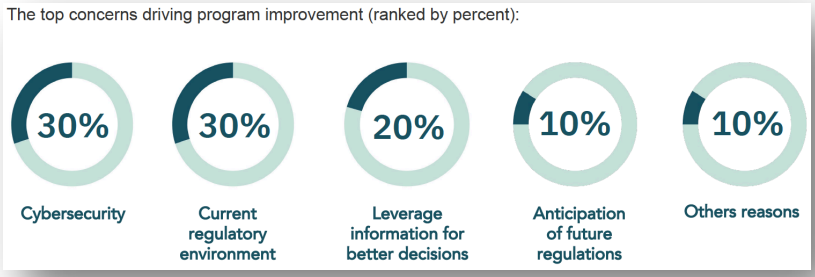
Information Governance 'Current State' Word Map [9]

[7] Ibid.

[8] CEB Gartner (2018). *Information Governance Benchmarking – 11 Key Findings* (Online). Available: https://www.cebglobal.com/member/data-privacy/

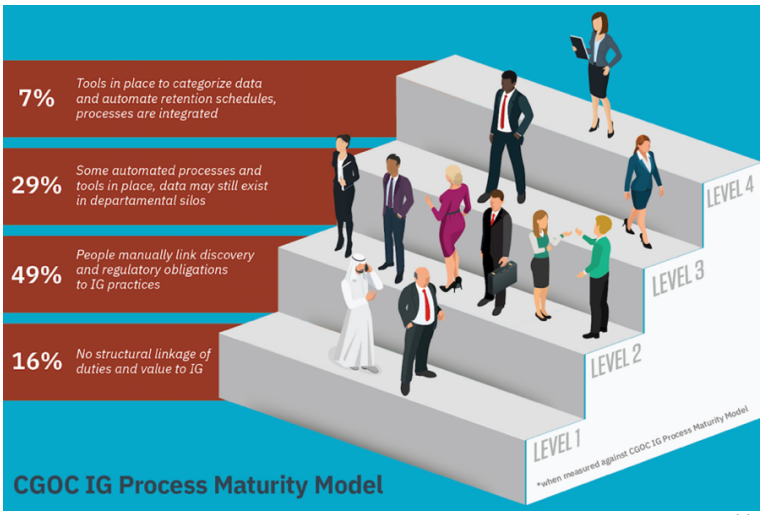
[9] Ibid.

The Compliance, Governance and Oversight Council (CGOC) identified the following key drivers, in their 2018 Information Governance Benchmark Report, that are spearheading momentum for organizations to either implement or improve their data governance programs or processes as the following,



Top Concerns Driving Data Governance Program Improvement [10]

Below is another graphic representation of the CGOC Council's benchmark survey results detailing where respondents believed their organizations fell within the CGOC's information governance process maturity model.



CGOC Data Information Governance Process Maturity Model [11]

[10] CGOC Council (2018). *CGOC Information Governance Benchmark Report 2018* (Online) Available: https://www.cgoc.com/resource/cgoc-information-governance-benchmark-report-2018/

[11] Ibid.

While progress is being made with the advancement of data governance, this analysis demonstrates that work remains to be done. The existing data governance gap is exacerbated by the fact that data creation and acquisition is growing at a massive rate along with the simultaneous proliferation of new data privacy and information security regulations and obligations. This "perfect storm" demands that organizations take steps now to implement, or mature, data governance programs in an effort to remain competitive and compliant. The following pages will delve into the "three questions" and what organizations can do to address them via the hybrid application of various existing and emerging tactics, techniques, and procedures.

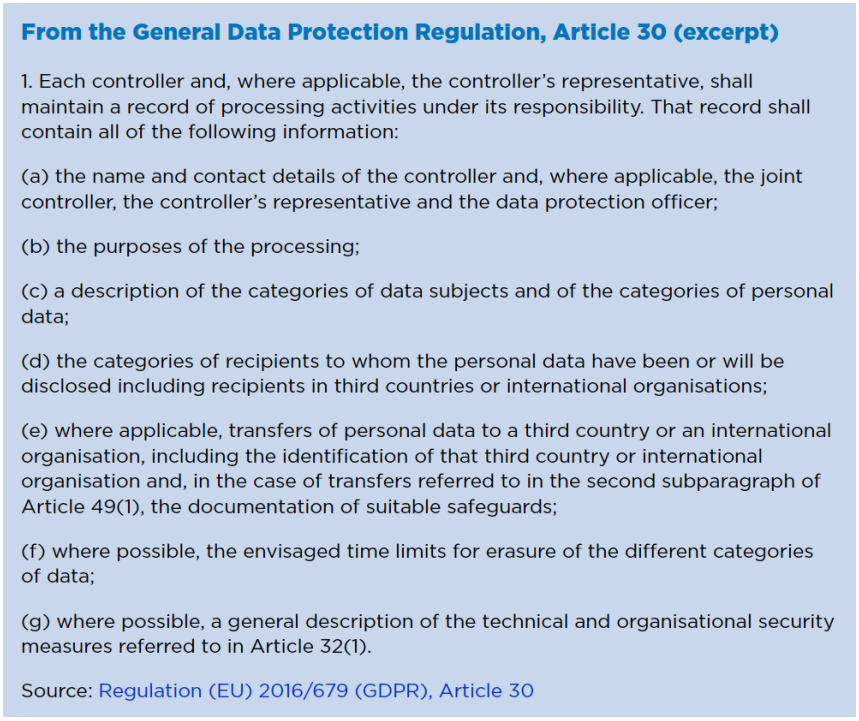
Data Inventory and Data Flows

Every organization should conduct and maintain a data inventory. A data inventory is a centralized inventory of all data an organization acquires, stores, processes, or transmits. It includes both structured and unstructured data, as well as, both electronic and paper-based data. The primary purpose of a data inventory is to understand what data is traversing an organization and what controls are necessary to efficiently and effectively manage the security and privacy obligations associated with the data. There are other data governance aspects that benefit from a data inventory, but as this analysis is limited to security and privacy implications, we will not cover those within this analysis. Robust data inventory practices aid in creating enterprise data that is accessible, timely, accurate, complete, secure, and utilized in a compliant and responsible manner. Cybersecurity teams spend significant effort securing and protecting an organization's data, however, these efforts are diminished (or fruitless) if one doesn't know where all the data requiring protection exists. This assertion is re-enforced by a 2017 ISACA survey where it was determined that, "Fifty percent of information security professionals worry that they do not understand the full extent of the data risk of their enterprise."[12]

Regulations are a key driver for maintaining a complete and timely data inventory. Some regulations require a data inventory outright. Other regulations may not require the explicit need for an inventory, however, without one, an organization will be unable to effectively deploy security and privacy controls given they do not know where the data requiring protection exists. One of the most recent (and robust) regulations requiring a data inventory is the EU's General Data Protection Regulation (GDPR). In it, Article 30 demands what the GDPR refers to as a,

[12] Guy Pearce (2017). *Boosting Cyber Security With Data Governance and Enterprise Data Management* (Online). Available: https://www.isaca.org/Journal/archives/2017/Volume-3/Documents/Boosting-Cyber-Security-With-Data-Governance_joa_Eng_0517.pdf

"Record of Processing Activities," or ROPA, the details of which are summarized in the graphic that follows. The GDPR uses a different naming convention, but they are basically requiring a data inventory to be executed, documented, and maintained.



From the General Data Protection Regulation, Article 30 (excerpt)

1. Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information:

(a) the name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer;

(b) the purposes of the processing;

(c) a description of the categories of data subjects and of the categories of personal data;

(d) the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations;

(e) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;

(f) where possible, the envisaged time limits for erasure of the different categories of data;

(g) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).

Source: Regulation (EU) 2016/679 (GDPR), Article 30

GDPR Article 30 Excerpt [13]

Another regulation that is on the horizon that puts a big emphasis on data inventory processes is the California Consumer Privacy Act of 2018 (CCPA). The International Association of Privacy Professionals (IAPP) has assessed that CCPA demands that organizations should, "prepare data maps, inventories or other records of all personal information pertaining to California residents, households and devices, as well as information sources, storage locations, usage and recipients, to add newly required disclosures to privacy policies, to prepare for data access, deletion, and portability requests, to secure prior consent for data sharing from parents and minors and to

[13] CEB Gartner (2018). *Ignition Guide to Data Classification* (Online). Available: https://www.cebglobal.com/member/data-privacy/

comply with opt-out requests to data sharing."[14] Additionally, some regulations, such as HIPAA or GLBA, point to the necessity for data inventory practices indirectly through the mandate to conduct security risk assessments. These risk assessments inquire, or assess, whether or not the organization maintains a timely and accurate data inventory to facilitate compliance efforts. Even the U.S. Government requires data inventories, as relayed to executive departments and agencies via the President and his Office of Management and Budget in a memo that states, "To protect Federal Government information and assets, agencies must first identify the value and impact of the information on their systems and networks. Agencies must also identify the IT assets used to store, process, and transmit that information. Further, agencies must identify those assets and capabilities that enable mission essential functions and ensure the delivery of critical services to the public."[15]

While the concept of a data inventory sounds fairly straightforward, achieving this goal is more of a long and winding path. Given the nature of today's distributed systems, on-premises versus cloud and hybrid information technology infrastructures, jointly held data, the emergence of IoT devices, and the ease with which data analytic platforms lead to data repudiation, the task of implementing and maintaining a data inventory is a significant effort.

To begin this journey, it is imperative to obtain buy-in from executive leadership as well as cross-functional senior leaders who will have a stakeholder role in the initiative. Data inventory and mapping projects require collaboration across various functions and lines of business to drive success. While Privacy and IT teams typically lead these initiatives, other

---

[14] Lothar Determann (July 2018). *Analysis: The California Consumer Privacy Act of 2018 - Broad Data and Business Regulation, Applicable Worldwide* (Online). Available: https://iapp.org/news/a/analysis-the-california-consumer-privacy-act-of-2018/

[15] Shaun Donovan & Tony Scott (October 20, 2015). *Memorandum for Heads of Executive Departments and Agencies - SUBJECT: Cybersecurity Strategy and Implementation Plan* (CISP) for the Federal Civilian Government (Online). Available: https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2016/m-16-04.pdf

groups have important value-add roles, such as the following: legal, information security, compliance, human resources, risk management, finance, quality, records retention, internal audit, sales and marketing, research and development, and key business unit data creators/consumers. If an organization shares significant data with key vendors or partners, they too may need to be included to ensure collaboration is maintained and compliance enforced.

When beginning the data inventory exercise, it is best to start with what you already know and to target high-risk and high-volume data assets. Existing resources such as record retention schedule or the organization's configuration management database are good places to start. This will allow for a "quick win" and a foundation with which the team can build upon. The data elements, or criteria, that is to be collected and documented should be aligned, stored, and communicated via a data inventory tool or form. Some examples of pertinent inputs to collect include the following: application name, server name, IP address, data elements/fields, data classification(s), data subjects (i.e., employee, customer, etc.), governing consent/notice, those with access to the data, information asset steward, data origination or acquisition (internal, third-party, geography, etc.), data storage location, data transfer locations and mechanisms (i.e., where does it go and via what means or authorization). Data elements included in a data inventory can be quite exhaustive if desired. Given limited resources, organizations should balance inventory breadth with whether or not the data captured in the inventory adds values and/or meets a compliance obligation.

Data inventories are often conducted initially via survey mechanisms. As such, there should be a robust training and awareness programs targeting survey participants to ensure they are informed and savvy enough to effectively execute the survey. Additionally, the project management team should have mechanisms to verify the information submitted via the survey

process (i.e., trust but verify). To augment the self-reporting nature of the survey process, organizations can utilize data loss prevention (DLP) solutions or emerging indexing software (such as Nuix, Guidance or Active Navigation) to pro-actively seek out data in systems that align with pre-configured data classifications. One must be sure to scan all IT assets and to search both structured and unstructured data stores, even assets such as your centralized logging suite, web analytics tools, or SFTP servers – no IT asset class should go un-surveyed or uninterrogated. Furthermore, the inventory effort will need to partner with the third-party risk management program and the procurement organization in an effort to identify all data stored outside the organization's physical assets. Data may reside with vendors, partners, in SaaS solutions, or in the organization's IaaS cloud environments. One vector that must not be overlooked is the potential unsanctioned use of public collaboration tools such as Google Docs or Dropbox. These tools can lead to rapid and significant data leakage and the potential inability to fully comprehend everywhere the organization's data resides. Processes that still utilize paper records must also be in-scope and be put through the same survey efforts at the digital data. Finally, the data inventory exercise can further augment the survey process and scanning tools with a third technique, interviews. The execution of interviews by the project team can be time and labor intensive but often result in a deeper understanding of data elements, usage, and flows. As one can see, in a large complex organization, this entire process can involve hundreds of constituents across numerous disparate teams. Therefore, the project team should consider executing the data inventory in phases to permit timely and quality support of the initiative. As the inventory process matures, and lessons are applied, the process can be scaled up and deployed in a broader manner.

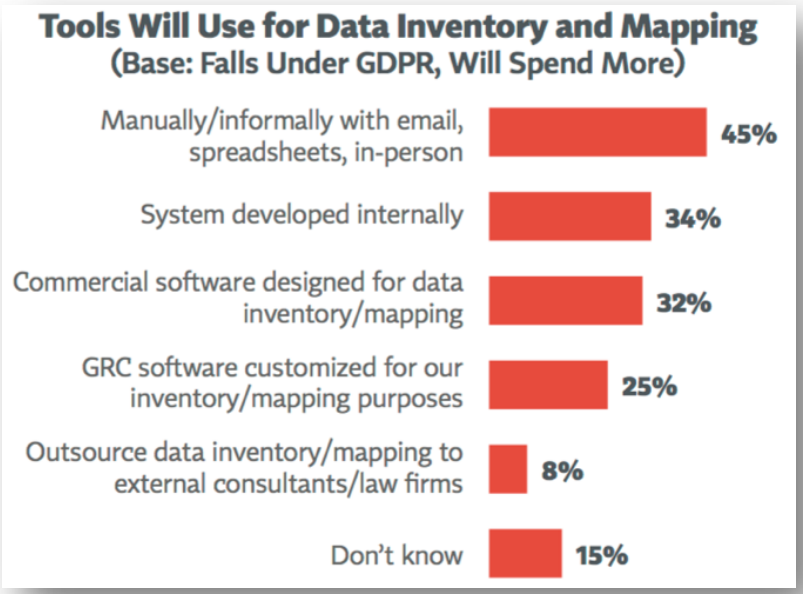
A common goal within data governance programs is to supplement the static nature of a data inventory with the more dynamic aspects of data mapping. It is pertinent to ensure that people understand the nuanced differentiation between a data inventory and a data map. Data inventories detail the information about the data assets, including what data assets the organization possesses, by whom they are owned, and the classification of said data. On the other hand, data maps detail where data assets are stored and how they traverse through your organization and third parties. Data maps are a vital tool that enables Legal and Privacy professionals to identify, manage, and monitor complex data flows that can cross not only organizational borders but also international borders that demand varied compliance obligations. The EU's GDPR regulation (Article 30) requires that, "among other records, details pertaining to data transfers must be kept electronically and in a format that can be made available to a supervisory authority upon request."[16] Additionally, Federal Rules of Civil Procedure (FRCP) require that companies in federal litigation provide a data map along with their initial disclosures. Data mapping can also result in faster, more efficient e-discovery efforts, as well as, more refined Data Subject Access Request (DSAR) fulfillment operations under GDPR.

Data mapping is far more difficult to identify and document in comparison to a static data inventory. Complex network topologies that are routinely in flux along with limited visibility and potential negative impacts on network bandwidth are just a few examples of the tertiary convolution involved with data mapping efforts. However, data mapping can have benefits that go beyond data governance. Visibility around the flow of sensitive data across a network can

[16] CEB Gartner Data Privacy Leadership Council (2018). *Ignition Guide to Data Inventory and Mapping* (Online). Available: https://www.cebglobal.com/member/data-privacy/

help facilitate behavioral access management, enable the ability to implement, or refine, network segmentation, or enhance SIEM monitoring and alerting capabilities.

There are a variety of emerging automated tools to assist data inventory and mapping efforts. Below is a chart detailing data inventory and mapping tools and their recent utilization for GDPR related efforts.



GDPR Data Inventory & Mapping Tool Usage [17]

Additionally, the IAPP has recently started publishing a "Privacy Vendor Tech Report" that highlights technology solutions that are emerging to address a variety of privacy-related demands. Data inventory and mapping tools are entering the market at a staggering rate (in excess of 200% per year). On the following page is the IAPP's Tech Report's current list of data discovery (i.e., inventory) and data mapping solutions.

[17] IAPP & EY (2017). *IAPP-EY Annual Privacy Governance Report 2017 (*Online). Available: https://iapp.org/media/pdf/resource_center/IAPP-EY-Governance-Report-2017.pdf

DATA DISCOVERY TOOL VENDORS:		
Advanced Metadata	HexaTier	Qixium
Alation	Immuta	Raptor Compliance
Automated Intelligence	Indica	Rever
AvePoint	Informatica	RISMA Systems
BigID	Information First	SAP
CENTRL Inc	Infosum	SAS Global Data Mgmt
CipherCloud	Integris	Senya Limited
Clarip	Janusnet	Senzing
Collibra	The Media Trust	Solidatus
CompLions-GRC BV	Mentis	Structure Systems
D.Day Labs	Miner Eye	SuitePrivacy
DataGravity	NextLabs, Inc.	Systnaps
Dataguise	Nortal	T Closeness
Direct Line To Compliance, Inc.	OneTrust	Tealium
DLP Assured	OptInsight	Transcend
DSS Consulting Ltd.	Opus	Varonis
Egnyte	Predesto	Veritas
EPI-USE Labs	Prifender	WireWheel.io
Exonar	Proofpoint	ZLTech
Global IDs	Protenus	
GTB Technologies	Proteus-Cyber Ltd	

DATA MAPPING TOOL VENDORS:		
Advanced Metadata	Exonar	Protenus Proteus-Cyber Ltd
Alation	Global IDs	Qixium
Aptible	GTB Technologies	Raptor Compliance
Automated Intelligence	Hellometrics	Rever
AvePoint	Immuta	RISMA Systems
BigID CENTRL Inc	Informatica	SecuPi
CipherCloud	Information First	Senya Limited
Clarip	Infosum	Senzing
Clearswi	Integris	Signatu
Compliance Technology Solutions bv	MEGA International	SkyHigh
Compliancelog	MetaCompliance	Smart Privacy
CompLions-GRC BV	Mighty Trust Limited	Solidatus
CUBE	Miner Eye	Spearline Risk & Compliance
D.Day Labs	NextLabs, Inc.	Stratrai Ltd
Data Solver	Nortal	Structure Systems
DataGracvity	Nymity	Surecloud Systnaps
Dataguide	OneTrust	T Closeness
Datastream.io	Opus	Tealium
Data Assured	Predesto	TrustArc
dpify bvba	Privacera	Trust-hub
DPOrganizer	Privacy Company	Usoft
Draftit Privacy	Privacy Lab	Varonis
DSS Consulting Ltd.	PrivacyAnt Ltd.	Veritas
ECOMPLY.io	PrivacyPerfect	WireWheel.io
eDatask	Proofpoint	ZLTech
Egnyte		

Data Discovery & Mapping Tool Vendors [18]

An area that requires special attention as it relates to data inventory and mapping efforts are cloud hosted and third-party shared data. One must be cognization of geographical limitations tied to regulatory requirements or contractual commitments, as cloud vendors may be storing your data in any number of countries. Some cloud services will even have your data moving between data centers in different countries periodically, often without the organization's knowledge. Another consideration is that an organization may share data with a SaaS vendor whose information technology infrastructure is outsourced to a tertiary cloud IaaS vendor. In this case, organizations must go the extra step to ensure that they know where their data truly resides. For example, a drug discovery company would be remiss to learn that their research and clinical trial data (i.e., valuable intellectual property [IP]) was residing on a trusted SaaS platform whose hosting was occurring on servers in China – the world leader in IP pirating. Another example would be an operation in the EU, processing EU resident data in the cloud. However, the vendor contract may permit the vendor to host the data at their lowest cost of

[18] Jedidiah Bracy, CIPP (2018). *IAPP 2018 Privacy Vendor Tech Report v.2.4e* (Online). Available: https://iapp.org/media/pdf/resource_center/2018-Privacy-Tech-Vendor-Report.pdf

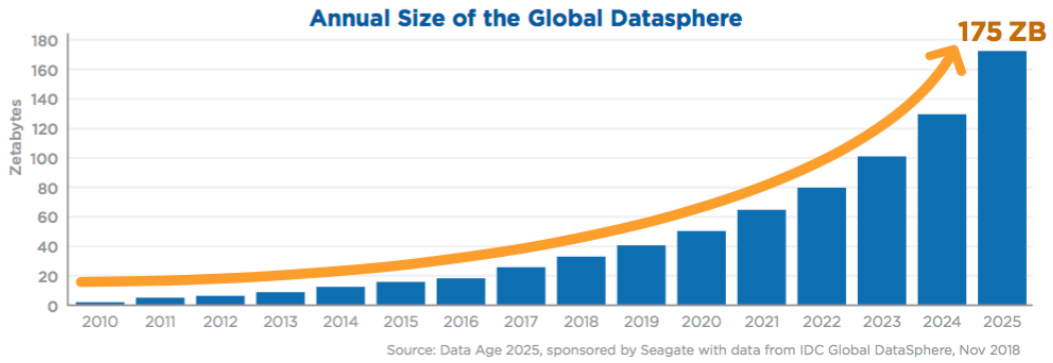
operation, in this case – in the Philippines. Now that organization has a major GDPR issue on their hands. Careful planning, visibility via data mapping efforts, contractual prohibitions, and ongoing monitoring are required to ensure that these scenarios do not impact organizations.

Data inventory and mapping exercises are required by most governing regulatory bodies, either explicitly or indirectly. Additionally, they are instrumental tools that enable data visibility, giving an organization a three-hundred-and-sixty-degree view of their data footprint and how data traverses its network and processing activities. In short, you cannot adequately secure and protect data you didn't know existed or weren't aware was crossing your network. Going beyond security and privacy, data inventories provide visibility of all the information assets an organization possesses, enabling them to maximize data value (i.e., "data is the new oil"[19]), data analytics, and data sharing.

[19] Clive Humby (2006). *The Offices of Clive Humby & Edwina Dunn – About Us* (Online). Available: http://www.humbyanddunn.com

Data Classification and Stewardship

The concept of data classification serves to consistently categorize data based on specific, pre-defined criteria, thereby enabling one to ensure the data is effectively managed and secured.[20] Data classification allows organizations to identify the business value of data (structured and unstructured), segregate valuable data that may be targeted from less valuable data, and make informed decisions about resource allocation to secure data from unauthorized access.[21] The need for data classification is driven by a number of contributing factors, to include regulatory obligations (PCI, GLBA, HIPAA, GDPR, etc.), corporate compliance requirements, and the need to simplify data security operations and strategy. IDC predicts that the global data sphere will grow from 33 zettabytes in 2018 to 175 zettabytes by 2025.[22]



Data Sphere Expected Growth thru 2025 [23]

That is a compound annual growth rate of 61%. With data creation and processing increasing at this rate, organizations cannot afford to a take a "one size fits all" approach to security and must

---

[20] Digital Guardian (2018). *The Definitive Guide to Data Classification – Data Classification for Data Protection Success* (Online). Available: https://digitalguardian.com/resources/whitepaper/definitive-guide-data-classification

[21] Thomas Eck & Anne Grahn (March 2018). *7 Steps of Effective Data Classification* (Online). Available: https://edge.siriuscom.com/security/7-steps-to-effective-data-classification

[22] David Reinsel, John Gant, John Rydning (November 2018). *The Digitization of the World From Edge to Core* (Online). Available: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

[23] Ibid.

be capable to classifying data in a manner that permits the application of limited resources to secure the most sensitive, or at risk, data.

Three common approaches, or methods, to data classification include context-based, content-based, and user-based. The context-based method examines the application, location, and creator along with other variables to derive the sensitivity of data. This approach attempts to answer, "How is data being utilized," "Who is accessing the data," "Where the data is traveling," and "When is the data being accessed."

Content-based data classification operates by inspecting and interpreting files in an attempt to locate sensitive data. This is typically accomplished via fingerprinting and aims to answer the question, "What is in this file/document?" As a point of reference, data loss prevention and email proxy scanners often rely on content-based analysis.

User-based data classification is a manual method that relies on the end-user pro-actively classifying data. While context and content-based data classification approaches can be implemented via automated technical controls, user-based classification relies heavily on a user's awareness and discretion at the creation, modification, or review of files to identify the appropriate sensitivity and assign the corresponding data classification. This approach has the highest risk for misclassification as there can be greater variances in subjectivity and awareness among large populations of end-users.

Data classification efforts also rely on various criteria that aid in determining the sensitivity and value of different types of data within an organization. Some common data classification criteria include:

1) Criticality – a measure of how important data is to resiliency and operational effectiveness
2) Sensitivity – how much harm could be done if this data reached the wrong people

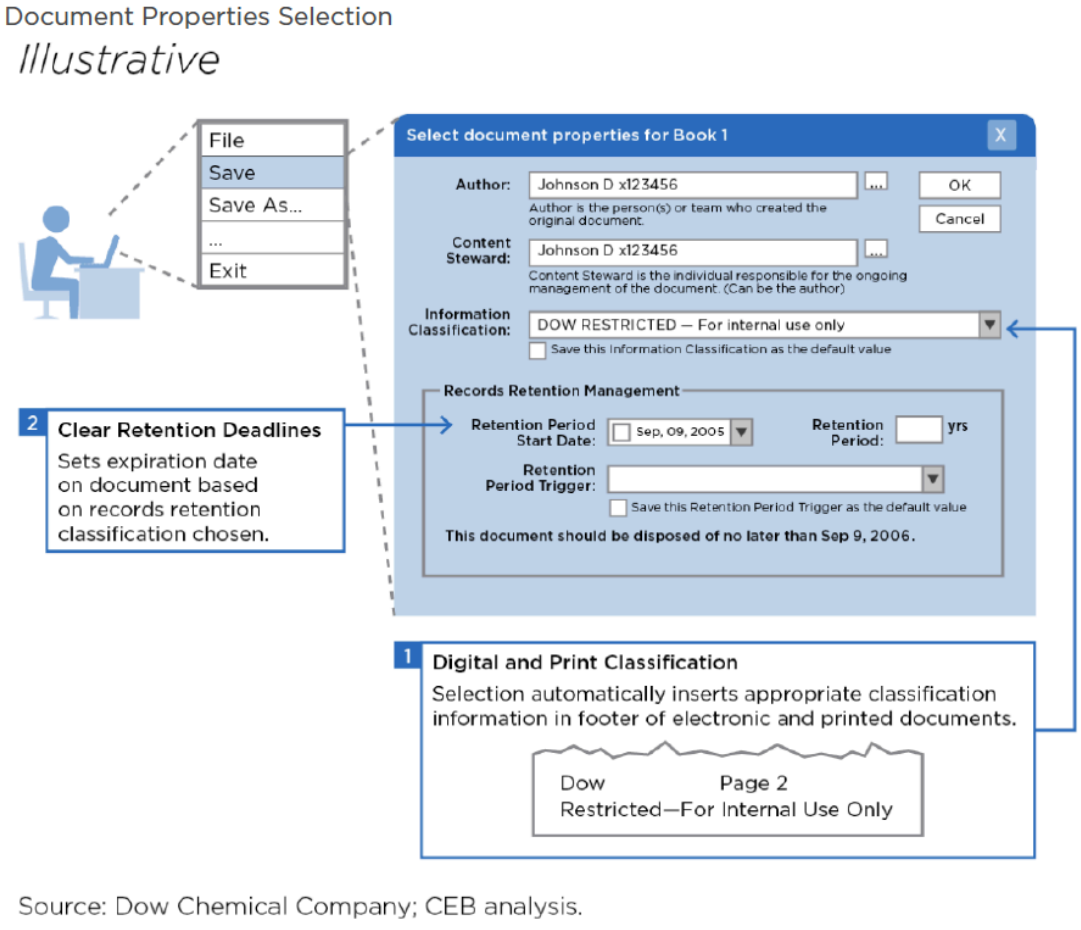
3) Scarcity – how readily available is the data in question (i.e., it may be PII, but is also publicly available)
4) Availability – the timeliness and reliability of access to, and use of, the data in question[24]
5) Retainability - requirements driving records retention and storage of the data
6) Integrity - assurance that data hasn't been modified or inappropriately disclosed while in storage, processing or transmission
7) Regulatory – what legal, industry, regulatory, or client contractual commitments exist around the data an organization collects, stores, processes or transmits.
8) Consent – the documented approval/authorization for data processing and sharing use cases

To deploy a data classification framework, organizations should first determine the scope and objectives of the overarching data classification policy to be documented. This policy should include input from a variety of stakeholders, to include but not limited to, the Chief Information Security Officer, Chief Privacy Officer, Chief Analytics Officer, Chief Data Officer, General Counsel, Chief Compliance Officer, Head of Human Resources, and key Business Unit Leaders (who create and utilize the data in question). This cross-functional team should aim to determine the appropriate data classification method, criteria, and schema based on input from all stakeholders. This effort will drive the creation of a data classification policy which outlines the people, processes, and technologies that govern data classification, as well as, the resultant classification levels/categories and how to assign, document, and monitor classification efforts.

Following the publication of the data classification policy, the cross-functional team must determine how best to operationalize the program across the enterprise. Key steps for this phase include training, awareness, and simple employee data handling practices (i.e., easy reference

[24] CEB Gartner (2018). *Ignition Guide to Data Classification* (Online). Available: https://www.cebglobal.com/member/data-privacy/

guides). Data tagging is another operational consideration one must assess. By embedding data tagging in employee workflows (and systems or applications), one can reduce the end-user burden while increasing compliance with the data classification policy.[25] It is imperative that data classification guidance be simple, clear, and concise, thereby enabling end-users to follow the policy during their normal day-to-day operations without undue effort. An example of an integrated data classification approach (at Dow Chemical) that embeds data classification via data tagging, for the lifecycle of a document, is witnessed below.

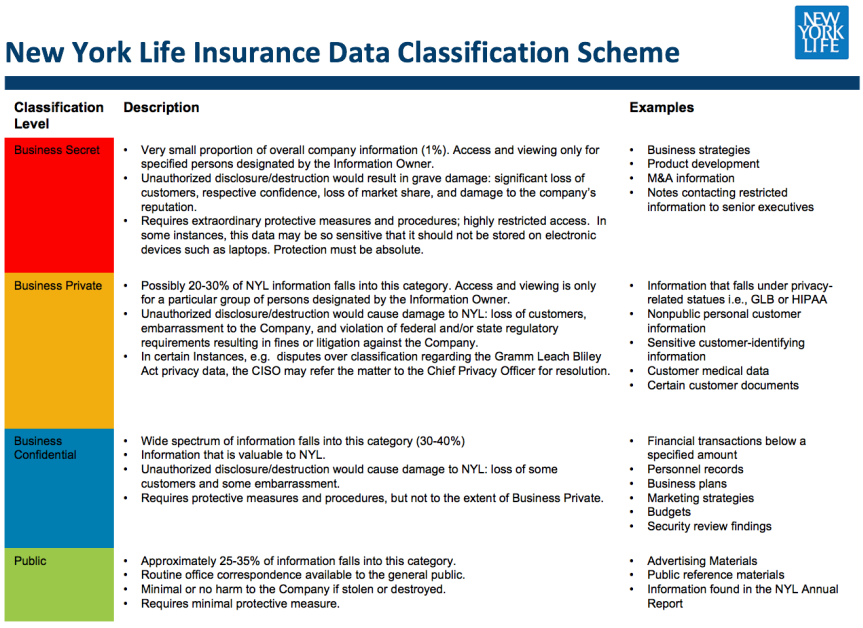


Dow Chemical Integrate Data Classification Example [26]

[25] Gartner Data Privacy Leadership Council (2018). *Ignition Guide to Data Classification: Step-by-Step Guide* (Online). Available: https://www.cebglobal.com/member/data-privacy/

[26] Ibid.

Typical data classification schemas often have a tiered approach. For example, the base tier may include broad classifications, such as public, for partners only, confidential, restricted confidential, etc. A second tier of classifications can serve to further define and align compliance obligations to specific data types. The second tier is often made up of more granular labels such as personally identifiable information (PII), protected health information (PHI), EU/GDPR member data, intellectual property, employee records, attorney-client privilege, pricing data, etc. The number of classifications in the second (and additional) tiers can be as broad or as narrow as required by the organization. Ideally, these classifications should allow information security controls to be applied in an escalated manner commensurate with the value or risk associated with the data. One should not expend the same resources to secure the organization's "crown jewels" as they do to secure data of lesser importance. Below is a high-level example of the top echelons of New York Life's data classification schema.



New York Life Insurance Data Classification Scheme

Classification Level	Description	Examples
Business Secret	• Very small proportion of overall company information (1%). Access and viewing only for specified persons designated by the Information Owner. • Unauthorized disclosure/destruction would result in grave damage: significant loss of customers, respective confidence, loss of market share, and damage to the company's reputation. • Requires extraordinary protective measures and procedures; highly restricted access. In some instances, this data may be so sensitive that it should not be stored on electronic devices such as laptops. Protection must be absolute.	• Business strategies • Product development • M&A information • Notes contacting restricted information to senior executives
Business Private	• Possibly 20-30% of NYL information falls into this category. Access and viewing is only for a particular group of persons designated by the Information Owner. • Unauthorized disclosure/destruction would cause damage to NYL: loss of customers, embarrassment to the Company, and violation of federal and/or state regulatory requirements resulting in fines or litigation against the Company. • In certain Instances, e.g. disputes over classification regarding the Gramm Leach Bliley Act privacy data, the CISO may refer the matter to the Chief Privacy Officer for resolution.	• Information that falls under privacy-related statues i.e., GLB or HIPAA • Nonpublic personal customer information • Sensitive customer-identifying information • Customer medical data • Certain customer documents
Business Confidential	• Wide spectrum of information falls into this category (30-40%) • Information that is valuable to NYL. • Unauthorized disclosure/destruction would cause damage to NYL: loss of some customers and some embarrassment. • Requires protective measures and procedures, but not to the extent of Business Private.	• Financial transactions below a specified amount • Personnel records • Business plans • Marketing strategies • Budgets • Security review findings
Public	• Approximately 25-35% of information falls into this category. • Routine office correspondence available to the general public. • Minimal or no harm to the Company if stolen or destroyed. • Requires minimal protective measure.	• Advertising Materials • Public reference materials • Information found in the NYL Annual Report

NY Life Data Classification Scheme [27]

[27] Gartner Data Privacy Leadership Council (2018). *Data Classification Scheme Creation Tool* (Online). Available: https://www.cebglobal.com/member/data-privacy/

The final phase of data classification is to develop and deploy a monitoring program to measure effectiveness and to identify non-compliance. An end-user feedback channel is also useful to solicit feedback from functional leaders regarding complaints, potential enhancements, and evolving business requirements that may drive the development of new classifications or amendments to existing classifications. A process must also exist to ensure that any modifications to the data classification program are communicated to related processes and systems to ensure that downstream efforts that are driven by data classification remain timely and compliant.

Ultimately, data classification enables the broader data governance program to more efficiently and effectively align and apply controls over the data that matters. Data classification can influence data access rights, information asset classification, data stewardship, data loss prevention efforts, streamline compliance obligations, and aid in the risk-based application of security controls in a resource-constrained environment.

Data stewardship is closely tied to data classification. The premise is that all data that an organization stores, processes, or transmits should have an owner. The Data Management Association International (DAMA), within their Data Management Body of Knowledge (DMBOK), defines data stewardship as, "the formal accountability for business responsibilities that ensure effective control and use of data. A data steward is a business leader or subject matter expert (SME) designated as accountable for these responsibilities."[28] So, once data has been identified, and subsequently classified, the next step would be to identify and assign a data steward. The data steward can have a variety of responsibilities, however, as it relates to data security and privacy, there are a handful of critical ones. Data stewards should be the ones who

[28] Earley, Susan. *DAMA-DMBOK: Data Management Body of Knowledge*. Basket Ridge, NJ: Technics Publications, 2017.

serve as the approver of access to data. They should also be responsible for any sharing of the data set they manage, be it with partners, via APIs and system interfaces, with universities and researchers, or for collective enterprise data analytic efforts. In theory, the data steward would have the best knowledge of the appropriateness of use for the data they manage and can serve as the gate-keeper who ensures that access is provisioned on the basis of minimum necessary (or principle of least privilege) and that the usage of the data is ethical, responsible, and in compliance with regulations, contracts and consent/notice.

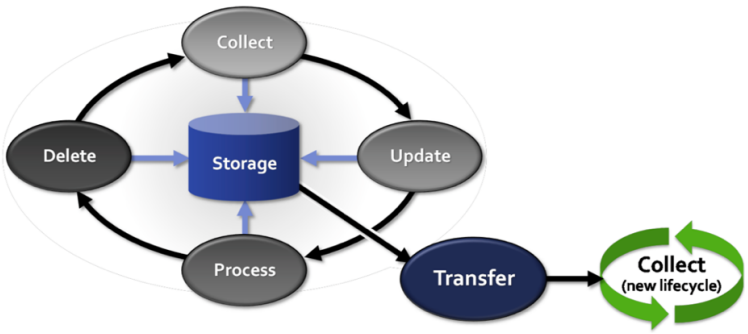
Data classification is the mechanism by which organizations can classify and segment data based on value, risk, and confidentiality, thereby permitting the measured application of finite security and privacy resources. Organizations cannot afford to "boil the ocean" when it comes to the application of privacy and security controls. Given limited resources, security and privacy controls must be deployed in a tiered, risk-based approach. This is not feasible if one is unable to ascertain, or differentiate between, the various risks or opportunities that data presents. Data classification permits security and privacy professionals to effectively and efficiently manage an organization's data in a compliant, value-add, differentiated manner that can enable trust, innovation, growth, and compliance. The following section will demonstrate how security and privacy professionals can leverage data inventories and data classification efforts in the application of privacy and security controls.

Application of Security Controls

The continued growth of data breaches and security incidents within the public and private sectors are evidence that data security practices require enhancement. To give some measure to this observation, below are some pertinent facts:

- Data loss accounts for 43% of data breach costs[29]
- In 2017, there were over 130 large-scale, targeted breaches in the U.S. per year, and that number is forecasted to grow by 27% per year.[30]
- The White House estimated that malicious cyber activity cost the U.S. economy roughly $109 billion in 2016[31]

Data inventory and mapping exercises, along with the resultant data classification efforts, permit information security professionals to plan and deploy their limited security resources in a manner that best addresses the risks their data poses. Security professionals must understand the information (i.e., data) lifecycle in order to develop comprehensive controls that span the data's persistence within the organization. Below is a common depiction of the general information lifecycle. Security planners must



Information Life Cycle [32]

[29] Ponemon Institute LLC. (2017). *Cost of Cyber Crime Study – Insights on the Security Investments That Make a Difference* (Online). Available: https://www.accenture.com/t20171006T095146Z__w__/us-en/_acnmedia/PDF-62/Accenture-2017CostCybercrime-US-FINAL.pdf#zoom=50

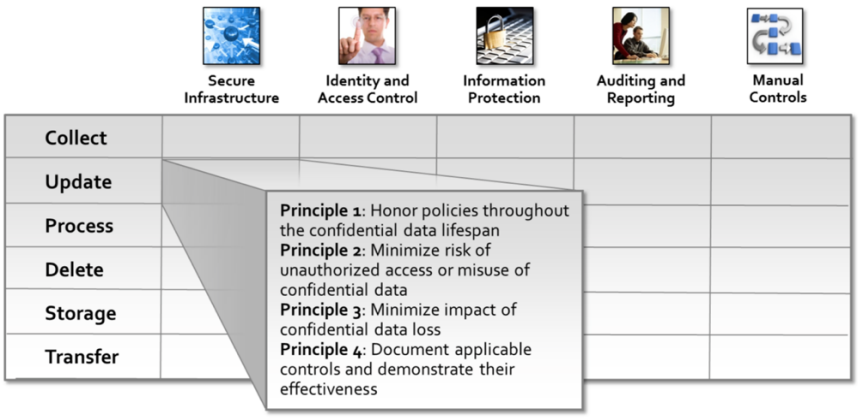
[30] Ibid.

[31] The Council of Economic Advisers (February 2018). *The Cost of Malicious Cyber Activity to the U.S. Economy* (Online). Available: https://www.whitehouse.gov/wp-content/uploads/2018/03/The-Cost-of-Malicious-Cyber-Activity-to-the-U.S.-Economy.pdf

[32] Javier Salido & Patrick Voon (March 2010). *A Guide to Data Governance for Privacy, Confidentiality, and Compliance – Part 3: Managing Technology Risk* (Online). Available: https://iapp.org/media/pdf/knowledge_center/Guide_to_Data_Governance_Part3_Managing_Technological_Risk_whitepaper.pdf

aim to ensure that sensitive data is secured and protected throughout the lifecycle and everywhere the data is stored, processed, or transmitted. Additionally, "traditional IT security approaches that focus on protecting the organization's IT infrastructure by securing the network edge and end points need to be augmented with protective measures that focus specifically on protecting the data that are stored and moved through that infrastructure."[33]

Security professionals are best served by performing a security risk assessment that encompasses both the IT infrastructure as well as the data stored or traversing it. This will highlight to security leadership the high-risk and high-volume operations where limited resources can be deployed to affect the greatest impact and, ideally, the best ROI. The security risk assessment will also highlight control gaps as well as areas where existing controls deliver a residual risk that lives outside the boundaries of the organization's acceptable risk tolerance. An example of a data risk gap analysis matrix that is a common deliverable of a security risk assessment exercise is outlined in the graphic below (this example is specific to data privacy and confidentiality concerns).



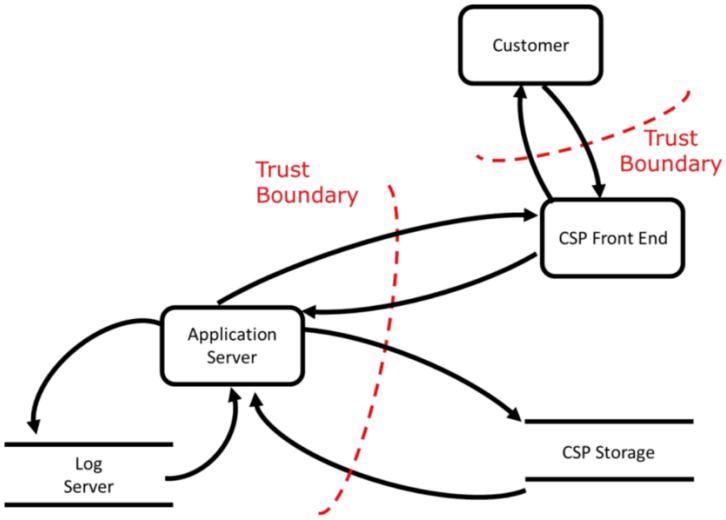
Security & Privacy Risk Gap Analysis Matrix [34]

[33] Javier Salido (2010). *ISACA Journal – Data Governance for Privacy, Confidentiality & Compliance – a Holistic Approach* (Online). Available: https://www.isaca.org/Journal/archives/2010/Volume-6/Documents/jpdf1006-data-governance-for.pdf

[34] Javier Salido & Patrick Voon (March 2010). *A Guide to Data Governance for Privacy, Confidentiality, and Compliance – Part 3: Managing Technology Risk* (Online). Available: https://iapp.org/media/pdf/knowledge_center/Guide_to_Data_Governance_Part3_Managing_Technological_Risk_whitepaper.pdf

A matrix such as this is beneficial as it can combine the information lifecycle, various technology or security domains, and data security and privacy principles within the context of targeted data processing flows to help assess the people, processes, and technology risks impacting the confidentiality, availability, and integrity of the function or asset.

Traditional threat modeling techniques can also be applied to data security and risk reduction efforts. Below is an example of how data mapping outputs can aid in identifying trust boundaries within threat modeling activities. Rectangles represent computer systems (i.e., IT assets), parallel bars denote storage platforms or media, and arrows indicate data flows.



Trust Boundaries Impact Application of Controls [35]

What follows are some targeted security considerations that derive significant benefits from data inventory, mapping and classification processes. A comprehensive discussion on information security controls could fill several volumes, therefore, this analysis will be limited to key controls that inherit the greatest benefits from data governance activities. It is relevant to add that each organization has varying threat and vulnerability landscapes as well as different

[35] Javier Salido & Doug Cavit (August 2010). *A Guide to Data Governance for Privacy, Confidentiality, and Compliance – Part 5: Moving to Cloud Computing* (Online). Available: https://iapp.org/media/pdf/knowledge_center/Guide_to_Data_Governance_Part5_Moving_to_Cloud_Computing_whitepaper.pdf

regulatory or business drivers that must be taken into consideration when designing and deploying security controls. Some of the most common regulatory standards that benefit from strong data governance practices include, but are not limited to, GDPR, HIPAA, FISMA, PCI-DSS, FedRAMP, GLBA, AICPA SSAE 16 SOC 1 & 2 reporting, NY DFS 23 NYCRR 500, HiTRUST, or CCPA to name a few.

De-identification of data is a security control that can offer significant value from an ROI perspective. Organizations can leverage encryption, tokenization, or data masking to accomplish this goal. Which solutions are utilized may be driven by relevant governing regulatory obligations. For example, under HIPAA, encrypted protected health information (PHI) is rendered null via encryption and the loss of encrypted data is not subject to data breach reporting obligations. However, under the PCI-DSS, encrypted data is still considered cardholder data and the application of encryption does not relieve the organization of any required security control obligations. Instead, the PCI-DSS relies on tokenization as a form of de-identification that will effectively remove cardholder data from the scope of PCI compliance obligations (assuming it is deployed in accordance with the standard). Data masking, on the other hand, is useful in de-identifying data in lower environments such as development, test, or quality assurance. The implementation of masked data in lower environments enables broader access and lowers the residual risk associated with large data sets living in potentially less-secure lower environments.

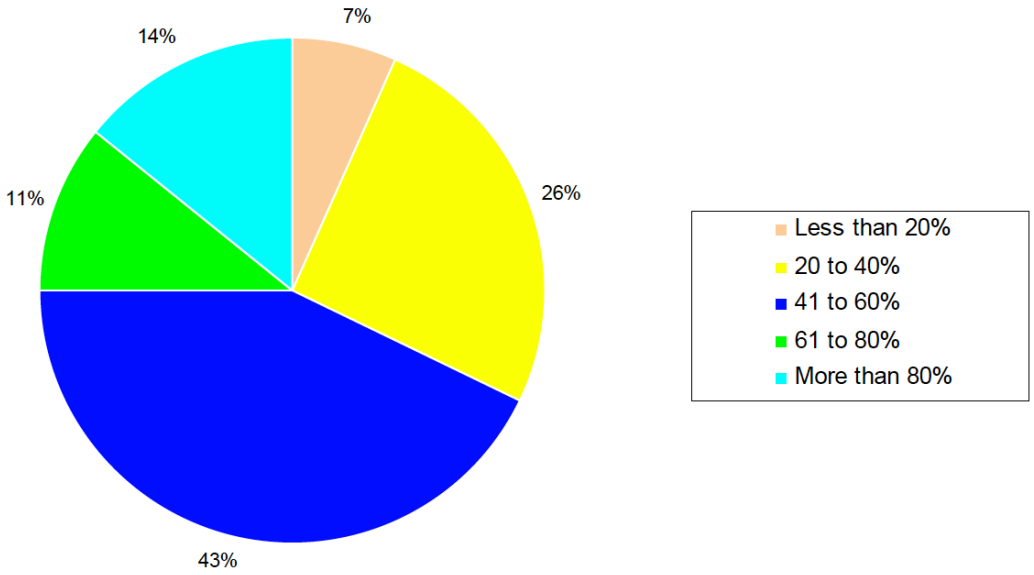
Data governance permits networking and security resources to more effectively architect network segmentation. Network segmentation permits organizations to cluster data by classification, thereby enabling the targeted application of security and privacy controls commensurate with the data's classification. Additionally, network segmentation can enable restrictions with regard to the paths that sensitive data use to traverse the organization's network.

This permits network engineers visibility into whether or not certain network paths require additional hardening based on the classification of data traveling across them. Finally, network segmentation can help support identify and access management (IAM) efforts and enhance role-based access controls.

IAM is the over-arching security process by which organizations manage user provisioning and role-based access. Strong data governance enables more precise provisioning and enables more refined access levels. Knowing where sensitive data resides and its data classification permits IAM to better deploy access under the principle of least privilege (POLP), also referred to as 'minimum necessary'. The proper combination of IAM solutions and principles along with a solid data governance program enable organizations to manage user access via the POLP to ensure access is limited to authorized users while also improving security, reducing complexities and making processes more efficient.[36] While IAM excels with structured data residing in applications and databases, it lacks similar effectiveness with unstructured data. Security and provisioning teams should aim to keep data within structured systems as much as possible. Failure to do so can lead to high-risk or high-value data landing in less secure ad-hoc locations that were never intended to store data of this nature, thereby increasing the risk of inappropriate access and sharing, or data breaches. A recent Ponemon Institute survey on the governance of unstructured data highlighted that unstructured data can often make up almost fifty percent of data within organizations, thereby highlighting the morass of less rigidly managed data stores. This is evidenced in the following chart from the survey.

[36] Robert F. Smallwood (September 11, 2015). *Information Governance: Concepts, Strategies and Best Practices – Chapter 11 Information Governance and Privacy and Security Functions* (Online). Available: https://onlinelibrary.wiley.com/doi/10.1002/9781118433829.ch11

How Much Organizational Data is Unstructured?



Survey on the Governance of Unstructured Data – Ponemon Institute LLC [37]

Data loss prevention (DLP) technologies are useful in augmenting data governance programs as they can both locate and identify data, by classification. DLP adds the most value when deployed across the enterprise to encompass data on servers, data on end-points (PCs/laptop), and data in motion (to include traversing firewalls). DLP can serve as a "trust but verify" monitoring or auditing mechanism and can also be deployed as a proactive prohibitive control (such as with email) that can prevent unauthorized or unsanctioned data leakage events. While DLP has many benefits, those who engage it must be cognizant of its limitations as well. For example, DLP relies on content inspection which is rendered null if network traffic or data-at-rest is encrypted. DLP can also create significant compute and network burdens in large, complex environments. Finally, if a DLP implementation is not well-tuned, it can produce high

[37] Dr. Larry Ponemon (May 22, 2008). *Survey on the Governance of Unstructured Data* (Online). Available: https://www.ponemon.org/local/upload/file/Governance%20Unstructured%20Data%20052208.pdf

volumes of false-alarms that will have a significant drain of limited security personnel who are tasked with investigating these events.

Information rights management (IRM) is an information security technology used to protect documents containing sensitive information from unauthorized access, copying, viewing, printing, forwarding, deleting, and editing.[38] IRM offers a major benefit in that protections persist even when files are shared with internal or external third parties. Even when files leave the corporate network, IRM rules are capable of continuing to protect the file. On-premise IRM solutions can be cumbersome and overly constraining if not implemented properly. Often, IRM is utilized to manage only the most sensitive classifications of data due to its heavy rigor and restrictions. A pharmaceutical manufacturer may find IRM useful in securing drug discovery and research and development data that can be valued in excess of $2 billion dollars and encompass 10 years of investment. A public company's treasurer or finance team responsible for SEC reporting would also benefit from IRM with regard to the management of sensitive financial reporting files. IRM is also emerging in cloud environments. For example, Microsoft's Office 365 application suite has IRM capabilities and are powered by Microsoft Azure Rights Management, Microsoft's IRM solution for the cloud.

Data governance practices also add value to some of the less prominent, but data-rich functions such as records retention and data deletion or disposal. Without solid data governance, an organization's ability to effectively manage record retention obligations (for both paper and electronic records) would be difficult. Data governance also helps facilitate efficient e-discovery for subpoena, warrant, and litigation investigation and fulfillment purposes. Data disposal is

[38] Ajmal Kohgadai (2018). *What is Information Rights Management (IRM)?* (Online). Available: https://www.skyhighnetworks.com/cloud-security-blog/what-is-information-rights-management-irm/

aided by data governance by ensuring data is accurately identified for deletion or disposal at the earliest available opportunity, thereby reducing the organization's data footprint and resultant threat attack surface.

When planning, designing, and deploying security and privacy controls in line with data governance objectives, it is important not to forget some more obscure data sources, such as: digital copiers and printers with hard drives; application, database and Syslog files; VOIP and IVR systems; recordings on call monitoring and quality assurance solutions; biometric data in physical security systems; video on CCTV DVRs; offsite or disaster recovery hot-site backups; and IoT devices data, to name a few. Data exits in many formats and is in a constant state of motion in today's operational environments. It is imperative that proper data governance is brought to bear in order to apply the required security and privacy controls to bring data risks in line with organizational risk tolerance levels.

Conclusion

COBIT 5 succinctly articulates the value of data governance in the following context, "Data governance reflects the practice of evaluating requirements and bringing direction and control over data and information so that users have access to that data and can trust and rely on it. Data governance also involves monitoring the performance of IT operations, especially those areas that relate to data and its availability, integrity, and confidentiality. The complexity of functions increases as enterprises grow and become more sophisticated in their operations and geographical dispersion, and as information security and other requirements evolve."[39] Or to put into simpler terms, data governance takes effort, but brings with it great benefits. With the prolific increase in data creation and resultant attack surfaces, the dynamic growth of threats and vulnerabilities, the sophistication of criminal and nation-state hacking elements, the increasingly complex regulatory landscape, and the obligation to honor trusted relationships with customers and partners, it is clear that without data governance, privacy and security teams will be deploying controls in a scatter-shot manner. Given limited resources, this is not a path that organizations can afford to follow.

Organizations that can timely and effectively answer the three primary questions below will be in the best position to structure their security and privacy programs for success and maximum ROI on control investments.

1) Do you know where all your data is?
2) Is it classified appropriately given its value, sensitivity and regulatory obligations?
3) Are security and privacy controls implemented in accordance with its classification?

---

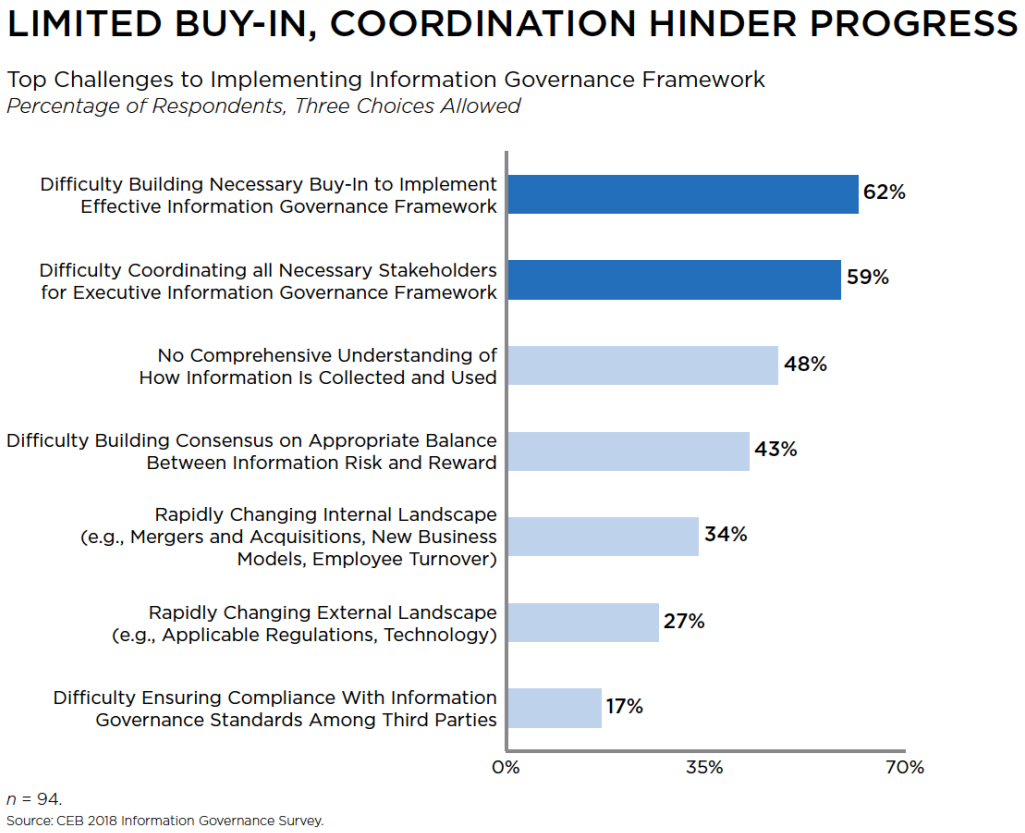
[39] Peter C. Tessin, ISACA (2017). *Governance – Getting Started With Data Governance Using COBIT 5* (Online). Available: http://www.isaca.org/Knowledge-Center/Research/Documents/Getting-Started-with-Data-Governance-Using-COBIT-5_res_eng_1017.pdf?regnum=481802

No two data governance programs will (or should) be the same and it is important that such a program aligns with the organization's mission, vision, and values, as well as their risk tolerance and legal obligations. The analysis above aims to support a holistic approach via a flexible, hybrid data governance framework tailored to an organization's needs. Trying to hammer a square peg (i.e., a rigidly structured data governance program) into a round hole (i.e., a complex organization requiring flexibility) is clearly sub-optimal. Organizations should start small, apply lessons learned, and then scale-up the deployment across the enterprise. All this should be performed with executive support and key stakeholders engaged.

In a world where data breaches have become the norm, it is the responsibility of public, private, and government organizations to take their data stewardship responsibility more seriously and invest in reasonable, yet effective, security and privacy programs to better protect data. Data governance is an instrumental tool that can help organizations do just that.

Appendix – A

Additional Findings from CEB Gartner's Benchmarking Survey (2018) [40]

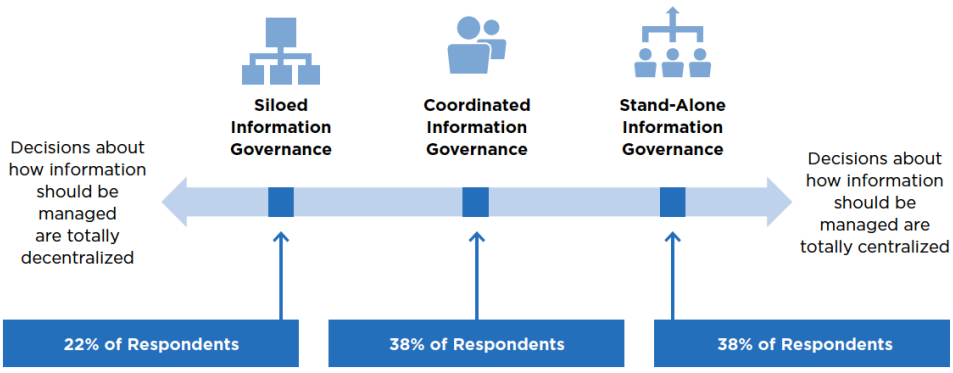


- Legal, compliance and privacy executives cite difficulty building buy-in and coordinating stakeholders as the primary obstacles to effective information governance.
- Buy-in and effective cross-functional coordination are crucial to effective information governance, given the number of stakeholders who must be motivated to devote time and resources and coordinated in their standards and activities.
- Legal, compliance and privacy executives report satisfaction with the collaboration among their three groups, with a mean score of 5.25 out of seven.
- The challenge is extending that collaboration to the broader stakeholder network, with whom close collaboration is more rare.

[40] CEB Gartner Data Privacy Executive Council (2018). Information Governance Benchmarking - 11 Key Findings (Online). Available: https://www.cebglobal.com/member/data-privacy/

Continued.

Structures of Information Governance Frameworks



22% of Respondents	38% of Respondents	38% of Respondents
Hallmarks of Siloed Information Governance: ▪ Functions plan and execute information governance in silo, coordinating with other functions on a case-by-case basis. ▪ Tends to be focused on more tactical objectives (e.g., creating policies).	Hallmarks of Coordinated Information Governance: ▪ Coordination most commonly takes place through a cross-functional committee comprising a diverse group of stakeholders from the legal suite to IT/IS, Internal Audit and ERM. ▪ Common mandates for the committee include: setting guidelines, advising on business decisions and managing information governance projects.	Hallmarks of Stand-Alone Information Governance: ▪ Can be a function, typically helmed by the chief data officer, chief information governance officer or the equivalent. ▪ Can also be a dedicated team within a function ▪ Function or team takes the lead role in establishing and enforcing standards, as well as coordinating the implementation by relevant functions.

Source: CEB analysis.

Strengths and Limitations, by Type of Information Governance Structures

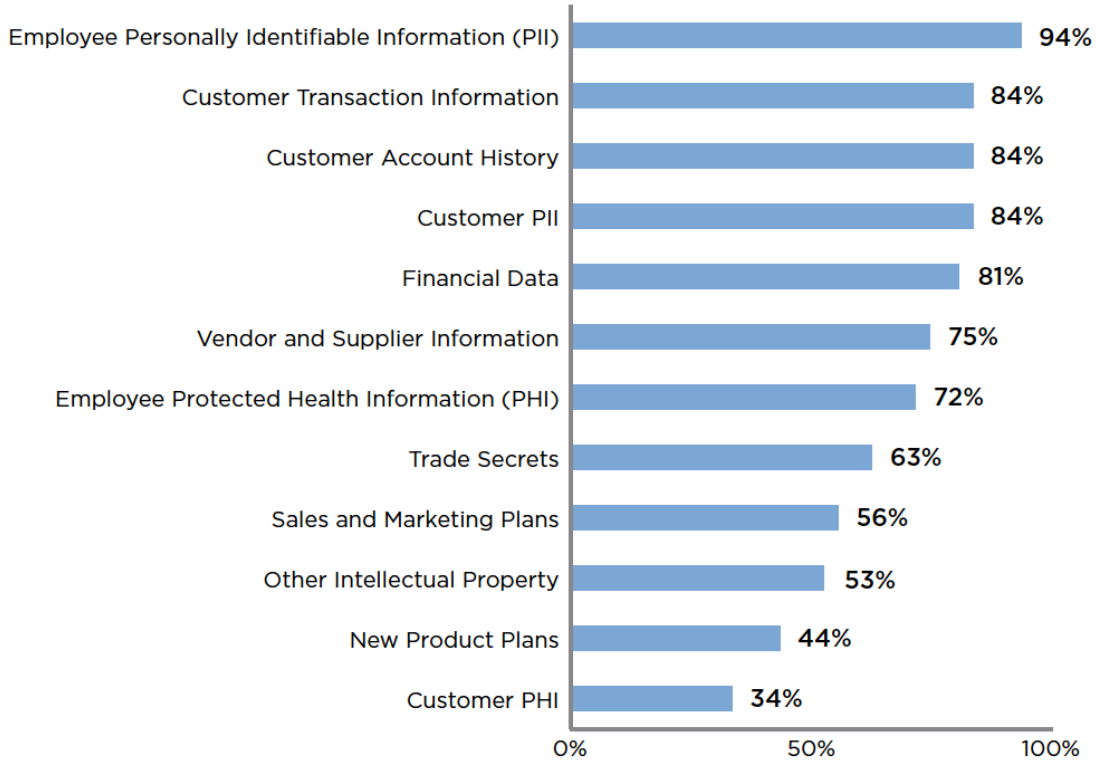
Siloed Information Governance	Coordinated Information Governance	Stand-Alone Information Governance
✔ **Pros:** ▪ Independence in decision making and execution allows functions to maximize to their objectives ▪ Less bureaucracy enables quicker, and potentially more efficient, execution of tasks	✔ **Pros:** ▪ Provides a platform for stakeholders to collaborate ▪ Uncovers opportunities to reduce duplicative work, close gaps and align strategy and messages ▪ Creates a forum for the business to seek feedback	✔ **Pros:** ▪ Clearly establishes information governance as a key organizational priority, putting teeth behind the initiative ▪ Clear ownership and accountability supports action ▪ Dedicated funding for information governance activities
✘ **Cons:** ▪ No clear owner or accountability for information governance outcomes can stall progress ▪ Limited functional coordination can result in duplicative work, gaps in coverage and/or conflicting messages	✘ **Cons:** ▪ Shared accountability often contributes to delays in action, as no one is empowered to drive progress or take action ▪ Well suited to high-level decisions and discussion, but less effective at operational coordination ▪ Proliferation of committee members and of committees themselves creates complexity	✘ **Cons:** ▪ A central function can struggle to gain visibility into how data is collected, used and retained across the organization ▪ The division of roles and responsibilities among the information governance function and the functions responsible for information management may become muddled

Source: CEB analysis.

Continued.

THE SCOPE OF GOVERNANCE IS TOO BROAD

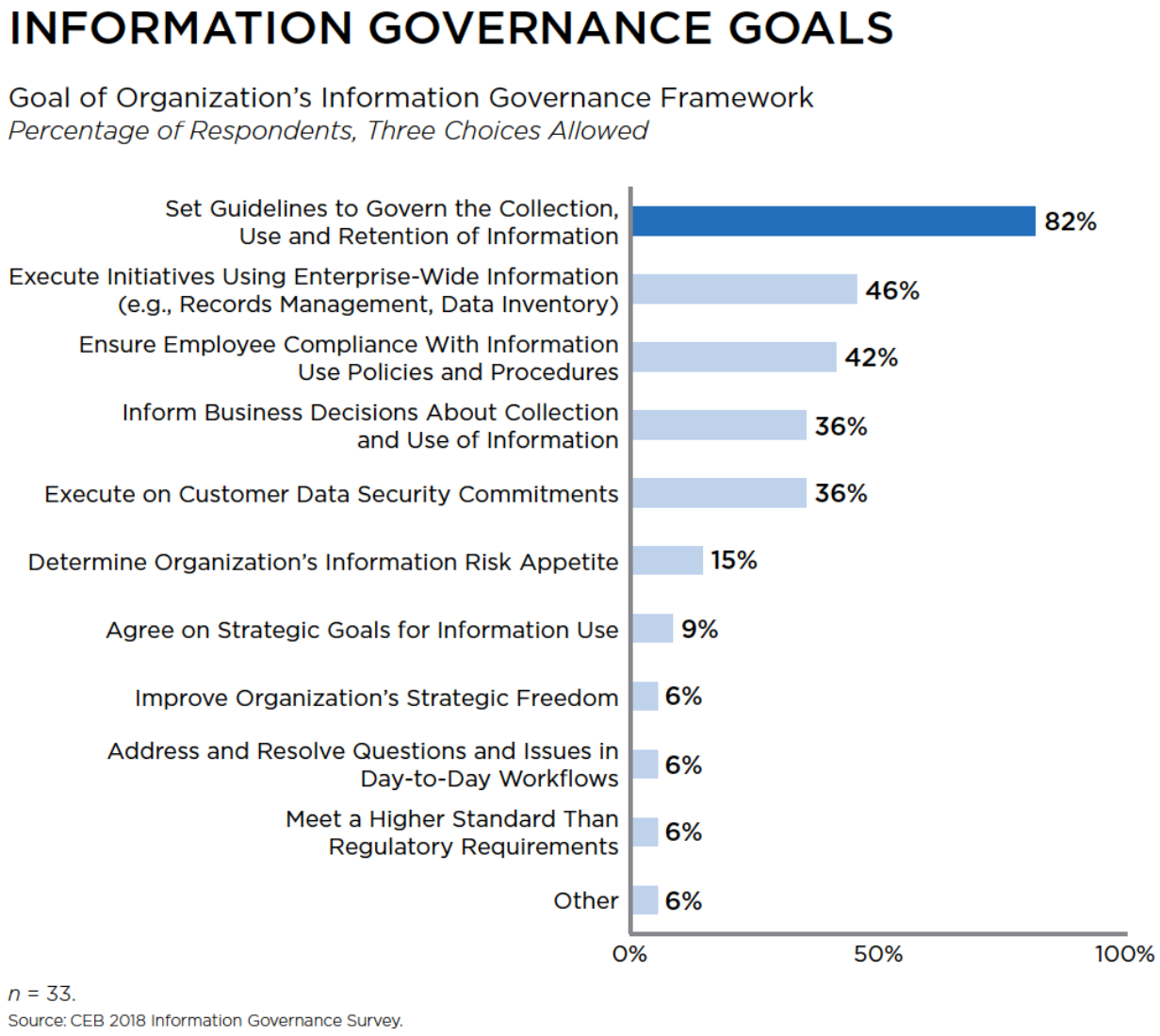
Data Types In-Scope for Information Governance
Percentage of Respondents



n = 33.
Source: CEB 2018 Information Governance Survey.

- Most organizations take too broad a scope for their information governance framework, creating several risks for its success.
- A scope that is too broad can contribute to stalled progress in two ways:
    - The initiative can become too resource intensive.
    - Stakeholders may perceive the initiative as too remote from their needs and responsibilities.
- An all-encompassing approach fails to recognize that not all data holds the same risk or the same value, creating the likelihood that some things are over or under governed.
- Organizations can narrow the scope of information governance by considering value (e.g., contribution to strategic initiatives) and risk (e.g., level of regulatory intensity).

Continued.



- The main goal of the information governance framework is to arrive at a set of guidelines to govern data collection, use and retention.

Continued.

CONTRIBUTORS TO INFORMATION GOVERNANCE

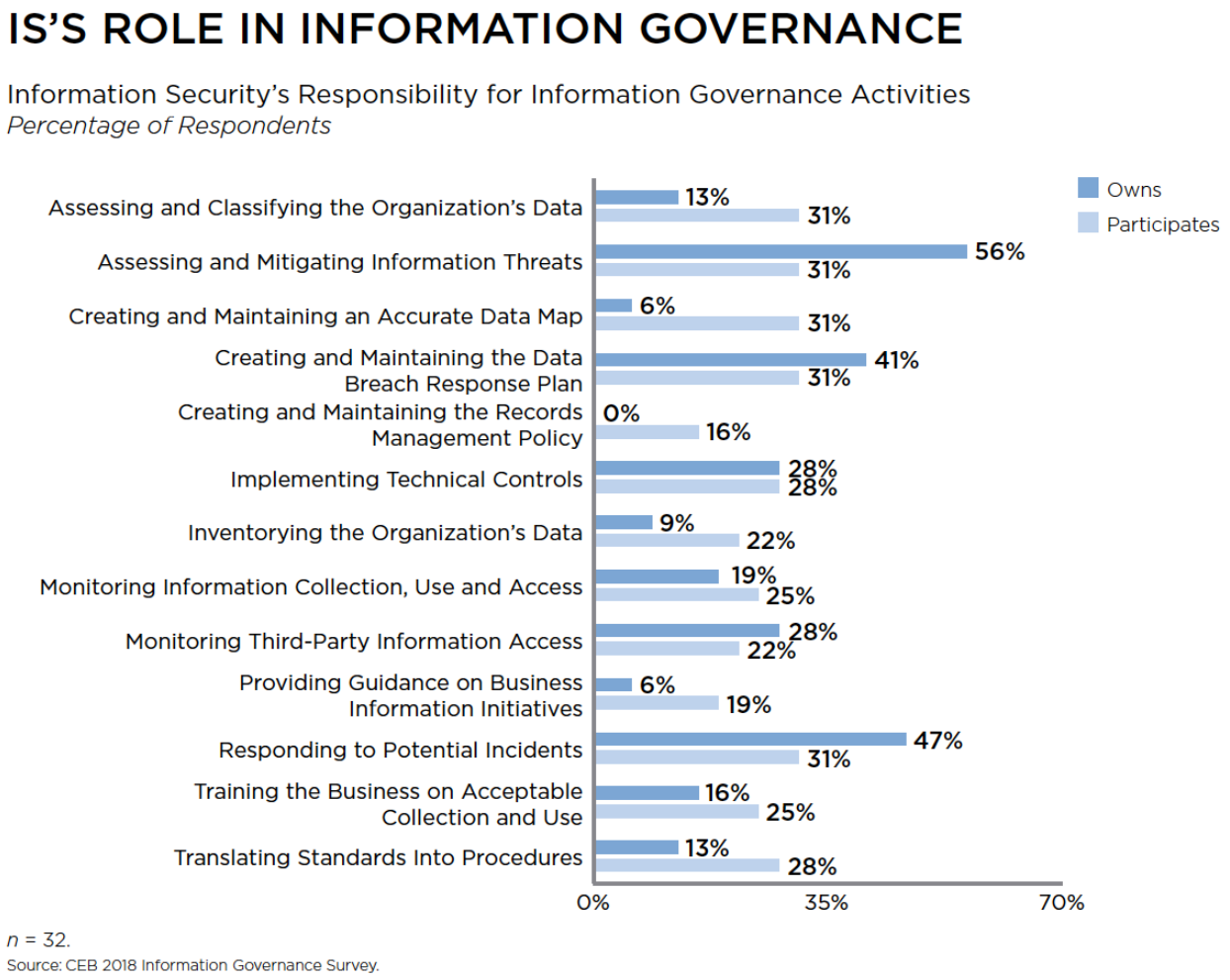
Participants in Designing and Implementing Information Governance Framework
Percentage of Respondents, Many Responses Allowed

Legal	**94%**
IT	**79%**
Information Security	**76%**
Compliance	**76%**
Privacy	**64%**
Human Resources	**46%**
Enterprise Risk Management	**36%**
Business Unit Managers	**36%**
Finance	**33%**
Records	**33%**
Internal Audit	**33%**
Marketing	**15%**
Research & Development	**15%**
Quality	**15%**
Other	**6%**

n = 33.
Source: CEB 2018 Information Governance Survey.

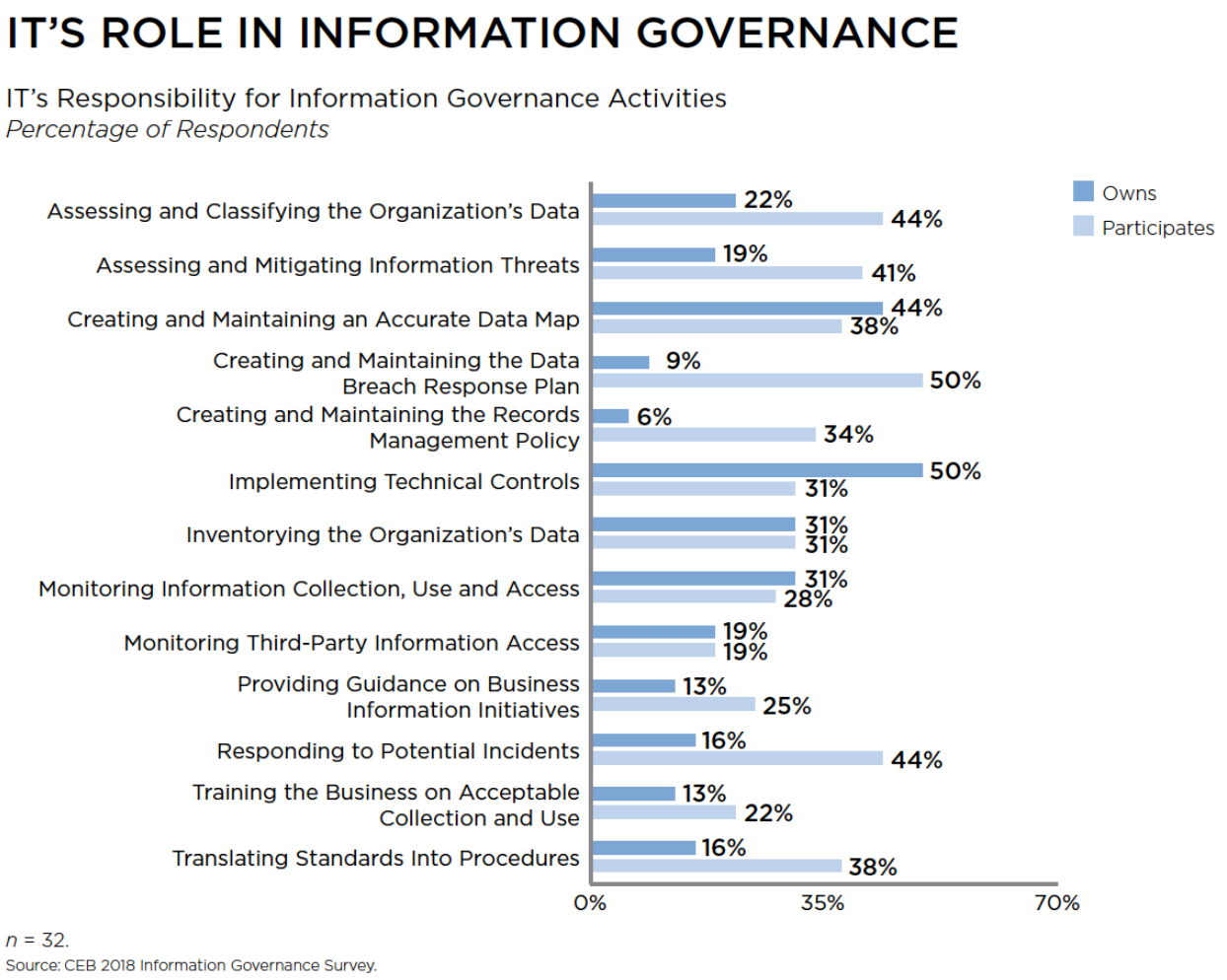
- Information governance is truly a cross-functional effort, with five different functions participating more than half of the time.

Continued.



- Information security has the highest across-the-board rates of owning information governance activities.

Continued.

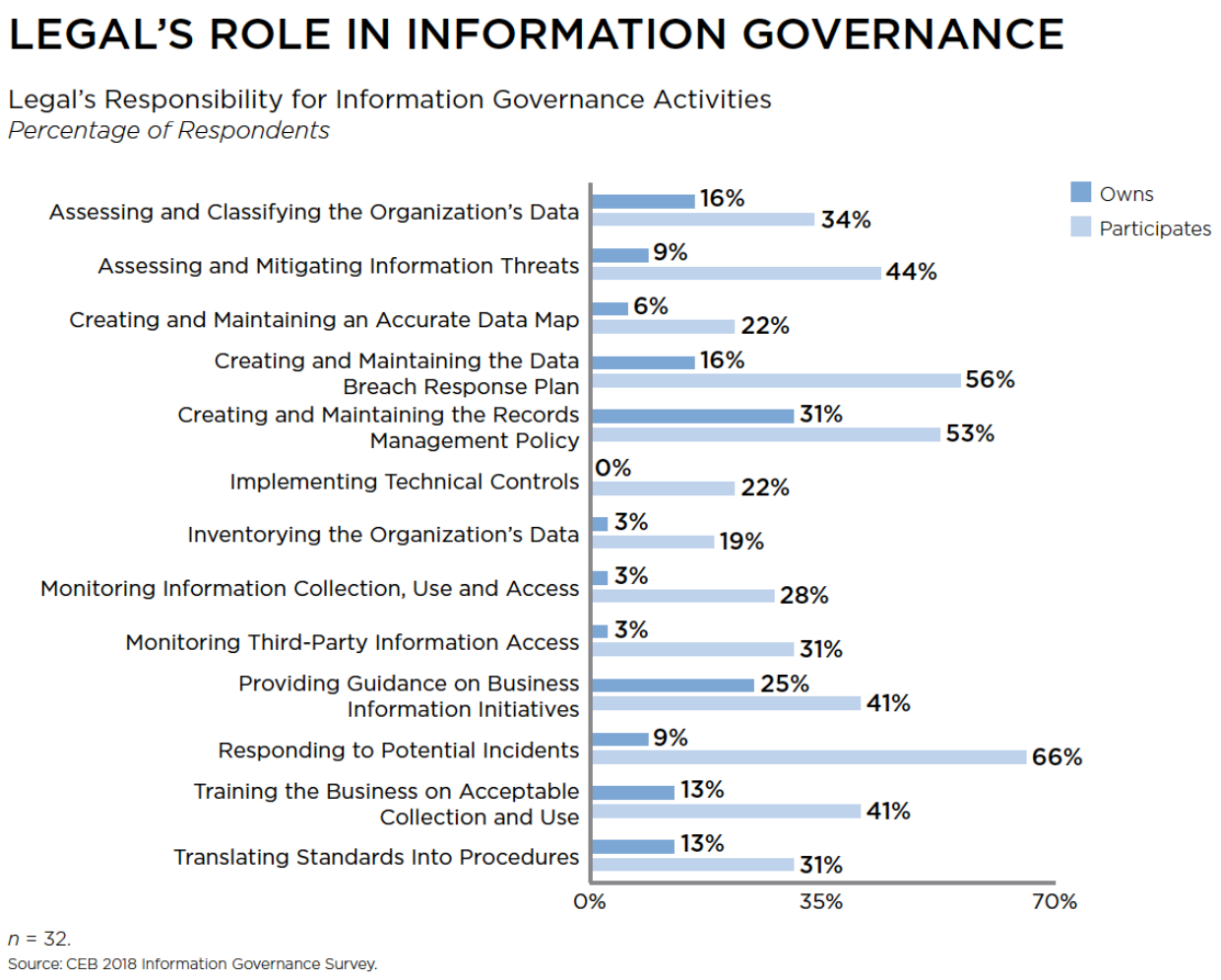


n = 32.

Source: CEB 2018 Information Governance Survey.

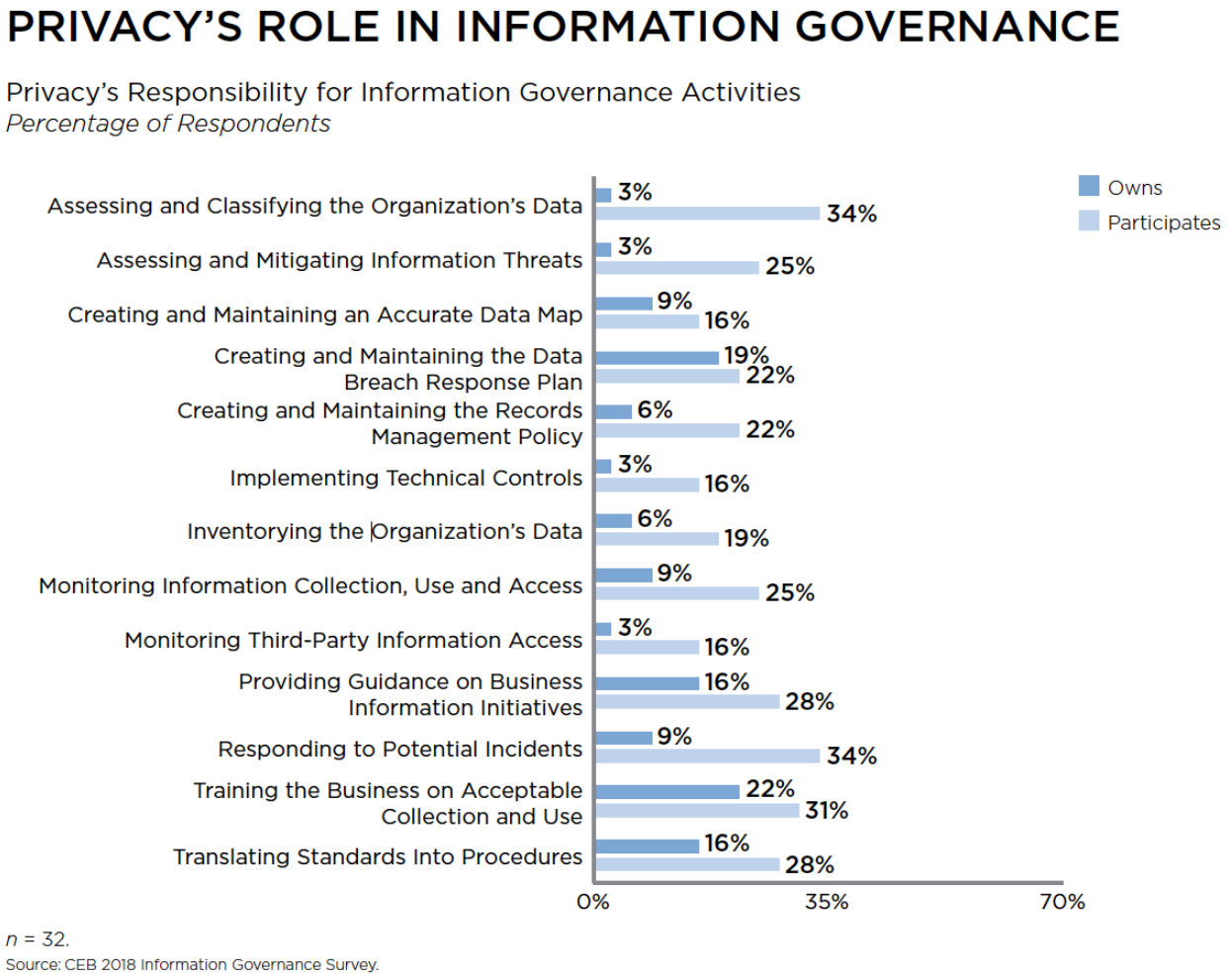
- IT also owns a high percentage of information governance activities and participates frequently in several others.

Continued.



- Legal has a surprisingly low rate of ownership over information governance activities, although most have high levels of participation in those activities.

Continued.



- Privacy also exclusively owns few information governance activities. It also has relatively low participation in many of the activities.