Brought to you by:

HITACHI
Inspire the Next

Intelligent Data Governance

for dummies®
A Wiley Brand

Understand the data governance imperative

Establish a data governance framework

Protect and secure crucial data

2nd Hitachi Vantara Special Edition

The Hitachi Vantara Team

About Hitachi Vantara

Hitachi Vantara, a wholly owned subsidiary of Hitachi, Ltd., helps data-driven leaders find and use the value in their data to innovate intelligently and reach outcomes that matter for business and society. Hitachi Vantara combines technology, intellectual property, and industry knowledge to deliver data-managing solutions that help enterprises improve their customers' experiences, develop new revenue streams, and lower business costs. Hitachi Vantara elevates your innovation advantage by combining IT, operational technology (OT), and domain expertise. Hitachi Vantara works with organizations everywhere to drive data to meaningful outcomes. Visit **HitachiVantara.com**.

Intelligent Data Governance

2nd Hitachi Vantara Special Edition

by Lawrence C. Miller

for dummies®
A Wiley Brand

Intelligent Data Governance For Dummies®, 2nd Hitachi Vantara Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-119-88638-9 (pbk); ISBN: 978-1-119-88639-6 (ebk). Some blank pages in the print version may not be included in the ePDF version.

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Project Manager: Carrie Burchfield-Leighton
Sr. Managing Editor: Rev Mengle
Managing Editor: Camille Graves
Acquisitions Editor: Ashley Coffey
Sr. Client Account Manager: Matt Cox

Table of Contents

Foreword

Today our lives are intertwined with data like never before. From personal communications to the most obscure interests born of an internet search, our digital footprint continues to grow. Beyond personal interests, data about us holds our most sensitive information, such as health records or finances. We, the digital citizens, trust in an organization's stewardship of our data. Mismanagement of our personal data breaks our trust, damaging our willingness to do future business.

Data permeates corporations as much as it does our personal lives. It can be said the companies today build their "digital transformation" on their data. Their control, access, and use of data enables them to reach insights on new markets and achieve greater efficiencies. The quality and veracity of data within a corporation is essential to maximize digital potential.

For digital citizens to maintain trust in organizations they choose to engage with, proper governance of data is a must. For corporations to fully leverage their digital assets, proper governance of data is a must. With emerging regulations applying to the personal information of the digital citizen, data governance is a fiscal imperative.

Simply storing this much data is a daunting task. Keeping it organized in a manner that allows the data to be found and used requires intelligence. Information that cannot be found is like a tool in an over-full junk drawer — what good is it? Intelligence allows us to know what is being kept and where. It allows us to know the "retention" appropriate to that information, such as who has access to it and how long it can be kept. Intelligence allows proper reporting on all such activities.

Data growth will continue for the foreseeable future — as will the sensitive nature of the information data contains. Thus, the need to intelligently govern data will only increase. This book is designed to provide you the necessary insights to move beyond simply storing data and into creating a framework under which data can be intelligently governed for the length of time it has value to your business or regulations require it to be kept (whichever is greater). Keep in mind that governance is not meant to be heavy-handed control backed by complex and stoic processes.

Rather, it is an effort to create a fluid and flexible information fabric that is designed to increase the value and relevance of the data you rely on.

We hope you find this book insightful — allowing you and your organization to accelerate your ability to enact intelligent data governance.

The Hitachi Vantara Team

Introduction

Data is your most valuable asset. As organizations move up the digital maturity curve, they increasingly focus on big data and analytics projects that can help create a single user view and deliver actionable insights to relevant stakeholders. Intelligent data governance is key to maximizing the value of data in the age of big data, artificial intelligence (AI), and machine learning (ML). Before you can put your data to work for your business, you need to enrich it with metadata that gives it meaningful context for better management and governance, ensures compliance with constantly evolving regulations, minimizes the risk of data breaches, and reduces the cost and complexity of discovery and reporting.

In this book, we explain how intelligent data governance solutions that extend across your private and public cloud data footprint can help drive more business value from your data and simplify compliance for your organization.

About This Book

Intelligent Data Governance For Dummies, 2nd Hitachi Vantara Special Edition, consists of six chapters that explore

- What data governance is and why it's important (Chapter 1)
- How to establish a data governance framework for your organization (Chapter 2)
- Data governance and management concepts, principles, and components (Chapter 3)
- Data protection and security requirements (Chapter 4)
- The business benefits of a value-based data governance program (Chapter 5)
- Data governance best practices (Chapter 6)

Foolish Assumptions

It's been said that most assumptions have outlived their uselessness, but we assume a few things nonetheless.

Mainly, we assume that you are someone responsible for data management and governance in your organization. As modern organizations evolve to create a culture in which data-driven decisions are made, everyone within the organization who has access to or uses data to make decisions is now responsible for data management and governance.

If any of these assumptions describe you, then this book is for you. If none of these assumptions describe you, keep reading anyway. It's a great book and when you finish reading it, you'll be pretty intelligent about all things data governance.

Icons Used in This Book

Throughout this book, we occasionally use special icons to call attention to important information. Here's what to expect:

REMEMBER

This icon points out information you should commit to your nonvolatile memory, your gray matter, or your noggin — along with anniversaries and birthdays.

TECHNICAL STUFF

You won't find a map of the human genome here, but if you seek to attain the seventh level of NERD-vana, perk up! This icon explains the jargon beneath the jargon.

TIP

Tips are appreciated, never expected — and we sure hope you appreciate these tips. This icon points out useful nuggets of information.

WARNING

These alerts point out the stuff your mother warned you about (well, probably not), but they do offer practical advice to help you avoid potentially costly or frustrating mistakes.

Beyond the Book

There's only so much we can cover in 48 short pages, so if you find yourself at the end of this book, thinking, "Where can I learn more?" just go to `hitachivantara.com/data-governance`.

IN THIS CHAPTER

- Looking at the data governance imperative
- Defining data governance
- Creating an intelligent data governance strategy

Chapter 1
Recognizing the Need for Intelligent Data Governance

In this chapter, you explore the market drivers and key challenges of data governance, what data governance is all about, and how to create an intelligent data governance strategy for your organization.

Understanding Market Drivers and Key Challenges

Digital transformation is disrupting businesses everywhere and data has become an organization's most strategic tool for understanding, predicting, reaching, interacting with, and retaining customers and influencing their behavior. Strategic and operational decision making is further driving data science, artificial intelligence (AI), and machine learning (ML) applications of data in the modern enterprise.

With the exponential growth of data, a strict regulatory environment, cyberthreats (including ransomware attacks) on the rise, and highly competitive and innovative upstarts continuously upending the market, understanding your data and protecting and extracting value from it is a business imperative — and a formidable challenge. Businesses must trust the data they use, understand end-to-end lineage, and get on top of data quality issues and the cost of bad data.

If businesses ever needed an incentive to govern data tightly, it's compliance. The cost of failing to comply with stringent regulatory requirements can add up quickly including legal costs such as litigation and settlement fees, as well as reputational damage and business losses.

Consider, for example, the European Union (EU) General Data Protection Regulation (GDPR), which requires companies that handle the data of EU citizens to comply with strict data privacy regulations. Organizations that fail to comply with the GDPR can face penalties of up to four percent of their annual global turnover. With potential penalties this severe, it's no wonder that many organizations are investing heavily in compliance initiatives. Some even view compliance as a key competitive advantage.

Yet, many factors make it increasingly difficult for organizations to maintain compliance with the growing number of regulations across multiple industry verticals. Case in point: Global banks work under more than 100 different regulations at any given time. These include Basel III and the Basel Committee on Banking Supervision (BCBS) 239, Securities and Exchange Commission (SEC) regulations, Dodd-Frank, EU GDPR, and the Markets in Financial Instruments Directive (MiFID), among others. Chief information officers (CIOs) and information technology (IT) directors in the EU are readying for yet another new banking regulation: The Revised Payment Service Directive (PSD2), which enables consumers and businesses to use third-party providers, such as Google or Facebook, to manage finances, pay bills, or analyze spending. The impact on EU banks is very real — they will be required to enable third-party provider access to customers' accounts through application programming interfaces (APIs). As a result, IT costs will likely increase due to new security requirements like stronger identity checks and API development.

Complicating matters further is an increasing number of producers and consumers of data and the growing mountains of

data across disparate systems. An IDC study, *Data Age 2025: The Evolution of Data to Life-Critical*, predicts that by 2025 global data creation will swell to a total of 175 zettabytes. By 2025, 60 percent of the total data will be created by the enterprise and 90 percent of this data will require to be secured. The big question for enterprises is what and how much data to store, manage, and protect.

Legislation and new regulations are also outpacing the capabilities of existing IT infrastructure investments and the budgets necessary to adopt adequate solutions. For example, the length of time that sensitive data must be stored to meet regulations may surpass a legacy architecture's physical capabilities. In response, many IT leaders over-implement data control processes, not only stifling innovation and productivity but also hampering the built-in flexibility needed to adapt to changes in the regulatory landscape.

WARNING

Exposure to risk will continually grow as the gap widens between applications, people, processes, systems, and existing technologies with each new or modified regulation.

Ransomware attacks have now become a global issue. According to a new study from ESG Global, *The Long Road Ahead to Ransomware Preparedness,* 79 percent of organizations have been exposed to ransomware, 41 percent been victim of a successful attack, and 32 percent have been hit sporadically every month. Attacks are on the rise because it's easier than ever for criminals to launch an attack and because organizations are paying the ransoms to recover their valuable data. According to Cybersecurity Ventures, ransomware attacks rose by nearly 93 percent in 2021 compared to 2020 levels, with 2,690 reported attacks in 2021 and 1,389 reported attacks in 2020. Per Cybersecurity Ventures, a business, consumer, or device is expected to be attacked by ransomware every 2 seconds by 2031, up from 11 seconds in 2021. Furthermore, Cybersecurity Ventures predicts that total ransomware damage will cost $265 billion per year globally by 2031.

Not all threats to data security are perpetrated by malicious outsiders. IT leaders must also maintain the integrity of data and protect data from being corrupted or irretrievably deleted by accident or destroyed by some unforeseen outage or event.

For IT departments, frequently running data backups and installing the latest antivirus software can help minimize potential damage from a ransomware attack, accidental deletion, system corruption, or outage, but it isn't enough.

Information security and identity management solutions can help prevent unauthorized access, but data security must extend from an organization's core to every endpoint. For example, a financial institution may be compliant with regulations, but as soon as the organization allows trade activity via a customer's personal device, it introduces a new endpoint and increases exposure to attacks.

To address these data governance challenges, businesses must capture, manage, and utilize the growing volume of data to reveal competitive insights, inform new product development, and attain a deeper understanding of their customers. Yet many business and technology leaders are ill-equipped to answer even the most basic questions about their data. Questions, such as

- Who owns what data?
- Do we understand it?
- Is it protected?
- How long should it be retained?
- Where is it located?
- Who has access to it?
- What does it contain?
- What does its quality look like?
- How can the data be used to produce a competitive advantage?

What Is Data Governance?

Data governance helps organizations better manage the availability, usability, integrity, and security of their enterprise data. With the right technology, data governance can also drive enormous business value and support digital transformation.

Data governance is about bringing data under control and keeping it secure. Successful data governance requires understanding of the data, policies for quality and metadata management, as well as knowing where data is located, how it originated, who has access to it, and what it means. Effective data governance is a prerequisite to maintaining business compliance, regardless of

whether that compliance is self-imposed or required by industry or government mandates.

Regulatory compliance generally adds to operations complexity, requiring the ability to properly search data, know every word or number it contains, and produce the right data point if requested for any purpose. However, regulatory compliance is just one of the key aspects of the much broader issue of enterprise data quality and data governance. The quality, veracity, and availability of data to authorized personnel can also determine whether an organization meets, or violates, stringent regulatory requirements.


TECHNICAL STUFF

Data governance moves beyond information management to support business processes and encompasses a broad set of data strategies and functions, including the following:

- Data delivery and access
- Data integrity
- Data lineage
- Data loss prevention (DLP)
- Data security
- Data synchronization
- Master data management (MDM)

Planning for Intelligent Data Governance Success

The pressures facing modern organizations have never been higher. Employees now expect access to enterprise data anywhere, at any time, on any device; business leaders demand data that is searchable, viable, and flexible enough to deliver actionable insights; and tight regulations require best-in-class compliance processes. This trend has accelerated exponentially since 2020 when organizations in many industries were forced to adopt more remote worker models.

Satisfying these expectations can give rise to considerable security and compliance risks. IT leaders must find a balance between driving business value and complying with stringent regulations,

all without disrupting workforce productivity or compromising business assurance.

A strong data governance strategy doesn't require you to build a new silo dedicated exclusively to compliance. Rather, it combines governance requirements with data analytics for a more dynamic data governance process. Traditional data management techniques make it nearly impossible to easily access and analyze data for actionable insights. Yet, organizations need a way to integrate and visualize data quickly to meet compliance requirements, while also driving better business decisions.

Using software and application programming interfaces (APIs) to move data to a central data hub and then managing access, protection, retention, and expiry of each object, organizations can create a win-win situation that delivers greater value from data and a superior customer experience, while driving a broader strategic and analytic plan at the executive level.

WARNING

APIs have become an increasingly common attack vector with threat actors using credential stuffing, cross-site scripting (XSS), injection, man-in-the-middle, and other attack techniques to compromise applications and data. Organizations must secure their APIs by using a combination of API gateways, API access management, and web application firewall (WAF) solutions, as well as implementing security best practices such as Transport Layer Security (TLS) encryption, least privilege, and input validation.

By gathering, analyzing, and gleaning insights from organizational data, business leaders can readily detect and respond to regulatory inquiries, perform early case assessments, and explore new business opportunities based on data with the highest referential value and quality. With the data under control and centralized, it's possible to defensibly delete data when its value to the business is no longer measurable, with a standardized and repeatable approach.

REMEMBER

Data governance is the process of managing the availability, usability, integrity, and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. An intelligent data governance strategy and technology solution empowers an organization to manage its data and meet regulatory requirements and can also support the organization's journey to digital transformation.

IN THIS CHAPTER

- Defining a new data governance framework
- Automating policies, standards, and processes
- Understanding unique data management needs

Chapter 2
Orchestrating a Single Source of Truth

In this chapter, you learn how to establish a data governance framework that helps you classify and manage data as a business asset, implement and enforce policies, standards, and processes, and manage unique data requirements.

Managing Data As a Business Asset

Forrester Research outlined a data governance framework, *data governance 2.0*, defined as an agile approach to data governance that focuses on controls for managing risk. This control enables broader and more insightful use of data, which is required by the evolving needs of expanding business ecosystems. Four types of data governance are identified:

- **Systems of record:** These traditional types of structured databases support online transaction processing (OLTP) and online analytical processing (OLAP) where governance focuses on quality, master data management, and compliance.

- **Systems of engagement:** This data governance is driven by new systems of customer, employee, and partner engagement, such as social media, chatbots, mobile apps, and so on. Here, governance deals with the need for data personalization versus the need for data privacy.
- **Systems of automation:** Governance is driven by the Internet of Things (IoT) integration, operational technology (OT) data, event correlation, analytics, and artificial intelligence (AI). Here, data governance must not only focus on data context but also the source and quality of the data that may be created and processed on the edge of your business or outside of your normal realm of responsibility.
- **Systems of design:** These systems are for creativity and innovation where the stakeholders may be product designers or researchers. Applying standards of data governance to creativity may not be as well-defined as in the previous systems. Data governance in this case may entail social, legal, ethical, and moral questions.

Setting Data Policies, Standards, and Processes

Intelligent data governance begins with clearly defined data policies, standards, and processes. For example, organizations typically have policies, standards, and processes for information security and privacy, data retention, records management, and more, but defining these and other important policies, standards, and processes is only the starting point for an effective data governance program.

Many organizations create policies as paper records or as electronic PDF documents. For example, a policy for retaining invoices may define the required retention period, what the procedure is for retaining invoices, and so on. However, manually implementing and enforcing paper-based policies, standards, and processes is impractical in most organizations. With an advanced software-based policy engine, organizations can automate the implementation and enforcement of these policies, standards, and processes in live information systems where the actual data or information governance rule exists.

Exploring Unique Data Management Needs

The need to rapidly and accurately store and manage unstructured data has never been greater. Enterprise IT organizations are trying to solve difficult business data issues, just as the digital universe hits vast growth spikes. Massive and perpetual data storage increases are predicted. Yet IT budgets worldwide have had a flat to marginal growth trajectory. Balancing these realities and supporting business governance mandates requires innovative IT thinking.

Enterprise IT organizations are looking for ways to effectively manage and secure sensitive records and big data far into the future. These organizations explore ways to cost-effectively archive and manage the flood of unstructured data types in order to derive valuable business information and insights.

IT leaders see object storage as the ideal technology solution for data governance. *Object storage* is a data storage architecture that manages data as variable-sized containers (known as *objects*) organized into a flat address space, rather than as more complex hierarchical *files* or *blocks*. Objects can contain both user-provided and system-generated metadata to enrich the data.


TIP

Object storage brings structure to unstructured data, such as audio, video, images, and documents. The structure provided is in the form of seemingly limitless and customizable metadata that can be used to refine how the file is described, what it contains, and what value it brings to the business, without first having to open the file. This metadata space on the object is where access, retention, preservation, mobility, and other policies are applied and enforced.

The complexity of governing unstructured data stems from both its variety and its difference from information found in a traditional database. Object storage makes it easier to store, protect, secure, manage, organize, synchronize, share, search, and analyze all data, including unstructured data. Although use cases vary, object storage is particularly critical for organizations in highly litigious and increasingly regulated environments. Regulatory compliance requires such organizations to store and organize large volumes of data, and these volumes simply overwhelm the capabilities of

many legacy storage systems and traditional transaction-based technology architectures. Fortunately, object storage architectures were designed with this fact in mind, becoming the data reservoir for relevant organizational data that must be managed and controlled for specific periods of time without suffering losses or corruptions.

Data quality and trust

To create a single customer view and deliver actionable insights, organizations must ensure that the data used for analytics is accurate. However, data quality is often a significant challenge for many organizations. The questionable quality of data is often complicated by the logical and physical silos that separate users from the data they need to support their business functions. Users must often access multiple separate data systems on a regular basis to perform their role within an organization.

Simply put, many organizations lack a central repository — a "single source of truth" — where information can be reconciled and aggregated for ongoing business operations. This leaves critical decisions in disarray where the consensus is mainly about questioning the quality and veracity of the data, rather than making the best tactical or strategic decisions based on the data. The failure to centralize, cleanse, augment, manage, and govern data results in disjointed interpretations of key business data and prevents the best use of metadata as a means of data categorization and classification.

REMEMBER

The failure to centralize, cleanse, augment, manage, and govern data makes it less transparent, harder to locate, difficult to control, and almost impossible to integrate. A key means to address this failure is to centralize knowledge around the data in a machine learning-driven data catalog. A data catalog can enable delivery of trusted data and knowledge that your organization's data is fit for purpose and ready to act on. You can quickly assess your critical quality metrics, define your data quality in the language of your business, and automate with holistically applied data quality policies.

Traditional enterprise data centers often end up as repositories for silos and disparate data sources, requiring time-consuming, hands-on management approaches by highly specialized IT staff. Data infrastructure has shifted from traditional on-premises

physical servers to virtual networks, and data centers can now communicate across multiple sites, both on-premises and in the cloud. The actual data exists and is connected across multiple data centers, the edge, and public and private clouds. Modern data centers are technologically advanced with fully autonomous machines that use AI and machine learning to support agility and innovation. The value of infrastructure standardization, data intelligence, and data centralization is key to delivering data of the highest quality and most referential value to the business.

However, the centralization of data shouldn't require physically centralizing it, and it shouldn't matter where data resides in an organizational structure. Instead, the modern data center facilitates a centralized form of control via a virtual data hub. By creating that centralized data hub, IT leaders can standardize data access, management, and governance and deliver greater convenience to users without compromising security or compliance.

Modern data centers with centralized data access don't focus on technology because the technology stack is so flexible. Instead, IT can focus on turning a regulation or requirement into a policy that protects the business and enables users to work with data in new and more insightful ways.

A 2002 study conducted by the Compliance, Governance and Oversight Council (CGOC) revealed that 69 percent of information being retained by companies is, in effect, data debris — information having no current business or legal value. A 2002 Data Warehousing Institute study titled *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data* found that 76 percent of flaws in organizational data were due to poor data entry by employees. These trends have continued since the original studies were published, further compounded by the proliferation of data and lack of effective data governance. It is much better to move data quality upstream and embed it into the business process, rather than trying to catch flawed data downstream and then attempting to resolve the flaw in all the different applications that are used by other people. Introduce data quality control into the business process instead of trying to catch flawed data downstream and then attempting to resolve the flaw in all the different applications that are used by other people.

Creating a centralized data hub starts with the ingestion of data, which includes the elimination of data debris and the cleansing of

flawed data. *Data debris* is the data that's identified as no business utility, not subject to legal hold, and has come to the end of the life cycle.

Metadata

Metadata can be two types: technical and business. Technical metadata is achieved in file systems, for example, using an object storage model. File systems typically only collect a few minor data points — such as filename; file type; creation, modification, and access dates; and system-based attributes — about the information they store.

Some applications — such as office suites and media creation apps, as well as various devices such as digital cameras and smartphones, and industrial Internet of Things (IoT) devices — generate additional, or custom, metadata to further describe the contents of the file. However, this metadata is typically embedded in the file, so it's hidden and underutilized.


TIP

With the right software tools, this embedded contextual information can be extracted and added to the metadata that's key in establishing the contents or value of the file itself.

Metadata defines the criteria needed to establish long-term tiering, retention, deletion, security, and access policies. Unfortunately, no industry-wide metadata model currently exists that establishes universal, business-related fields that address these policy needs, as well as other practical fields such as ownership, nation of origin, privacy, regulatory controls, legal hold, and other tags that address specific business needs.


TIP

Because of the lack of industry standardization of metadata, organizations should partner with data storage vendors that understand the challenges of business data and that can help define a long-term metadata framework for the organization.

While practically all data storage vendors offer some form of metadata tagging, few offer advanced enrichment, indexing, search, or metadata-driven disposition capabilities. Successfully integrating these technologies becomes the foundation for deriving the business value enabled by rich metadata. But customers need to understand the business metadata in order to know exactly how to govern it better and make sure the right data is available to the right teams at the right time.

A data catalog can use AI and machine learning to create a set of business glossaries above this data to semantically enrich this data and add business meaning to it. Now the governance policies can be applied to obfuscate personally identifiable information (PII), personal health information (PHI), or other company-specific sensitive information.

Reference data

Reference data — such as units of measurement, country codes, calendar dates, and fixed conversion rates — define a set of values, statuses, or classifications that can be used by other data fields, for example in a financial, customer relationship management (CRM), or industrial asset performance management (APM) application. Reference data is commonly used to support key business processes and does not change often, yet many organizations lack specific roles for accountability over the data and definitions that are core to the business.

Without explicit control over reference data, organizations must face the arduous task of combing through a mountain of team projects that organically spring up to create, define, and interpret the semantics and values deemed to be "reference data." Lacking a coordinated and centralized enterprise control mechanism for reference data only gets worse as data sets from different producers, repositories, and versions are aggregated, elevating the differences in what each source deems as reference data. This can make data difficult to integrate or consolidate without a high degree of manual interpretation and intervention.

REMEMBER

When an intelligent form of data governance is applied to the business, reference data can be defined and associated with unique data repositories in a way that separates the reference data from the source that relies upon it. This removes the need for individual ownership and centralizes the responsibility and control to a foundational layer supporting all relevant business functions equally.

Mobile data

According to Statista's "Mobile worker population in the United States 2020 and 2024" study, the population of mobile workers in the U.S. is forecast to increase from 78.5 million in 2020 to 93.5 million by 2024. As their numbers grow, many of these

workers will transition away from slow and cumbersome in-house file-sharing solutions to more nimble cloud-based collaboration tools. However, the more employees send and receive sensitive data via the public cloud, the greater the threat to security.

At the same time, employees need to access files from a variety of devices and share those files quickly and easily with dispersed teams, but not at the cost of compliance.

Making your data mobile means greater exposure to risk. With more access points, more devices, and more data overall, securing mobile data is a multifaceted challenge. The ever-changing regulatory landscape and data sovereignty issues add to the complexity of this challenge.

With the right enterprise-grade tools, organizations can enable data mobility while ensuring proper data governance. A strong, enterprise-ready content mobility solution provides a consistent set of tools for protecting, securing, and governing your content across datacenters, remote and branch offices, public and private clouds, and mobile devices. Such a solution integrates with third-party technologies that focus on the needs of specific industries or lines of business and incorporates the following elements:

- Access control and identity management
- Data loss prevention, protection, and retention
- Encryption
- Information life cycle management
- Legal holds
- Mobile device management and more
- Mobile app security

TIP

Organizations can leverage innovative technology approaches to meet workforce demands for mobility without compromising the data governance requirements of the business.

IN THIS CHAPTER

- Understanding the data services platform components
- Designing for quality, modeling for relevance
- Analyzing, processing, and managing access to data
- Addressing content protection and security
- Enabling compliance and mitigating risk

Chapter 3
Architecting a Data Services Platform for Governance

In this chapter, we explain how to establish a foundation for servicing data to the edge of your organization with a platform that can scale with your business, provides capabilities to extract insights from your data, and is flexible enough to mold to your data governance requirements. This chapter also shows you how to identify and avert the all-too-common IT pitfalls of selecting the tools to use before the goals, strategies, and processes related to data governance are in place.

Exploring Concepts, Principles, and Components of the Platform

Strategic business initiatives often trigger one or more projects that involve IT, such as consolidating business applications, synchronizing disparate information systems to support a

core business process or, in this case, supporting data governance requirements. Yet many IT organizations struggle to deliver on the expectations of the business, especially when it comes to keeping up with the requirements of how data is managed and controlled. This struggle stems from the chaotic nature of data — shifting sensitivity or importance criteria, the changing roles of data producers and consumers, new or changing governance requirements, and the degree to which data travels, just to name a few. All of these issues stem from a single root cause: data fragmentation.

REMEMBER

Because data resides in disparate silos throughout the enterprise, the content, quality, structure, and definitions of the data are as variable as the silos that host them. To ensure that business decisions and operations are based on data that is trustworthy, timely, complete, accurate, and insightful, IT must facilitate timely access, integrations, and data delivery. Furthermore, today's stringent regulatory environment demands the prioritization of data governance to support auditing and reporting requirements over all other business functions, making the challenge for IT still more overwhelming.

The solution to these challenges is a data services platform (see Figure 3-1).

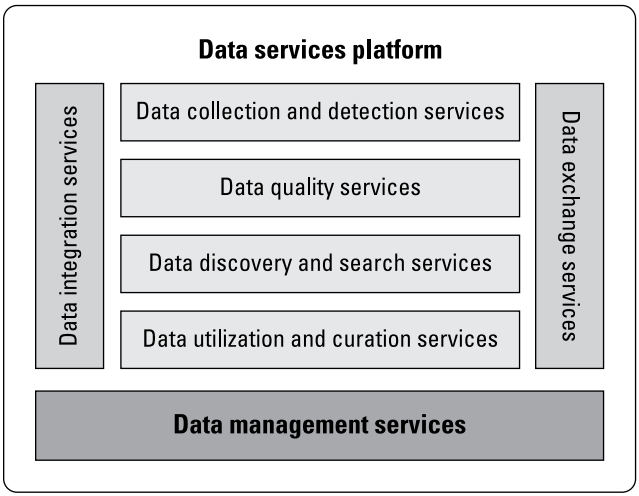

FIGURE 3-1: The data services platform.

A data services platform is a technology architecture that delivers features specific to how the data it is managing is utilized, regardless of who or what created it and needs to use it. To some, this may be a relatively new concept or shift in thinking. But as you consider the speed and complexity behind data growth, to continue to evaluate the competitiveness of your business through the applications that support your vision and strategy is tantamount to transformative extinction.

Modern data services platforms are core technologies that organizations use to pivot their offerings and behaviors from the perspective of the data they deem most valuable and relevant to the business. Product ecosystems, business processes, and data governance become more valuable when they are associated with a platform. This is because data itself becomes the commodity rather than the technology.

TIP

Data services platforms must include, at a minimum, the following kinds of services to be relevant to the business, especially for data governance initiatives:

- **Data integration services:** These services allow the organization to access all of its fragmented data regardless of where it resides, who has access to it, or what application created it. It is a service that starts the process for creating an accurate and consistent view of the core data assets that the business should be relying on by ensuring they are leveraged across the enterprise.
- **Data collection and detection services:** Integrated data streams enable the platform to identify correlated data elements and aggregate them into a centralized index. The act of aggregation affords the platform the added ability to detect characteristics of interest to the business — for example, identifying and notating data sets that contain patterns of information that would be considered sensitive information aligned to an internal or industry regulation.
- **Data quality services:** Data interrogation that focuses on ensuring the data adheres to how the organization has defined quality. This could require specific tagging, data categorization or classification, cleansing the data to normalize specific data types (such as date formatting, currency conversion, distance, measurements, geo-coordinates, and so on), and enriching the data with

additional metadata information to further increase its relevance and value to the business.

- **Data discovery and search services:** A major benefit of data integration is that all organizational data is routed through this service. While some data will remain in its original location (think of this as application-controlled data), other data elements are likely to be aggregated and centralized in this platform. Regardless of where the data lives, flowing through the service means that an index of all of the organization's data can be created. This index is a representation of the source data (the most important and relevant parts, if you will) that can be easily searched to discover the most relevant information based on keywords, natural language, or categorical criteria. The workforce efficiency improvements are often the most impactful, but the most tangible benefit is the reduction in infrastructure investments because of the elimination of duplicate data that results from data fragmentation.
- **Data utilization and curation services:** These services allow the organization to continually assess the usefulness of the data, what the data is and where it is located, what to do with the data once its usefulness expires, who can have access to the data, and who is responsible for the data. This service is tightly coupled with the data application programming interface (API) to ensure that data presentation is accomplished in the context of the user who needs it versus the application that generated it.
- **Data exchange services.** Also known as data APIs, these are the industry standard interfaces that allow applications to interact with the data from a centralized and scalable location. These are also the entry points that enable developers to integrate other systems and services into the information fabric that the platform creates. From a data governance perspective, this means that the processes and rules set forth on the data element become inherited and applied in the application or service that is trying to consume and use it.
- **Data management services:** The foundation of the platform architecture is responsible for storing the data and exposing interservice connection points to affect the data it is responsible for. Through this service, data can be controlled, secured, protected, made readily available, retained, and disposed of. The policies (global, localized, and file-level)

that govern the data the degree to which the data is allowed to be moved from the core to the edge of the business (for example, what data can reside on your smartphone or laptop) can be managed. In data management services, the real magic of intelligent data governance happens: the definition of the policy that governs the data and the application of the policy to the data element. This is what make the service so valuable to data governance initiatives and strategies.

Designing for Quality, Modeling for Relevance

Most businesses recognize the importance of having access to quality data, but many wait until poor quality data takes its toll on the business, or results in a compliance or regulatory issue, before taking action to address it. Thus, many recent concerns about the quality of data likely stem from past indifference, compounded by the vast quantity of data with which businesses must contend.

Similar data quality challenges arise when an organization opts for a quick-fix or ad hoc approach to address a data quality problem. These quick fixes are often implemented to address a single data fragment or silo but fail to comprehensively address the inherent weakness that the poor-quality data creates across the enterprise.

These two points highlight the necessity for organizations to first design their information supply chain around the quality metrics they will measure their data against, before bringing applications and users into the stream.

REMEMBER

For data quality to be most effective, the methodology that defines and measures it must be overseen and implemented by a data governance body, and automated and enforced by a data services platform.

The first step in designing for quality is to profile the data to

- Continuously discover its strengths and weaknesses
- Pinpoint errors and problems within the data's characteristics

- Isolate inconsistencies and remove redundant or irrelevant information
- Communicate with the other services of the platform to remediate these discoveries or apply specific policies relevant to the data, based on its referential value

The next step is to evaluate the overall quality of the data by using the platform to check for completeness, consistency, accuracy, and conformity to the defined quality standards. This is where you rely on data modeling techniques to define a timeline during which the data has relevance before requiring it to be part of a data quality audit or assessment.

These two activities drive the design and implementation of policies that enforce the business rules for each data element. These policies are based on the assessed quality level of the data and categorized by relevance (or risk) to the business, with the platform handling the automated management and control of this criterion.

Analyzing, Processing, and Managing Access to Data

As you might imagine, all of the focus on data management and control, categorization and classification, and data characterization is meant to elevate the data to a state where the systems, services, and people who require access to the data can be strictly enforced based on

- **Confidentiality** (used by authorized personnel)
- **Integrity** (data editing controls)
- **Availability** (access to authorized data when it's needed)

The problem that most organizations face stems from the organic nature of their data landscapes — relying on the systems that own the data fragments to manage access and use. Too often in such a scenario, anyone has access to the data whether they need it or not. Instead, the tasks performed by the employee should determine the characteristics of his or her access. This principle of data access management can only be enforced consistently and effectively with a data services platform.

Inherent in the data management services platform are the core security services that manage access to the information for which the platform is responsible. These access controls permeate associative services as well. These become the gateway that codifies "acceptable" and "authorized" data access and use, based on the structure created by the data governance rules when the element is first collected. These policies specify precisely which users have access to the data, which applications can modify it, and which contextual elements can be exposed to people based on their role in the organization. Imagine being able to dynamically redact a specific data pattern, deemed sensitive to non-HR employees, in real time.

This approach allows you to establish controls to data access based on audited rights of the person or systems. This access is managed by authentication (who is attempting to access, commonly controlled by a security system), authorization (whether or not they are in a group with rights to the data elements), and auditability (how the data is being used).

Addressing Content Protection and Security

Of course, no matter how much governance is applied to the data or how easy it is to integrate with, if authorized users and systems cannot access the data or rely on it, then all of this effort is for naught. The data services platform is the foundation for an intelligent data governance approach.

Within the data services platform, the data management service translates the logic into explicit commands that are executed on a physical or software layer — especially with those underlying and supporting systems where the data is being stored.

Choosing the right storage layer for the data services platform requires one that is scalable, performant, and extensible enough to support the data APIs discussed earlier in this chapter. Ideally, the underlying technology component is based on an object storage foundation. Without going into too many technical details, the object storage layer provides direct alignment to the services mentioned previously, but brings with it system level automated

controls for how the data it holds is protected from corruption and isolated from breaches.

Protection of data at this layer is facilitated in three ways (either individually or in combination):

- **Data replication:** The creation of managed copies of data stored in specified locations, from which recovery operations can be supported.
- **Data versioning:** Saving new copies of files when changes occur, so that previous versions can be retrieved, reviewed, and/or promoted at a later time.
- **Erasure coding:** A form of protection where data is broken into smaller fragments that are expanded and encoded and stored across different locations with a configurable number of redundant pieces.

Enabling Compliance and Mitigating Risk

Without proper aggregated and accurate data for analysis, compliance leaders are left to fly blind when it comes to developing their risk-based testing and monitoring approaches. Consequently, most organizations often recognize that their first priority in realizing better monitoring and testing capabilities is to understand and potentially enhance their technology infrastructure. Therein lies the value of the data services platform.

When the data services platform acts as the central information hub for the enterprise, the requirements of compliance responsibilities can be supported while continuously driving risk exposure often buried within data or processes. For the compliance leader, the central information hub delivers three key benefits due to the separation of the data from the processes that created it or the applications that rely on it:

- **A deeper understanding of organizational data and relationships:** Centralizing the data allows compliance leaders to develop a better understanding of their data without getting lost in the infrastructure. Instead, they can focus on the identification of data sets that require further

remediation based on the value of the data and the potential risks it may cause. In addition, they can better assess whether the quality gaps are the result of inconsistencies in data feeds, inputs, or user interactions. Finally, centralization of data allows the compliance leaders to consistently analyze the root causes of any issues related to data quality.

- **Better data coordination — requests, extractions, and uses:** Centralizing and aggregating data sets results in an integrated approach to data extraction. By utilizing metadata values as a form of data classification and categorization, organizations can minimize data requests through a shared repository and tool set that can be controlled and governed consistently across the organization. More importantly, this kind of coordinated and integrated approach means that data extractions and collections can help minimize multiple requests for similar data from the IT or operations teams involved in the data extraction process. It also creates a consistent starting point where all data discovery and quality evaluations requests begin.
- **Continuous assessments of compliance controls:** Data centralization provides compliance leaders with the ability to validate data feeds into and out of their various systems, develop more predictive analytics in order to proactively identify potential data conduct problems, and enhance the risk and performance indicators derived from formally disconnected or unrelated repositories.

Increasingly, compliance leaders are expected to understand the current data state of the organization in order to support regulatory responsibilities and continuously improve quality and compliance controls. Without a matching surplus of budgets, resources, and funding to match the data being generated, compliance leaders must have a data services platform that is designed to address the needs of multiple systems and platforms, data rectification, data quality validation, and data integrity checks/corrections across multi-structured data types.

IN THIS CHAPTER

- Simplifying data governance management
- Introducing seamless data governance
- Keeping users productive
- Managing legal activities
- Ensuring continuous compliance

Chapter 4
Enforcing and Monitoring Data Governance

In this chapter, you find out about data governance and management issues.

Avoiding Dashboard, Report, and Alert Overload

With so much data available, so many silos containing data, and so much of the technology supporting the information supply chain generating logs and messages, a seemingly endless set of dashboards, reports, and alert messages can be, and often are, created. These dashboards, reports, and tools attempt to visualize the data in different ways to tease out details that might be relevant to one group, but can often be contradictory to the understanding or needs of other groups. Worse, dashboards, reports, and alerts

that are generated from inaccurate data or that offer ineffective insights suffer from many of the same confidence issues that data governance seeks to remedy.

TIP

Avoiding dashboard, report, and alert overload starts with ensuring that the source of these visualizations is based on data that is already in a state of effective governance. This allows you to use quality descriptors (metadata tags) to ensure the accuracy and consistency of the data underlying the dashboard or report. Building visualizations from data that users have confidence in will help promote the same confidence in these dashboards and reports, even if they don't favor the ideal decision.

Beyond basing your visualizations on data that is under effective management and control, additional considerations include these:

- **Know your audience.** Even if the data is accurate, dashboards, reports, and alerts are effective only if they are relevant to your audience.
- **Set priorities and classify alerts based on relevance.** Remember, not all systems or events are critical. Impose granular reporting controls with auto-escalation rules to ensure the right message is seen by the people who can take the appropriate action on it.
- **If everything is quiet or the metrics aren't changing, something is likely wrong.** Said differently, test, test, and re-test your reporting systems regularly to ensure that they remain accurate and drive awareness.
- **Timestamps can denote relevancy.** The last thing you want is for the business to take action on data that is old. When creating the dashboard, report, or alert, ensure that you include a date and time stamp that is easy to reference so that the viewer can evaluate a course of action based on the time elapsed from when the report was generated.
- **Document everything.** It is crucial that you document how you have set up the dashboard, defined the report, or configured the alert, as well as who it was done for and why.

Enabling Non-Invasive Data Governance

A successful data governance program is a seamless, non-invasive data governance program — one that minimizes the level of manual effort required and does not impede individual productivity, yet remains transparent, supportive, and collaborative.

What sabotages data governance the most is attitudes and perceptions. Many often perceive data governance as a set of policies and rules that result in more work to do on top of the work the employees are already responsible for. This misconception can foster attitudes that make it difficult for organizations to get people to adopt data governance best practices.

Although it is true that your data governance strategy mandates how data will be managed and identifies what is and is not permitted with regard to the data, the approach does not have to be — and should not be — intimidating or invasive. However, governance, by definition, requires that something is going to be administered. In the case of data governance, your data and how it is used are what is being managed.

It is possible to have the command and control of your data that you need, yet accomplish it in a way that is non-invasive to the rest of the organization and does not create additional work for your users. This can be accomplished with the automation and logic controls of a data services platform (see Chapter 3).

At an individual file level, an example of a non-intrusive strategy is the use of metadata tagging to encapsulate data governance rules, quality assessments, and usability metrics. As data flows through the information supply chain, systems of record can automatically interrogate these descriptors and apply other forms of control based on relevance factors, sensitivity of the content, or retention requirements.

The need to fence data geographically to adhere to country-specific data privacy rules provides another example of how non-intrusive data governance can be implemented. By automatically injecting geo-coding details into each file that can be accessed by a mobile workforce, access to the files can be validated based on where in the world the device is operating. This means that if a user travels outside the regional boundaries, the data remains in place, but the access to the data is automatically terminated until

that device returns to a location within those borders. In this case, data governance is simply the way you instruct and train your staff on their responsibilities.

You must also educate and train the users of your data. The most effective method is top-down. Rely on your leaders to keep the relevance of data governance and the importance of being an effective data steward at the forefront of everything they do, rather than adopting a "because I said so" approach. The rules must outline what is and is not permissible with respect to organizational data, especially data that contains sensitive information or personal details. A clearly defined and articulated set of policies, standards, and processes for your organizational data (we talk about this in Chapter 2) is an absolute necessity.

Empowering the End-User without Compromising the Business

The era of the digital enterprise has arrived. Businesses seek to drive higher levels of productivity, gain competitive advantage, and attract and retain top employees. The modern workforce wants to be productive from anywhere, using familiar personal applications and devices. Every IT department strives for visibility and control over corporate data, regardless of where it resides.

The digital workplace promotes new, more efficient and flexible ways of working, but it presents a whole new set of challenges for organizations because they need to maintain security, control, and governance in this new digital landscape. Further, IT leaders want to deploy a cohesive approach in accordance with business goals. Because goals often change, the digital workplace requires agility and efficiency to move with the times.

REMEMBER

The digital workplace is collaborative and dynamic. It must be productive and risk-averse by design.

Most organizations today are hampered by siloed or disjointed resources. Add the turmoil of "shadow IT" (frustrated business users find, implement, and use their own IT solutions) and unstructured data growth, and you face untenable business consequences.

Today's digital workplace requires productivity tools underpinned by intelligent data management. IT must evolve to deliver on the promises of the digital workplace with a modern data foundation that delivers the tools that employees want and the controls that IT needs (see Figure 4-1).

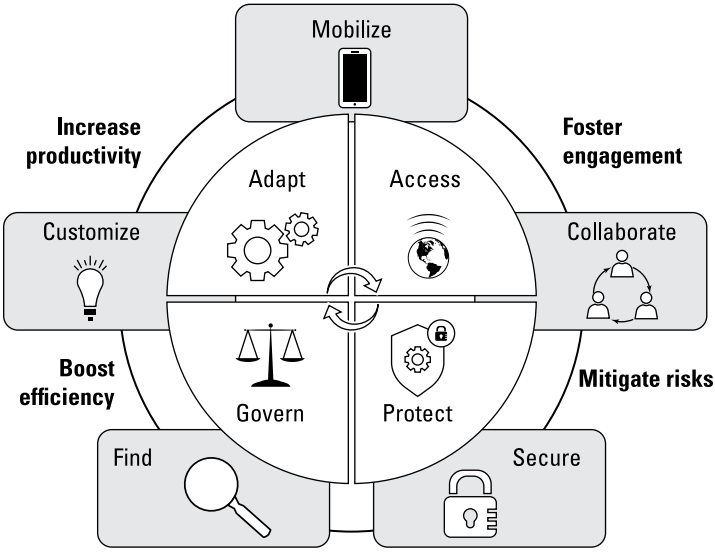

FIGURE 4-1: Empower your digital workplace and optimize productivity without increasing risk.

Supporting Data Governance for Legal Activities

A survey conducted by the Compliance, Governance and Oversight Council (CGOC) showed that most companies retain significantly more information than they need for business or legal reasons.

The Electronic Discovery Reference Model (EDRM) is a coalition of consumers and providers working together since 2005 to create practical resources to improve e-discovery and information governance. Published originally in 2014 by EDRM, the white paper titled *Disposing of Digital Debris* is still applicable today and provides recommendations for defining and identifying digital

debris, as well as proposing a coherent information governance strategy for cleaning up digital debris as follows:

"A digital disposal program must be defensible and requires:

- **People:** Leadership and commitment to guide transformational change.
- **Policies and processes:** Rules, regulations, and procedures that link information duties and value to data assets; and information demand to infrastructure supply.
- **Technology:** Tools that enable IT to implement and execute information governance policies and procedures.

With this three-step approach, organizations can begin to reduce the risk and overhead costs associated with the risky retention of digital debris."

The white paper emphasizes that while most organizations have records management policies in place and published on a website, it does no good if they do not put technology in place to support this initiative. Technology tools to implement and execute are often missing. It recommends that organizations leverage technology that can "automate legal holds, records retention, de-duplication, storage tiering, and deletion of data with no business, legal or regulatory value. To simplify overall implementation, it is desirable to use technologies that support a number of these capabilities within a single platform. Ideally, the chosen technology platform must also provide a central catalog itemizing the classes of and sources of data of end-users. Policy makers in legal, records, business and compliance must be able to view, understand and share this catalog."

Such a platform can help you gain actionable business insights with intelligent exploration of all your data, so you can do the following:

- Locate and identify the most relevant data regardless of its type or location.
- Identify data value with automated cataloging, transformation, and augmentation.
- Access relevant data with richer context available where and when you need it.

Leveraging Data Governance for Compliance Initiatives

The exponential growth of data, combined with increasing regulation, has left many organizations struggling with the complexity of compliance requirements and the most effective way to manage the issue. Many organizations are struggling to get a grip on data compliance for several key reasons:

- The total amount of data created, captured, copied, and consumed globally is forecast to increase rapidly, reaching 180 zettabytes in 2025, according to Statista. Therefore, the volume of data businesses must store, manage, and report on is becoming more difficult to administer.
- The types of unstructured data that must be captured are increasingly complex and include instant messaging (IM), short message service (SMS) text, voice, video, and social media interactions. Many businesses do not have adequate technology, tools, and policies in place to support this data.
- There are potentially thousands of pieces of legislation affecting companies, with more than 100,000 legal requirements relevant to multinational companies. This regulatory environment is constantly evolving and varies across markets, making compliance more difficult, especially for businesses that operate across geographies.

Until now, organizations have had a fairly well-defined set of sources of information to manage in traditional applications running databases (structured data). Structured data includes data that is managed centrally and is relatively easy to access and filter.

For example, a bank may have to govern, index, search, and provide content to auditors to show it is managing data appropriately to meet Dodd-Frank regulatory requirements. In the past, the information would have been retrieved from a database or email. Now, however, that bank needs to produce voice recordings from phone conversations with customers, show the Reuters feeds coming in that are relevant, and document all appropriate IMs and social media interactions between employees. These are all disparate data sources that the organization has never had to consider before, which creates new data and complexity

challenges. They are islands of information that seemingly do not have anything to do with each other. Yet, all have a significant impact on how the bank governs itself and how it saves any of the records associated with trading or financial information.

Coping with the sheer growth of unstructured and related external data is one issue; what to keep and what to delete is another. There is also the issue of what to do with all the data once you have it. The data is potentially a gold mine for the business, but most organizations just store it and forget about it. And, regardless of the medium, this storage is not free.

Legislation, in tandem, is becoming more rigorous and there are potentially thousands of pieces of regulation relevant to multinational companies. The EU, in particular, is subject to an increasing amount of regulation. There are a number of different regulations, including Solvency II, Dodd-Frank, the Health Insurance Portability and Accountability Act (HIPAA), the Gramm-Leach-Bliley Act (GLBA), Basel III, and new tax laws. In addition, companies face the expansion of state-regulated privacy initiatives and new rules relating to disaster recovery, transportation security, value chain transparency, consumer privacy, money laundering, and information security.

Many of these regulations vary across jurisdictions and, if you are an organization operating across several markets, a collaborative and integrated international approach is required for security, retention, and disposal. To compound the challenge, even when you might think you have a hold on a piece of legislation, you will find it has evolved and the policies and processes you had put in place to address it are no longer applicable or effective. It is not effective to address each legal requirement separately. Instead, a holistic approach to data governance is needed, which enables you to adapt to an increasingly regulated environment, yet remains seamless and transparent across your operations and business systems.

IN THIS CHAPTER

- Gaining business insights through effective governance
- Defining data quality and governance metrics
- Working with your executive team

Chapter 5
Enabling the Future of Your Business with Value-Based Data Governance

In this chapter, we explain how to close the proverbial "data governance gap" and use the single source of truth that data governance can create to make sense of the data you have and expose new insights into your business.

Transforming from Data Governance to Business Insights

Historically, data governance has focused on data security, protection, management, and control, along with the processes necessary to support regulatory compliance requirements. Like everything else today, data governance is transforming. Certainly, it will continue to address these topics and impose data quality

standards, but its primary focus is shifting toward delivering value to the enterprise with practices that are helping to ensure that data is optimized for use in analytics functions.

REMEMBER

The upward trajectory of data growth and complexity continues to accelerate year-over-year as organizations continue to demonstrate the ability to capture and store vast amounts of data at unprecedented rates. With all of this data, ensuring it is accurate, discoverable, and accessible is at the heart of the data quality process and is the key to successful analytics processes meant to deliver business insights to stakeholders and senior leaders.

The amount of data, coupled with the responsibility to drive value, can be paralyzing, but keep in mind that business insights are best created from accurate data that can easily be referenced in a timely manner by authorized users (we discuss those data governance attributes in Chapter 3). Then it becomes as easy as commitment, support, and focus on achieving the highest quality data underlying critical business functions. Improving the quality of these critical data assets begins with prioritizing data with the biggest business impacts and developing a lexicon of reference data that is consistent in its content and its use for different lines of business within the organization. This is often done through a good machine learning data catalog that can use artificial intelligence (AI) to discover and profile the data and assist data stewards in curating and adding business meaning to it.

Once the critical systems are in order, data quality issues must span to include data sources that exist across multiple organizations. Remember that not all data is created equally — meaning that data that lacks a defined owner, that has lain dormant for years at a time, or that has been superseded by other systems of record are prime candidates to be offloaded and culled from the information landscape. Everything else that remains should have a tangible relevance to the business and fall under the same data quality standards we discuss in Chapters 2 and 3.

Accept that groups, departments, and lines of business are going to operate independently — a necessity to support agility and respond quickly to changing business conditions. With consistency in how high-quality data is managed and governed across departmental and organizational boundaries, the shift from data governance to business insights can be realized. Sophisticated data quality scores can be applied to this data through a

data catalog accelerating insights to how fit-for-purpose the data really is.

REMEMBER

However, technology alone cannot solve data quality issues, and no amount of machine learning will magically clean up human-generated data in the middle of the night. Employees must be on board with your data governance processes — they must be involved and take their stewardship responsibilities to heart. Only with this combination of the right technology, supporting the right processes, driven by everyone, can governed data truly be operationalized into insights upon which actions can and should be based.

Working with Data Quality Metrics and Governance Measurements

First, some good news — the data governance process that is continually referenced is the starting point necessary to determine if information is complete and accurate. Depending on the characteristics of the data and how detailed your governance rules are, it is likely that much of the process can be automated using data profiling services inherited in the technology supporting the governance process.

For all other data, especially data that stretches across different areas of the organization, measure the quality of data on the following:

- **Completeness:** The percentage to which data fields contain the data expected. For example, if a field is labeled DoB (date of birth) but the field is empty or contains letters, it is fair to say the data is not complete and there is an upstream problem in the data collection process.
- **Uniqueness:** Measuring the contents of a data source against another to understand where data redundancy occurs, as well as if that redundancy is based on duplication or obsolete data. The same measurements can be applied at a file level by comparing two files.
- **Timeliness:** The length of time the data in question is valid before it is superseded by updates and content changes.

This is a key metric to watch, especially in highly transactional workloads, and in cases where copies of the data are used.

- **Validity:** Does the data conform to the governance standards that have been defined for its type and use?
- **Accuracy:** Does the data refer to or make use of the taxonomies in established reference data? Does the data correctly reflect that which it identifies?
- **Consistency:** This is a measurement of the degree to which data is standardized in format and content. How well does the data align to preconceived data patterns that are expected to be seen or automated against in the business? For example, do all dates share the same format and time zones, are currencies consistent, and so on?

This list can be extended to incorporate data governance specific metrics related to authorized access, authorized use, the degree to which data is mobilized, data recoverability, data security characteristics, and more.

Providing Data Governance Support for the C-Suite

Providing data governance support for the C-Suite starts with the C-Suite providing support for data governance.

Enterprise data governance policies must be established and driven at the C-Level. A top-down approach with executive backing has the power necessary to reign in recalcitrant stakeholders and mandate enterprise-wide consensus. Strong C-level backing can be the difference between an organization that's riven by competing and irreconcilable data governance standards and one in which data governance is consistently practiced across all domains.

WARNING

Avoid using C-suite power to demand that everyone must boil the data governance ocean by trying to go too big or run simultaneous governance initiatives across all domains. An iterative approach works best and serves executives with more control over the data they need to take informed action on the challenge or opportunity at hand.

With executive support behind the data governance program, it's time to reap the rewards. Some of the benefits that data governance provides to executive management include

- **Better decision making:** In both the organizational process and decision making, well-governed data is more discoverable, making it easier for the relevant parties to derive useful insights. This also means that decisions will be based on the right data, ensuring greater accuracy and overall trust.
- **Operational efficiencies:** As an asset of the business, the data governance process continually validates the fit and function of the data and its referential value to the business, as well as the decisions that are made by the departments for which these executives are responsible. Those who must make department-specific business decisions can rely on the same complete and accurate data.
- **Improved data understanding and lineage:** Executives can be assured that the governance process can point to what data they have, where it is stored, how it is being used, who is using it, and who is responsible for it. This understanding translates into timely responses to audits, more effective early case assessment activities, and a more proactive approach to preventing data corruption and breaches.
- **Regulatory compliance:** Whether self-regulated, regulated by industry or vertical, or regulated by a government entity, data governance is a critical aspect of ensuring and proving organizational alignment to the rules set forth in the regulatory requirements.
- **Increased revenue:** A welcome side effect of improving operational efficiencies and the culmination of all of the benefits of intelligent data governance is that it should help make the business better with faster decisions and more certainty in its actions.

REMEMBER

Although intelligent data governance provides more benefits to the C-Suite, the most important thing to remember is that to be a truly data-driven and transformative organization, data governance is not optional.

IN THIS CHAPTER

- Ensuring data completeness, accuracy, and alignment
- Eliminating confusion and enabling real-time decision making
- Sharing insights and fostering collaboration

Chapter 6
Ten Keys to Intelligent Data Governance

Although many reasons exist to adopt an intelligent data governance approach, adopting a robust governance model enables data accessibility, data confidence and understanding, and data activation, and it delivers the following ten benefits:

- **Data consistency ensures completeness and accuracy.** These are the foundations of data that can be trusted and are the basis for continually improving process models, data categorization/classification, and enterprise definitions that all business decisions will be based on.
- **Proactive data quality checks ensure data alignment.** With intelligent data governance, the common problems that arise from disjointed data points and siloed data repositories can be addressed in an automated fashion, with actionable insights related to adherence of data to defined quality and governance rules.
- **Data alignment is critical for regulatory and compliance responsibilities.** Intelligent data governance standardizes data quality standards, which reduces the risks and unplanned costs associated with basing decisions on misleading data, and also ensures accurate and timely adherence to regulatory and compliance requirements.

- **Removes confusion over data meaning and clarity.** Data confusion is a data governance problem whereby the data is incomplete or the processes supporting the data do not do enough to balance the rigors necessary for completeness with the speed at which the business must operate.
- **Analysis and decisions are based on well-defined and accurate data.** Intelligent data governance guides the structure and flow of data through the information supply chain — especially during analytics processes. Governance ensures that your data capture mechanisms are set up to collect what data you need and establishes alignment between the tactics of the lines of business and the organization's larger strategic goals.
- **Fact-based decisions become real-time events throughout the organization.** Intelligent data governance is key to ensuring data veracity, which in turn builds the confidence needed by the users of that data to achieve the real-time goal for decision making.
- **Data trust encourages the sharing of insights.** An organization that has a strong data-centric culture and fluency in data will only be able to encourage the sharing of information and insights if the data is complete and accurate. Intelligent data governance is much better for business because it requires a sharing of the burden of data and the insights gleaned from it throughout the organization.
- **Intelligent data governance fosters collaboration and establishes accountability.** Managing and controlling the use and proper maintenance of data based on a standardized set of rules or policies eliminates inefficiencies in the system, boosts collaboration between business units, and fosters a greater degree of accountability regarding who's responsible for data.
- **Data is kept clean and relevant based on its referential value.** The amount of data stored doesn't matter if it's relevant to the business and devoid of inaccuracies. An intelligent data governance approach makes everyone stewards of the data, responsible for keeping it in good shape.
- **Intelligently governed data gives you a competitive advantage.** Whether data is centralized or dispersed throughout the organization, when it is effectively managed and controlled, the process of gaining valuable insights and unlocking new opportunities is easier to achieve.

Derive more value from your data

Data is your most valuable asset, and you need to make the most of it. This book explains how intelligent data governance solutions that extend across your private and public cloud data footprint can help drive more business value from your data and simplify compliance for your organization.

Inside…

- Create an intelligent data governance plan
- Automate policies, standards, and processes
- Design for data quality and model for information relevance
- Ensure continuous compliance
- Enable the future of your business with value-based data governance

HITACHI
Inspire the Next

Go to Dummies.com®
for videos, step-by-step photos, how-to articles, or to shop!

ISBN: 978-1-119-88638-9
Not For Resale

for dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.