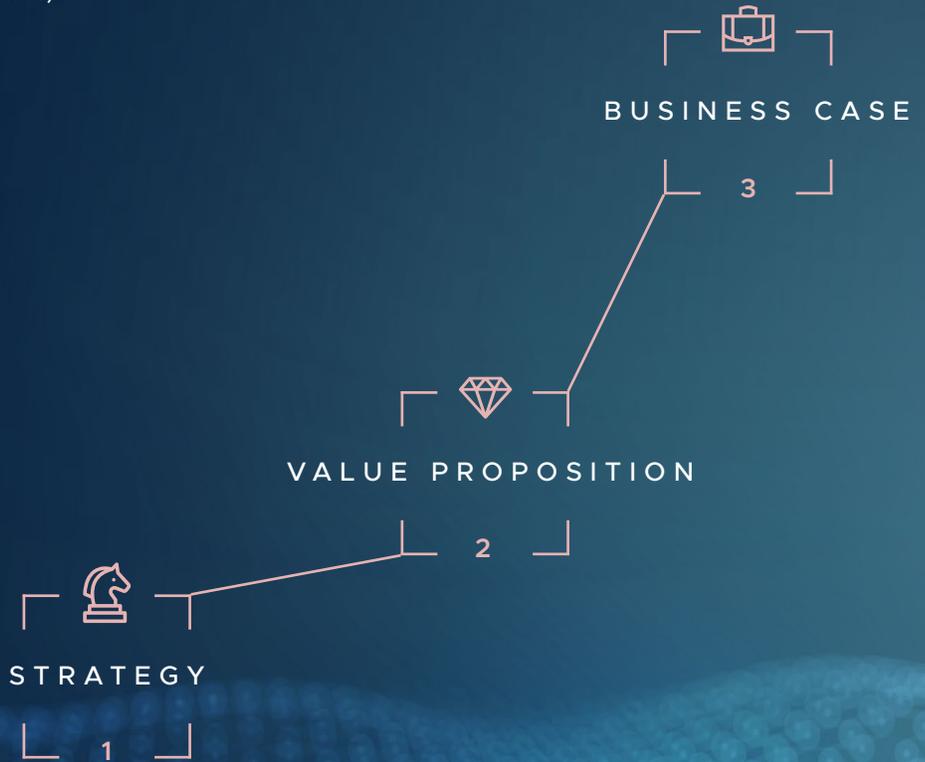


# Implementing Data Governance

How to Develop a Strategy,  
Determine a Value Proposition,  
and Build a Business Case



AUTHORS:

Sharad Varshney, Kieron Allen

11/2020

# Contents

**3 Executive Summary**

---

**4 Data Governance Defined: What it is and Why it Matters to Modern Organizations**

5 *The Rise of Collaborative Data Analysis*

7 *Data Privacy Regulations are Increasing*

---

**9 What is Traditional Data Governance and Why is it So Difficult to Run?**

---

**18 Modern Value-Driven Data Governance: Utilizing the Power of Smart Technology**

19 *Data-Driven Decisions and Innovation*

22 *Ensuring Data Compliance*

24 *Improving the Efficiency of your IT and Data Teams*

---

**27 What is the ROI of a Modern Data Governance Program?**

27 *Efficiency*

28 *Compliance*

29 *Better Business Decisions and Innovation*

---

**31 Conclusion**

---

**32 About the Authors**

## Executive Summary

 **Traditional data governance strategies and toolsets cannot effectively manage the terabytes of data accumulated since the advent of the internet.**

Since data was first stored and managed, systems that enable users to locate and work collaboratively on it have been essential—think of the antiquated referencing systems that still exist in some libraries today.

However, traditional data governance strategies and toolsets cannot effectively manage the terabytes of data accumulated since the advent of the internet.

In the age of big data, modern data governance techniques make it possible for multiple teams and departments to access data and use it to innovate. Innovation using data accessed through self-service platforms is a crucial driver for growth and today, modern companies needn't rely solely on dedicated data teams to achieve this success.

At the top-end of the scale, the most cutting-edge data governance programs allow for progressive implementation, enabling users to develop customized data governance programs at their own pace.

In this whitepaper, we'll define data governance, compare and contrast traditional and modern data governance techniques, and determine how data-driven innovation is fundamentally altering the mechanics of growth.

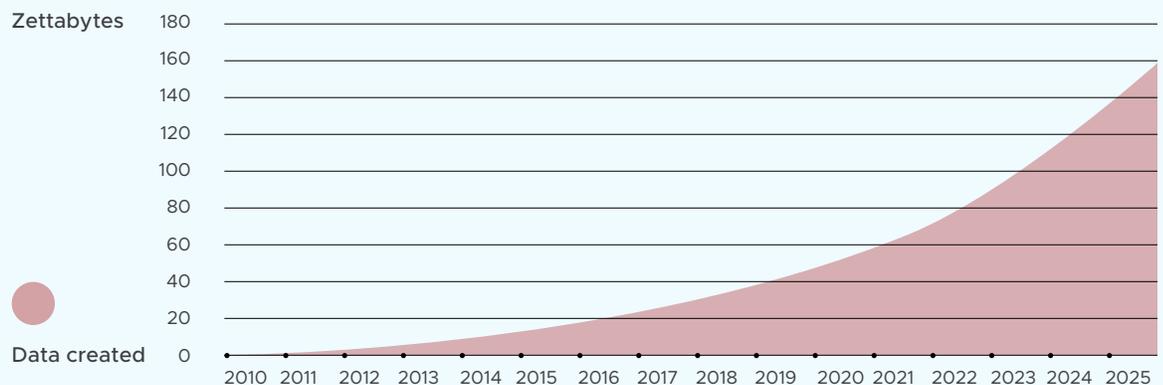
# Data Governance Defined: What it is and Why it Matters to Modern Organizations

“Today, data is king. This ‘digital gold’ is facilitating a significant shift in business practices and driving unprecedented growth.”

Data governance is the process of organizing, securing, managing, and presenting data using procedures and technologies that ensure it remains accurate, consistent, and available to verified users.

Effective data governance frameworks encourage data-driven decision-making, growth, and innovation, help organizations achieve compliance with company policies and privacy regulations from various jurisdictions, and enable IT and data teams to work more efficiently.

Annual Size of the Global Datasphere<sup>1</sup>

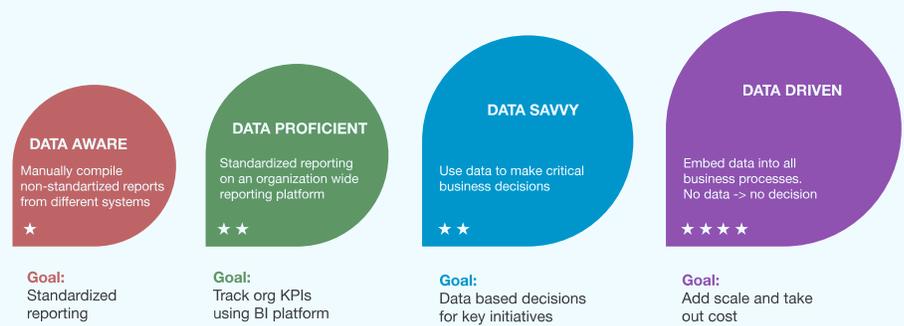


Today, data is king. This ‘digital gold’ is facilitating a significant shift in business practices and driving unprecedented growth. As a result, business executives, data owners, and departments are under enormous pressure to make the most of the data in their possession.

<sup>1</sup>Reinsel, David, Gantz, John, Rydning, John. “The Digitization of the World From Edge to Core.” Seagate, November, 2018, <https://www.seagate.com/files/www-content/our-story/trends/files/dataage-idc-report-final.pdf> (PDF version downloaded November, 2020)

But this is just one pressure point—data privacy is another. Data protection is becoming increasingly important to users, governments, regulatory institutions, and the business community.

Consequently, a comprehensive alignment process is essential for a modern company wishing to use its data to become more business efficient. This process is also vital to develop innovative products and services while conforming to regulatory requirements.



**“To achieve modern digital objectives, smart solutions are required. And that’s where data governance comes in.”**

The primary aim is to achieve a level of data maturity that enables companies to go from data-aware to data-driven organizations. But for this transformation to happen within regulatory boundaries, a company’s data (which on average, equates to around 400 different data sources), and the tools, strategies, and frameworks required to manage it, must be professionally implemented.

To achieve modern digital objectives, smart solutions are required. And that’s where data governance comes in.

## The Rise of Collaborative Data Analysis

IT and data decisions are now analyzed collaboratively at many of the most influential organizations in the world. There are some obvious benefits: a more democratic working environment, a culture of innovation, and the value of projects exceed expenditure.



**With so much data and so many advanced analysis tools available, it's possible to obtain more and more value from combined data sets.**

But regardless of collaboration, when business users are unable to make and implement critical decisions harmoniously, negative consequences soon begin to outweigh any original advantages.

For example, a company decides to move certain operations over to the cloud to save money on operational costs. The trouble is, cloud providers are not responsible for a company's data management commitments. So, despite the migration effort, co-ordination requirements would remain on-prem.

Splitting the storage and management of data like this can only lead to confusion—unless there is a substantial governance strategy in place.

Organizations face similar obstacles when rolling out company-wide data-focused solutions too. For example, single departments may require specific datasets at a particular time. This data is likely to concern a different customer group or product to another department.

Even if the product or customer data is from the same source, it could have undergone subtle changes that expose personal identifiable information (PII) or alter its meaning.

Data is continuously duplicated and used by various teams within an organization. Most enterprises hold at least ten different copies of their structured data, but this number is often much higher.

Despite this abundance of data, often, different teams within an organization don't have a full, or in some cases, even a partial understanding of the data available to them. Insufficient data visibility is an issue that many organizations face—especially when it comes to combined data sets.

With so much data and so many advanced analysis tools available, it's possible to obtain more and more value from combined data sets. Instead of focusing on independent sources, data teams look at historical records to acquire the most relevant and reliable information available to them.

But to facilitate a culture of democratic data consumption, organizations must first implement the technology to manage their data successfully.



## Data Privacy Regulations are Increasing

Today, data is a tightly regulated commodity, and in many instances, regulatory compliance is taking precedence over ingenuity. The General Data Protection Regulation (GDPR)<sup>2</sup> is probably the best-known data privacy law—with the threat of fines totaling 4% of revenue, this is hardly surprising—but outside of the EU, there are many more regulations.

Today, close to 70% of countries have data privacy laws in place<sup>3</sup>, but with constant violations and data misuse incidents across the board, more regulations will come.

Businesses are trying to promote a culture of innovation through creative data usage. However, they must also ensure that innovative practices don't expose sensitive data and force them to fall foul of regulatory requirements.

**It's a delicate balance, and modern businesses need to address two key questions:**



**1** With such diverse and disconnected data sources, how can we introduce holistic data management processes?

**2** How can we support a culture of innovation while ensuring everyone has the appropriate level of access to the data they need to succeed in their efforts?

**In this whitepaper, we answer these critical questions.**

---

<sup>2</sup> European Commission. "Data Protection in the EU." Accessed November, 2020, [https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en)

<sup>3</sup> UNCTAD. "Data Protection and Privacy Legislation Worldwide." Accessed November, 2020, <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>

**“ Progressive data governance enables businesses to implement data governance practices at their own pace on a modular basis.**

We discuss the ways traditional data governance failed to keep up with the demands of data-driven economies and how companies like OvalEdge are deploying modern methods to make the process easier, quicker, and more effective.

At the forefront of data science, [OvalEdge operates a progressive data governance model](#). Businesses can avoid many of the hurdles that arise during the implementation phase by using this method. When a company migrates onto a full-fledged data governance platform, it can often fail to succeed. The organization doesn't have time to learn the mechanics of the programs on offer or the opportunity to hone in on the areas where a data governance program may be most relevant.

Progressive data governance enables businesses to implement data governance practices at their own pace on a modular basis. This method allows them to learn strategies, dedicate resources as and where they are needed, and develop effective, scalable data literacy programs that ultimately lead to a self-service culture within an organization.

# What is Traditional Data Governance and Why is it So Difficult to Run?

**Traditional strategies are implemented over multiple aspects and fail to encourage data literacy among non-data-focused teams.**

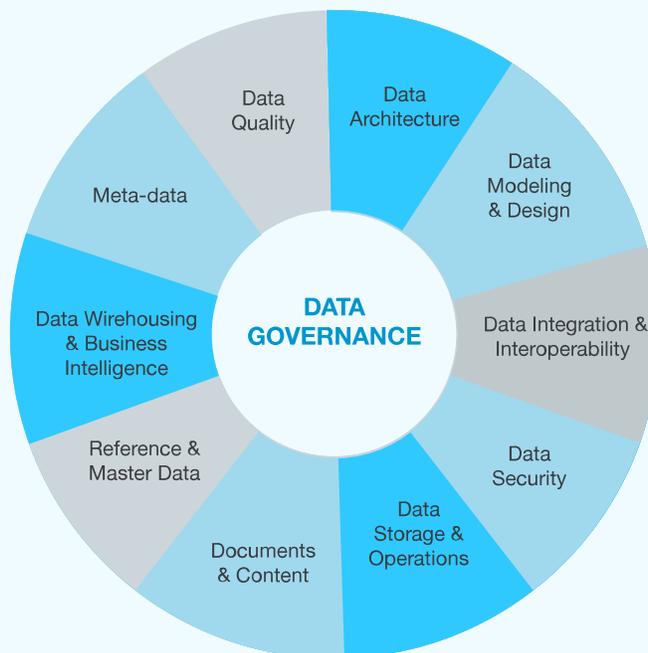
One can describe traditional data governance models as academic processes. They are incredibly useful, in-depth, and give data governance managers a great deal of control over the data in question. However, the requirement for expertise makes these procedures very difficult to implement company-wide. It is even more challenging to enforce and manage on an ongoing basis.

Traditional strategies are implemented over multiple aspects and fail to encourage data literacy among non-data-focused teams. Although these approaches are powerful, they neglect one of the most critical governance outcomes—the encouragement of data use.

Traditional data governance follows the DAMA framework. The biggest problem with this framework is that it is prescriptive and intrusive. Data departments do not generally follow the framework verbatim as their deadlines drive these departments. And when it’s obstructive, the DAMA framework creates a bottleneck in the execution of data projects.

The following terms define traditional data governance practices<sup>4</sup>.

The DAMA framework of Data Governance



<sup>4</sup> Dataversity. "The Difference Between Data Governance & Data Management." Accessed November, 2020, <https://www.dataversity.net/the-difference-between-data-governance-data-management/>

## Data Architecture

In simple terms, data architecture is about identifying the data needs of an enterprise and designing and maintaining the master blueprints to meet those needs. These master blueprints guide data integration, control data assets, and align data investments with business strategy.

### Data architecture group goals →

- » Identify data storage and processing requirements
- » Design structures and plans to meet the current and long-term data requirements of the enterprise
- » Strategically prepare organizations to quickly evolve their products, services, and data to take advantage of business opportunities inherent in emerging technologies.

### Data architecture group responsibilities →

- » Define the current state of data in the organization
- » Provide a standard business vocabulary for data and components
- » Align data architecture with enterprise strategy and business architecture
- » Express strategic data requirements
- » Outline high-level integrated designs to meet these requirements
- » Integrate with the overall enterprise architecture roadmap

From a governance perspective, this group plays a vital role in the overall data governance strategy. In most cases, this group is either:

- » Responsible for all data governance duties in an organization
- » Required to work closely with data governance team members, like data stewards

### Governance responsibilities →

- » Maintain and enforce standards
- » Manage architecture designs
- » Oversee projects



## Data Modeling and Design

An excellent way to understand data modeling and design is to compare it directly to data architecture. Data architecture processes constitute the overview of data management requirements, while data modeling and design is a secondary outcome.

Essentially, data modeling and design is the production of graphs, diagrams, and other documentation—physical, logical, or conceptual—that demonstrate and communicate a company's data assets. Data modeling is a highly complex process for which data scientists or data engineers are responsible.

### Governance responsibilities →

- » Create standards for data modeling and ensure the entire organization follows these standards
- » Maintain the quality of data models and database designs
- » Maintain version controls
- » Ensure that these data models are available company-wide

## Data Storage and Operation

Data storage and operation refers to the design, implementation, and support of stored data to maximize its value. Most organizations have various databases (SQL, No-SQL, data lakes, etc.), maintenance systems, backups, encryption protocols, and various other activities.

### Key goals →

- » Manage the availability of data throughout its life cycle
- » Ensure the integrity of data assets
- » Manage the performance of data transactions



Database administrators (DBA) play a vital role in data storage and operations. The role of the DBA is the best-established and most widely adopted role in the data industry, and database administration practices are the most mature of all data management disciplines.

**From a governance perspective, the following should be available** →

- » Various storage metrics, like the number of databases, transactional statistics, capacity metrics, storage metrics, number of requests, and improvement services
- » Performance metrics like transaction frequency and query performance
- » Services metrics like issue submission KPIs and resolution times
- » Information asset tracking tools that ensure the company meets license requirements and the cost of ownership is understood
- » Data audit and validation processes to evaluate stored data against established acceptance criteria to determine its quality and usability

## Data Security

Once an organization decides on its data storage methods, the challenge is ensuring the data remains secure. When data is stored on-prem, it's down to dedicated IT professionals to develop security systems that prevent third-parties from accessing, maliciously appropriating, or altering it.

But this challenge doesn't stop at external threats. Data security protocols should also prevent unauthorized users within an organization from intentionally or unintentionally accessing or manipulating prohibited data sets too.

**Data security goals** →

- » Enable appropriate and prevent inappropriate access to enterprise data assets
- » Understand and comply with all relevant regulations and policies for privacy, protection, and confidentiality
- » Ensure that the privacy and confidentiality needs of all stakeholders are both enforced and audited



To achieve these goals, IT security teams use various tools and techniques like encryption, antivirus software, malware attack prevention, and more. Database administrators and IT security teams are usually responsible for managing data security.

Today, regulatory guidelines are making it more difficult for security teams to find all available privacy-focused data assets and govern them appropriately. Furthermore, granting access to individual users to specific data creates a backlog that the security team is required to manage.

The objective of a governance team is to ensure that all the procedures are well documented and encryption and standards met.

## Data Integration and Interoperability

Data integration is the process of funneling data between numerous data stores, applications, and organizations. It is the most common business need and a vital requirement for any data solution. Data engineers are usually responsible for creating and managing these data pipelines.

### Data integration goals →

- » Provide data securely, with regulatory compliance, in the format and time frame required
- » Lower the cost and reduce the complexity of managing solutions by developing shared models and interfaces
- » Identify meaningful events and automatically trigger alerts and actions
- » Support business intelligence, analytics, master data management, and operational efficiency efforts

### Data integration requirements →

- » **Data Sharing Agreements:** Before developing interfaces or the provision of data electronically, there was a requirement to establish a data sharing agreement or memorandum of understanding (MOU). This agreement stipulates the responsibilities and acceptable use of data to be exchanged and is approved by the business data stewards of the data in question.
- » **Data Lineage:** Data lineage is a crucial requirement for data governance as, without it, you cannot conduct any impact analysis when making data changes. Analysts must document the data lineage and confirm how data is flowing from one system to another.
- » **Metrics of Data Integration:** Metrics are required to measure the scale and benefits of a data integration solution on availability, usage, volume, cost, and speed.



## Document and Content Management

Data exists in many formats such as PDF, text file, JPG, or one of many other document types. Everything else, like MP3 or MP4, can be considered content and is stored in any format other than RDBMS (Relational Database Management Systems).

Several crucial steps include organizing and categorizing data, developing storage solutions, implementing workflow protocols, editing the data, publishing, and archiving.

**You must have governance in place for unstructured data for the following reasons:**



- 1** Legal and regulatory compliance
- 2** Defensible disposition of records
- 3** Security of sensitive information

## Reference and Master Data

Although similar, reference and master data are two separate things. Master data is the core data contained within an organization and could be customer data, data referring to inventory or stock, primary analytical data, or something similar. You can characterize master data by how it is stored (on multiple systems) and shared (by numerous members of an organization). For example, in the retail industry, consolidating customer information is a master data management (MDM) activity.

Reference data is the set of values used to structure this master data with a focus on shared or common indicators. For example, in the global stock market, a trader may well be aware of the tickers that represent each stock even if they have no other detailed information about it.

Aggregating master data and categorizing reference data is a complex process that requires various tools and techniques.

**The following are the MDM activities** →

- » **Define MDM drivers and requirements by analyzing your business**
- » **Evaluate and assess all data sources**
- » **Define and design a data architecture approach**
- » **Model master data attributes, define enterprise-wide definitions and subject areas**
- » **Define the maintenance process**
- » **Establish governance policies**

All MDM activities should have a governance overreach. Without governance, reference and master data solutions are just additional data integration utilities, unable to deliver their full potential.

## **Data Warehousing and Business Intelligence**

Business intelligence (BI) refers to the strategies and technologies deployed by organizations to analyze business-critical data. Data warehousing is a vital component of BI. A data warehouse can contain all of a company's data, current and historical, from numerous sources inputted by multiple users.

From here, data analysts are theoretically able to access any data they need to make crucial business decisions. Traditionally, IT teams used an Extract, Transform, and Load process (ETL) to upload and store data in a data warehouse. In this manner, data is moved in batches and on daily schedules. But there's a limit to how much you can move at once, so traditional data warehouses often require updating. This method also requires a lot of resources like CPU, memory, and bandwidth.



Commonly, governance in BI groups is the primary driver for an entire data governance program.

### Governance objectives for BI groups →

- » **Enable business acceptance by:**
  - Documenting data models
  - Ensuring a data quality feedback loop
  - Providing end-to-end metadata
  - Offering verifiable data lineage
- » **Customer user satisfaction**
- » **Define Service Level Agreements**
- » **Ensure a reporting strategy for the entire data landscape**

## Metadata

Metadata is data in the fine print, the information used to find and categorize data. As well as making data discoverable, you can use metadata to find common relationships between data sets too.

Metadata is intrinsically linked to data quality because the information contained within it gives data provenance. But without a system in place that automatically analyzes metadata and uses it to categorize and qualify this provenance, it's impossible to get the most from metadata on a large scale.

### Metadata management objectives →

- » **Establish standard business terms or develop a widely-distributed business glossary**
- » **Collect metadata from all data sources**
- » **Standardize the way business users access metadata**
- » **Ensure the quality of metadata**

It is the role of the governance teams to establish metadata standards and guidelines.



## Data Quality

Good data quality improves the overall usage of data and makes data-driven decisions a reality. Good data quality is one of the primary objectives of a data governance program. If you follow all of the processes explored in this section, you end up with quality data. Teams need to tag data accurately for provenance, it needs to be stored safely and securely, and it must be well-referenced.

### Data quality team objectives →

- » **Develop a governed approach to make data fit for the needs of consumers**
- » **Define standards and requirements to achieve data quality**
- » **Define and implement processes and procedures to measure the quality of data**
- » **Identify and champion data quality enhancement through various process improvements**

As you can see, traditional data governance deals with multiple departments and its various functions, and alignment with all the departments and their data sources is a daunting task. Furthermore, traditional data governance approaches do not provide a mechanism to measure the success of a data governance program. Consequently, it is hard to justify the investment.

### So, what's the alternative?

Although incredibly useful, traditional data governance fails to achieve the efficiency, cost-effectiveness, and simplicity of modern data governance tools. Even though early data governance models deal with multiple aspects, there is no step-by-step approach. Multiple toolsets are required, and the implementation process is lengthy and convoluted.

What's required is a centralized, value-driven platform, like the OvalEdge data catalog and governance tool, that's easy to implement and manage. In the following section, we'll explain how modern data governance, through the use of smart technology, enables any company that aspires to grow into a data-driven organization to do so.



# Modern Value-Driven Data Governance: Utilizing the Power of Smart Technology

“Today, data is everywhere, and data governance protocols have had to adapt.”

Most traditional data governance practices were developed in the late 1990s and driven by the banking industry’s compliance requirements. Since then, the scope of data distribution and the industries that deal with data has grown tremendously.

Today, data is everywhere, and data governance protocols have had to adapt. Now, organizations define good data governance by the value it can bring—especially when quality data is the foundation of this value. With a shift in both data volume and strategy, an organization must govern its data smartly using modern methodologies, use cases, and technologies.

In the early days of data governance, there was a great deal of focus on developing specific data architecture. Now, using modern data governance technologies, like the OvalEdge data catalog and governance tool, architecture diagrams are automatically built using the raw data.

## Modern data governance has three core objectives

→

**1** To advance data-driven decision making in an organization through trusted insights

**2** To ensure data compliance across various data privacy laws and internal data policies

**3** To improve the efficiency and productivity of IT and data teams





**Individual departments and employees can use the wealth of data and insights at their fingertips to develop game-changing ideas with company-wide implications.**

## Data-Driven Decision Making

When presented and managed correctly, it's possible to convert data into trusted insights that lead to better business decisions. This process also drives innovation in an organization.

Individual departments and employees can use the wealth of data and insights at their fingertips to develop game-changing ideas with company-wide implications. But before an organization can innovate with its data, several processes need to be completed, or at least activated, first.



### DATA LITERACY

Data literacy is all about education. It's the process by which an organization puts in place measures to ensure all data users within that organization receive education to a level that enables them to consume data confidently.

In general, regular business users cannot utilize the data available to them as effectively as a dedicated data team can. When widespread data literacy and clearly defined data terms and frameworks aren't in place, communication channels can break down, and innovation stalls.

Developing and implementing a comprehensive data literacy strategy enables companies to avoid mixed messaging and cross-department confusion. When data is misrepresented or misunderstood, not only is it hard to get the best out of it, it can lead to potential conflicts too.

Having a dedicated data team is a powerful asset, but give everyone in a company the means to locate and utilize this data, and you can transform a business from the inside out.

Suppose business users are unaware of the data available to them. In that case, they will struggle to ask helpful questions based on it, find useful answers, and come up with innovative solutions for growth. Education is vital for progress, and in a BI context, data literacy is the educational process required to drive growth in a modern company.





## TRANSPARENCY AND TRUST

To build a culture where users can effectively utilize data, the way you distribute, store, and manage it must be transparent. This transparency leads to trust in data, and when users consider data trustworthy, they will do more with it.

You don't have to make all the data in your organization available to everyone to be transparent. This action would lead to scenarios where protected or confidential data is made available to unauthorized users, in opposition to internal and external regulatory requirements.

Instead, a company should clearly state where and what its data is, where it's coming from, who is using it, who owns it, and whom to contact if you need access to it. The OvalEdge data catalog is one such way for organizations to present the data available to users while withholding access to restricted information.

With technology like this in place, organizations can build trust by displaying the location and content of the available data without necessarily giving everyone access.

Trust in data is also intrinsically linked to consistency and classification. Well-classified data is far easier to verify than inefficiently organized information.



## ACCESS

Another crucial driver of innovation is access to data. Once your staff is data literate, and data sets are transparent and trusted, the next step is to make it accessible. Without access to the data they need to develop new concepts and approaches, users can't innovate.

In an ideal scenario, all users have equal access to the data they require—as long as no restrictions are in place to protect PII or other information. This scenario is achievable through smart cataloging and classification.

The aim is to make data searchable so users can access it easily. When dealing with teams of people who are becoming more data literate but aren't proficient data analysts, accessibility is incredibly important. Not only does it instill confidence in newly literate business users, but it also encourages experimentation.





## SELF-SERVICE ANALYTICS

Self-service analysis takes place when a regular business user develops into a business analyst. It's best to regard self-service analytics as the result of a well-developed, modern data literacy program. After giving users access to data, they can train, experiment, and innovate. Eventually, these users will transform into analysts using that data to make better business decisions.

At the point where a business user becomes literate enough to access data without assistance from an organization's data team, they can begin to undergo self-service analysis. In modern data governance, self-service is one of the primary use cases for end-users. The other two are data literacy and data discovery—a concept we will explore further in this whitepaper.

Self-service analytics requires a platform, like the OvalEdge data catalog, from which users can access the data they need and analyze it as and when they need it. These tools foster a quicker decision-making process where innovative ideas can be created as a team or independently.

Self-service provisions enrich and empower team members providing them with new skills to make better business decisions.



## DATA QUALITY IMPROVEMENT PROGRAM

The more systems that are put in place to streamline the data governance process and track KPIs, the better the quality of data becomes.

Organizations must make a concerted effort to improve the quality of their data for users to get the best from it. However, they must also determine what they want from the data before deciding how good it is.

Every company department has a specific business purpose for the data they access. When data literate staff can access and analyze this data, they can determine the correct KPIs required to track how well it's performing.

Again, modern smart technology enables users to set and access these KPIs once data is unified and organized in one place.

## Ensuring Data Compliance

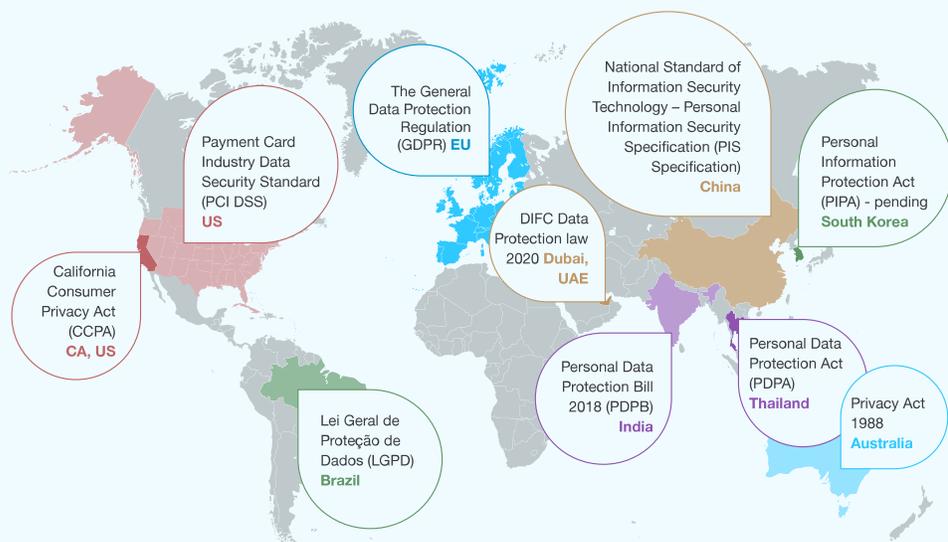
“Laws governing data privacy now exist in different countries, regions, and even states

As we mentioned earlier on in this whitepaper, the first data governance strategies were developed by the banking industry to aid compliance. Although not the only factor, compliance is still a driving force behind data governance practices today.

The difference is, the number of laws and regulations has erupted, and so too have the internal data policies of many organizations. Laws governing data privacy now exist in different countries, regions, and even states, and as businesses operate at a global level, it’s crucial to comply with these laws—regardless of where you are based.

We’ve highlighted some of the most significant regulations here:

Data Protection Map



“Today, failing to follow data protection regulations could result in staggering fines.

Today, failing to follow data protection regulations could result in staggering fines. For example, any organization that breaks the GDPR could face fines up to 20 million euros or 4% of global annual revenue<sup>5</sup>.

There are three key areas to consider if you wish to address compliance issues in modern data governance: standardization, knowledge, and lineage.

<sup>5</sup> GDPR.EU. “What are the GDPR Fines?” Accessed November, 2020, <https://gdpr.eu/fines/>



## STANDARDIZATION

Standardizing data is a crucial step to ensuring compliance. When data is standardized, it is easier to track and compare.

A vital compliance requirement is to ensure an organization can confirm the location of user data and its security. By standardizing data, analysts can easily categorize and trace it, ensuring it conforms to various regulatory requirements.



## CLASSIFICATION

Once data is standardized, it becomes easier to identify and enables organizations to classify and tag it.

Understanding data is vital if you want to ensure compliance. You need to know what it is and what it means, and classification is an important part of this process.

The next part is knowing where to put it. With a data catalog like OvalEdge's, users can easily search for and find information about the data they are using. This process is incredibly useful when it comes to ensuring compliance.



## LINEAGE

Data lineage refers to the lifecycle of data—where the data originates from and where it has been.

Lineage is vital for tracking data. With processes in place to confirm data lineage, it's possible to see if and how anyone has altered the data, who it belongs to, and what important information it contains.

In a compliance sense, data lineage enables organizations to achieve several objectives such as more efficient regulatory reporting, the ability to improve data governance through access to historical data, and the ability to expose any discrepancies or potential security threats.

Within the OvalEdge data catalog, there is a tool dedicated to tracking data lineage and displaying it visually.



## Data Lineage in OvalEdge



## Improving the Efficiency of your IT and Data Teams

“Most business executives expect Chief Information Officers (CIO) and Chief Data Officers (CDO) to do more with fewer resources.”

Today, most business executives expect Chief Information Officers (CIO) and Chief Data Officers (CDO) to do more with fewer resources. Budgets are getting cut, and there is a powerful drive to transform existing data models into modern systems.

This outcome is now the key objective, but data teams will only realize this goal by following several fundamental processes first.



### DATA DISCOVERY

Data discovery processes ensure that an organization’s data is easy to find, access, and understand regardless of where it is stored. The best way to achieve this is through a data discovery platform—like the OvalEdge data catalog, but how do provisions like these improve efficiency?

Generally, in an organization, data is neither located in the same place nor accessed in the same way. Instead, the opposite is true. Data is stored in multiple locations with countless different access measures in place. This method makes it incredibly hard to find and access data.

With efficient data discovery systems in operation, data is easy to locate because it is searchable. These processes slash the time it takes to find a particular data set. On top of this, comprehensive categorization makes the data within an organization easier to understand as the requirement to trawl through countless data sets to find specific information disappears.

Finally, when data is discoverable, users can collaborate on it. Data and IT teams can work together to use the data available to them to develop data-driven growth strategies. At the same time, business users can access the platform independently without help from these specialized staff members.



## IMPACT ANALYSIS

Impact analysis, in this context, concerns the processes IT and data teams undertake to determine the impact of data governance decisions downstream.

Impact analysis enables these teams to work more efficiently because before rolling out a data governance tool or protocol, they can systematically weigh up the pros and cons of this impending decision.

The first step of impact analysis is to do a business assessment. Using smart tools, both data and IT teams can quickly assess how introducing proposed changes will impact profits, workflow, and more. It is a critical BI process, and one that is becoming easier and easier thanks to the development of AI-driven tools that can quickly analyze the available data.

Of course, it is necessary to have data correctly stored, categorized, and managed first.



## METADATA MANAGEMENT

Earlier, we explained what metadata is and why it's so important—but let's recap now. Metadata provides context to information and enables users to work with it more effectively.

You must fully understand the data you are using if you want to get the most out of it. For this reason, one of the most important drivers of efficient data analysis is managing this metadata.

The vital point that data and IT teams need to grasp is that more metadata does not equate to more useful information. Instead, the primary utility of metadata comes from how it's managed.

Modern metadata management programs provide links to context, cataloging metadata in a way that makes finding these links straightforward and fast.

Metadata management requires a strategy, just like any other data governance process, and the standardization process used to make raw data more efficient can be applied to metadata too. When combined with a data catalog, this process boosts what can be done with metadata independently and collaboratively.

When managed correctly, metadata makes every aspect of modern data governance more effective. It provides accurate information to calculate impact analysis, it includes detailed user information required for regulatory compliance, it enables advanced users to track data lineage, and it provides analysts with more comparison points.

You can also use well-managed metadata to create a business glossary. This tool is incredibly useful for data literacy, and it's best to develop a business glossary after the classification of raw and metadata.

# What is the ROI of a Data Governance Program?



**ROI for modern data governance is measured by use case. The most important of these are innovation and data-driven decisions**

Determining an exact return on investment (ROI) for a data governance program is difficult. There are some areas where it's possible to calculate ROI and others where it isn't—right now. However, when looking at ROI as a whole, you can combine all of these elements from different use cases to confirm the benefits of stringent data governance provisions.

As we have explained, modern data governance is value-driven and always use-case specific. Smart technology calculates and comprehends pain points and develops solutions based on value.

ROI for modern data governance is measured by use case. The most important of these are innovation and data-driven decisions, the ROI on which can be exponential if you implement strategies correctly. However, the ROI of data-driven innovation is only calculable a reasonably long time after the event.

For this reason, let's start by documenting the ROI of increased efficiency, the use case that's easiest to calculate.

Using the three pillars of modern data governance—data-driven decisions, compliance, and efficiency—we will explain the ROI of a modern data governance strategy.



**Calculating the ROI of data governance in regards to efficiency is straightforward—you should recuperate your initial investment relatively quickly.**

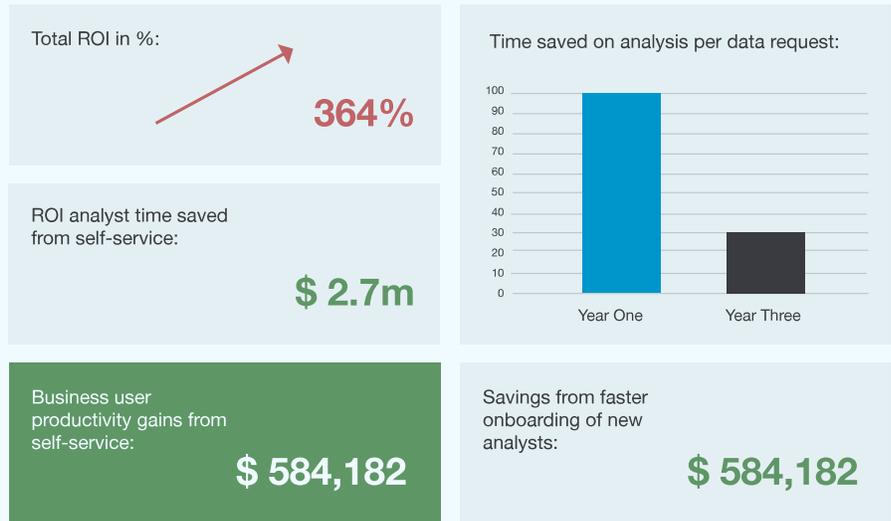
## Efficiency

Calculating the ROI of data governance in regards to efficiency is straightforward—you should recuperate your initial investment relatively quickly.

Improved efficiency within data teams leads to time saved within these specific teams and, because there are platforms in place to make data processes more efficient and lead to increased data literacy, organizations record savings downstream too.

Self-service has a significant impact on ROI. When users can access and use data themselves without the need to go via a member of the data or IT team, the ROI is staggering. One report by Forrester included analysis from seven companies that had used a modern data governance tool similar to the OvalEdge data catalog. Here are the findings over three years:

## Economic Impact of Modern Data Governance Tools<sup>6</sup>



“The EU’s GDPR has the highest penalty, but there are some other significant fines too.”

## Compliance

It’s almost impossible to calculate the ROI through compliance. However, what you can do is calculate the risks you mitigate by ensuring that you have a data auditing strategy in place. The best way to illustrate this is to look over the most comprehensive data protection regulations and evaluate what you stand to lose if you don’t comply with them—bear in mind any internal data or privacy protection protocols too.

Unsurprisingly, the EU’s GDPR has the highest penalty, but there are some other significant fines too.

Any intentional violations of the California Consumer Privacy Act can result in civil penalties of up to \$750 per incident<sup>7</sup>, while the DIFC Data Protection Law’s administrative fines range from \$20,000 to \$100,000<sup>8</sup>.

<sup>6</sup> Forrester. “A Forrester Total Economic Impact™ Study Commissioned By Alation.” Alation, October, 2019, <https://www.alation.com/wp-content/uploads/Forrester-TEI-Alation-Final-10.08.2019.pdf> (PDF version downloaded November, 2020)

<sup>7</sup> State of California Department of Justice. “California Consumer Privacy Act (CCPA).” Accessed November, 2020, <https://www.oag.ca.gov/privacy/ccpa>

<sup>8</sup> Dubai International Financial Centre. “Data Protection Law DIFC Law No. 5 of 2020.” Dubai International Financial Centre, October, 2020, [https://www.difc.ae/files/6115/9358/6486/Data\\_Protection\\_Law\\_DIFC\\_Law\\_No.5\\_of\\_2020.pdf](https://www.difc.ae/files/6115/9358/6486/Data_Protection_Law_DIFC_Law_No.5_of_2020.pdf) (PDF version downloaded November, 2020)

Failure to follow the Lei Geral de Proteção de Dados (LGPD)<sup>9</sup> can result in penalties equal to 2% of a private legal entity's, group's, or conglomerate's Brazilian revenue minus taxes for the previous tax year, or a fine of 50 million reals (\$9 million).

Finally, a fine of up to to \$1 million is payable if you breach Singapore's Personal Data Protection Act (PDPA)<sup>10</sup>. And these are not just threats. The list below displays the largest fines handed out to corporations for breaking GDPR guidelines as of January 2020.

British Airways – 204.6m Euros.

Marriott International Hotels – 110.3m Euros.

Google Inc. – 50m Euros.

Austrian Post – 18.5m Euros.

Deutsche Wohnen SE – 14.5m Euros.

1&1 Telecom GmbH – 9.5m Euros<sup>11</sup>.

## Data-Driven Decisions

Value through innovation is the most difficult ROI to calculate. Firstly, this is a company by company consideration and every single entity will use its data to innovate in different ways. How much they get from the data is dependent on how much they are willing to invest in data governance strategies. Secondly, data-driven decisions drive value over the long-term so it takes at least a year—more likely two to three years—to get to a stage where it's possible to analyze outcomes.

Business leaders make decisions every day. When these decisions are backed by trusted insights, it leads to better outcomes. This directly affects a company's top or bottom line. In terms of innovation by creating new data products, this is entirely determined by the sporadic nature of innovation and the independent speeds that it progresses at.

Even with a solid data delivery platform in place and all the data required to innovate at a user's fingertips, it's very difficult to predict when and how innovation will occur.

**“Business leaders make decisions every day. When these decisions are backed by trusted insights, it leads to better outcomes.”**

<sup>9</sup> GDPR.EU. “What is the LGPD? Brazil's version of the GDPR.” Accessed November, 2020, <https://gdpr.eu/gdpr-vs-lgpd/>

<sup>10</sup> Singapore Statutes Online. “Personal Data Protection Act 2012.” Accessed November, 2020, <https://gdpr.eu/gdpr-vs-lgpd/>

<sup>11</sup> Alkudmani, Fares. “GDPR: The 6 Biggest Fines Enforced by Regulators So Far.” Secure Privacy (blog), January, 2020, <https://secureprivacy.ai/gdpr-the-6-biggest-fines-enforced-by-regulators-so-far/>

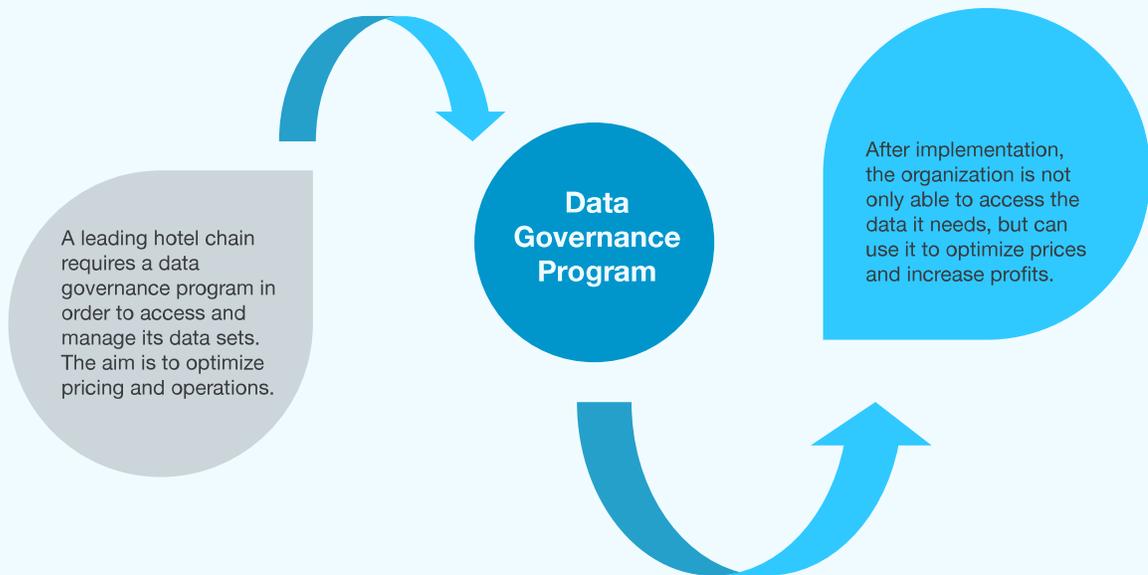
Over time, teams will build more use cases with the technology available to them and eventually there will come a pivot point where this new use case, say a recommendation engine, for example, is rolled out.

Even at this point, however, it's better to analyze the ROI only after lots of people have begun to use the technology. Although it's difficult to calculate ROI in regards to innovation, what is certain is that data is the foundation for digital transformation and every company is aware of its importance.

Imagine if you owned a plot of land and underneath this land, unbeknownst to you, was an untapped reserve of crude oil. You know the value of crude oil but you don't know that this great wealth is right below you. Until you explore it you will never be able to realize its value.

We don't know exactly how data-driven decisions and innovation will affect the way businesses operate in the coming years, but we do know that it will be transformational.

### Calculating ROI: OvalEdge Case Study



## Conclusion

With big data playing such an integral part in the growth, success, and evolution of businesses in every sector, data governance is already one of the most important considerations for management teams.

A strategic level initiative with far-reaching consequences, data governance practices have both short-term and long-term benefits. However, the extent to which a company can realize these benefits is entirely dependent on the degree to which a company is willing to innovate, using the data at its disposal.

Businesses can increase productivity using modern data governance techniques by reducing the cost of existing operations and building new data products faster. Company-wide innovation accelerates, and organizations can avoid penalties by remaining compliant with data protection laws and regulations.

Modern data governance is incredibly efficient, but even data governance providers should continue to innovate. Today, a large majority of data governance platforms operate using a one-size-fits-all approach. Organizations are introduced to data governance strategies all at once, regardless of their initial requirements or previous experience.

Implementing data governance in this manner can have negative consequences for businesses. Organizations are unable to scale their data governance procedures at a comfortable speed, and this can increase the chances of failure.

Progressive data governance enables companies to implement strategies gradually, at a pace they desire. It's possible to specify the scope of data used, the specific data governance programs you wish to implement, and how users will access these assets through self-service.

As the data landscape continues to mature and develop, so too must the way organizations implement modern data governance strategies.



**Businesses can increase productivity using modern data governance techniques by reducing the cost of existing operations and building new data products faster.**

## About the Authors



**Sharad Varshney** is a trailblazing entrepreneur and technology executive and voted one of the top 50 Indian entrepreneurs in the US. Before co-founding OvalEdge, Sharad was a principal architect at the innovative, California-based software firm, Hortonworks.

*Sharad Varshney, CEO, OvalEdge:*

<https://www.linkedin.com/in/sharadvarshney/>



**Kieron Allen** is a highly-experienced journalist, copywriter, and editor well-versed in both print and digital content, with experience writing for internally-distributed magazines, websites, and pioneering technology firms.

*Kieron Allen, Freelance Journalist and Editor:*

<https://www.linkedin.com/in/kieron-allen/>