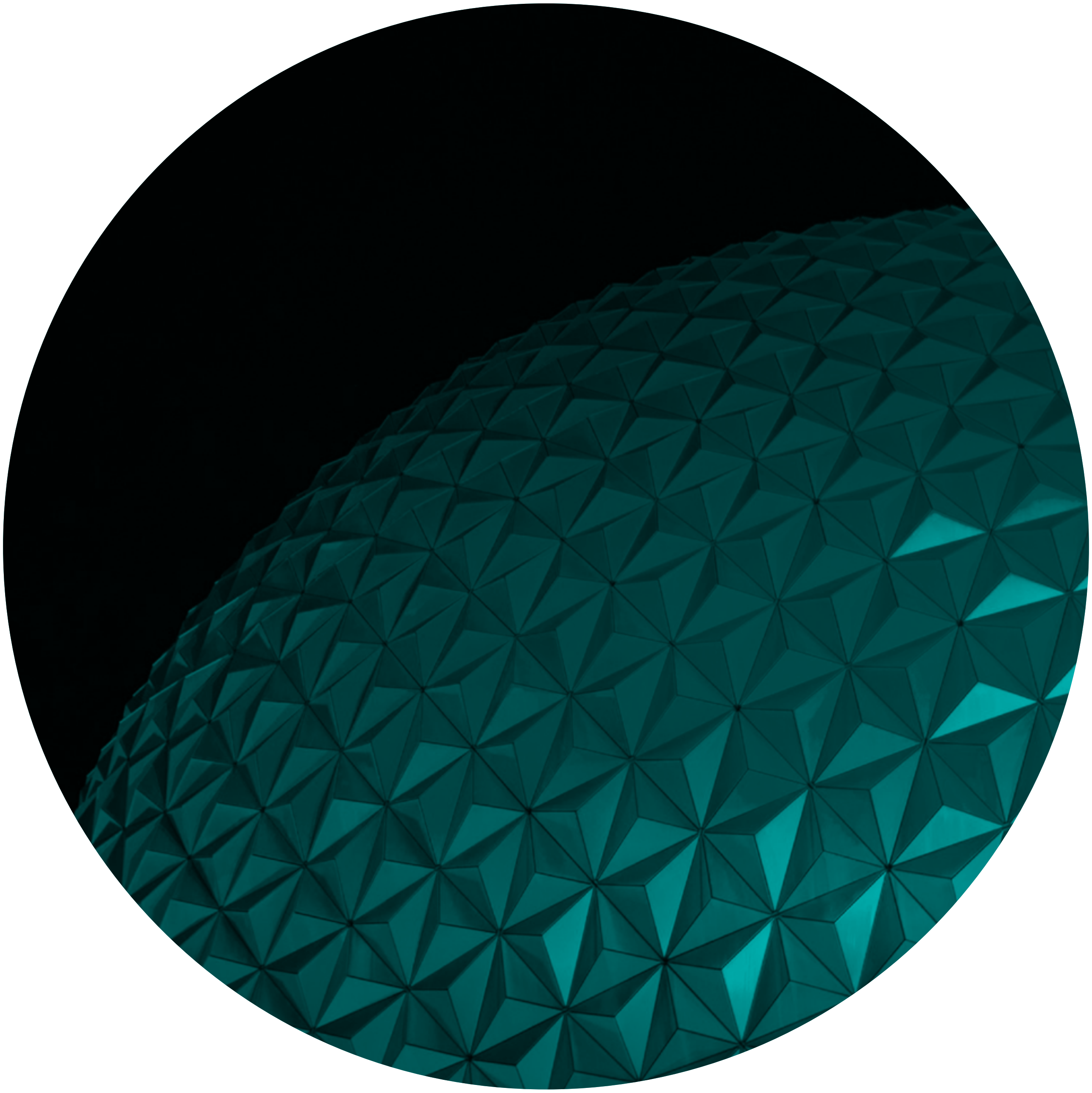
EBOOK

Data Governance: Establishing a Trustworthy Enterprise Data Resource for Innovation and Production



Kevin Lewis

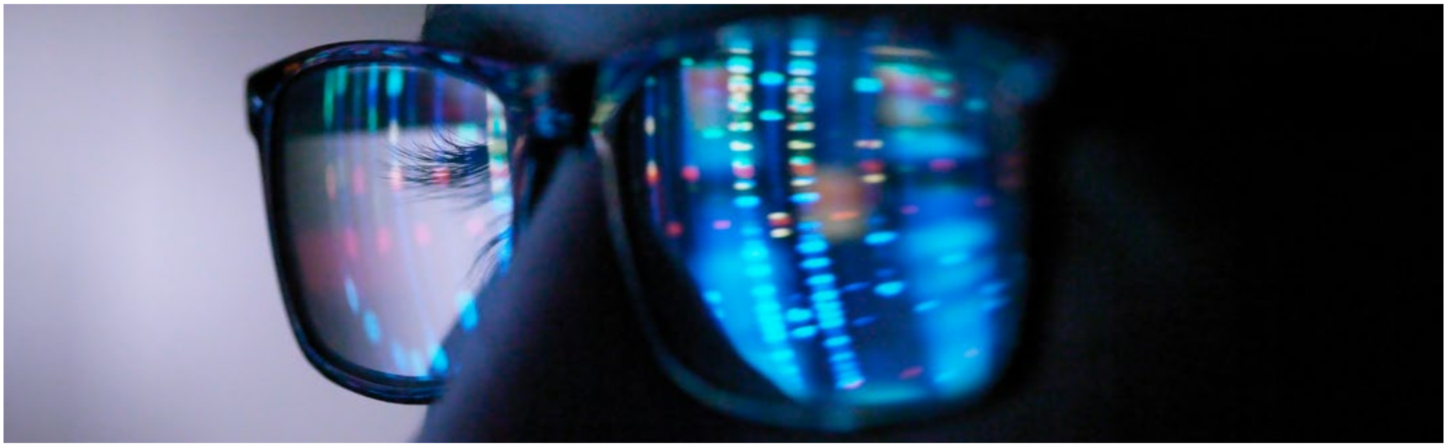
07.21 / DATA GOVERNANCE

Table of Contents

Data governance is often presented as an arcane and complex topic, but at its heart, it's quite simple. The purpose of data governance is to get data in the condition needed to support valuable action. This definition is important for what it says, but also for what it doesn't say. It doesn't say that the purpose of data governance is to create policies and procedures, establish "good" quality data, or implement data management practices. Those things are important, but only to the extent that they help data governance do what it needs to do – support valuable action.

In this ebook, we'll describe how to do that.

First, we'll discuss common pitfalls and the best way to avoid them. Then we'll present a framework for enterprise data management that includes the drivers for data governance and the data management capabilities that data governance uses to support those drivers. Finally, we'll present a simple but effective structure for a comprehensive data and analytics roadmap, and we'll explain how to position data governance activity within the roadmap for maximum effect.



Common Pitfalls of Data Governance Implementations

Unfortunately, most data governance implementations struggle to achieve their goals. And the reason they struggle is because they focus too much on data governance itself and not enough on valuable business action. This mistake manifests in many ways, but generally comes in two patterns.

The first pattern positions data governance merely as a mechanism to provide loose guidance to various people and efforts throughout the organization as they deploy data. The result is what many people inside these organizations self-deprecatingly refer to as the "wild west." Sometimes starting under the name of "data democratization" or a similar well-intended idea, every data scientist and application development team is left on their own to prepare data, manage the quality, link disparate data domains together, secure the data, and perform all manner of data management activities that distract time and attention from producing real value from analytics and applications. These distributed efforts duplicate work by collecting data from the same sources over and over again, using different processes for different requirements, resulting in a bewildering and vulnerable data landscape. This leads to excessive costs, slow projects, and inconsistent and untrustworthy data.

The second pattern often occurs after experiencing frustration with the first. After noticing the proliferation of data and all the issues that go along with it – sometimes in one jarring moment, such as a security breach, or a slow-building realization that data is becoming unmanageable and unreliable for business needs – data governance decides to take control. Here, the data governance organization, under the direction of a chief data officer (CDO) or a similar role, centralizes data deployment so core data can be shared by everyone.

The problem occurs when the program over-centralizes by detaching from identified and named analytic efforts and application development work, believing this to be a necessary step as data is deployed for all possible uses. This mistake is sometimes hard to notice because data governance teams typically do maintain relationships with the end user and application development community, perhaps interviewing many of them, inviting them to data governance meetings, delivering communication to them, and so on. The data governance team may even document a seemingly respectable return on investment.

This all seems reasonable. But if data is not delivered in direct support of targeted, data-hungry application and analytic use cases within business initiatives, then each iteration of data delivery balloons in size as it attempts to include all data elements and solve all data quality problems within the in-scope data domains. Without a strong mechanism for scope control – targeted application and analytic use cases – there is no way to effectively decide what's in and what's out for each data delivery project. Everything goes. This results in projects that take too long, cost too much, and deliver data and capabilities that are severely underutilized. To make matters worse, the first pattern – the "wild west" approach – continues unabated as application and analytic projects work around the centralized data governance function, understandably having neither the patience nor the motivation to wait for the data they need.

The Right Approach that Meets Business Initiatives

Data governance is successful to the extent that it quickly and efficiently ensures the readiness of data to meet the needs of important business initiatives. To make this happen, data governance must find a pragmatic middle ground between simply providing polite advice on one extreme end and taking on excessively large data delivery projects – disconnected from the major business initiatives of the company – on the other extreme end.

Think of other assets and how they are managed in a large enterprise – money, for example. What is the responsibility of the chief financial officer (CFO)? Imagine a finance department that merely offered advice to all other departments on how to manage a general ledger, keep track of accounts payable, balance debt and investment, and so on.

Or, on the other extreme end, imagine the CFO insisting that only employees of the finance department may possess a company credit card and other forms of payment, and all other employees must wait for a finance representative to approve and carry out any purchase of any size. Of course neither of these extremes would make sense. Instead, the CFO takes direct control of core financial processes while also providing guidance to everyone else on their financial rights and responsibilities. To connect the two, the guidance includes policies and systems to support a reliable interface between decentralized transactions and centralized accounting.

That's how the chief data officer (CDO) – and data governance under his or her control – should manage data. The CDO should take direct control of data deployment and management for highly shared, core, enterprise data, delivered in appropriately scoped efforts based on identified business drivers. The CDO should also provide guidance to the organization on how to contribute to and leverage that data responsibly with the freedom to access shared data, experiment with new and innovative datasets, perform analysis, write reports, and build solutions – without the need to wait in line.

Accelerating Self-Service with Data Governance

To support both shared data deployment and distributed innovation, it is often assumed that highly coordinated and proactive enterprise data governance must be balanced with self-service data provisioning and analytics. But this is not the case. A carefully planned and implemented data governance program *accelerates* self-service by providing reliable core data, thus allowing analysts and application teams to focus their efforts on unique work needed for their use cases. This understanding becomes increasingly important as self-service or **"data democratization" is further enabled by friction-reducing cloud technology**.

Effective data governance further enables self-service by offering data access tools, usage guidance, and policies to protect sensitive data, automated wherever possible. In addition to the services offered directly by data governance, data catalogs allow analysts and developers to participate in "crowdsourcing." This empowers the wider community to rate data sets, provide suggestions on use, and share data curation processes, reports, models, and so on.

While a successful enterprise data management framework applies to both shared data and self-provisioned data, different levels of rigor are required for each. For example, data quality for self-provisioning is typically the responsibility of the end user, whereas data quality for shared data must be certified for the production use cases it is intended to support.

This ebook focuses primarily on data governance in its role to support shared, production data deployed to serve the most business critical initiatives while simultaneously enhancing shared data available for innovation, self-service analysis, and solution development.

Data Governance Within a Framework for Enterprise Data Management

Data governance is best understood within two contexts:

- Drivers that motivate action and determine scope (the outer circle on Figure 1).
- Data management activities that contribute value to those drivers (the six small circles surrounding the center of Figure 1).

The data governance/data stewardship ring is positioned between the two, implying that they orchestrate and participate in data management practices to contribute to identified business drivers

Let's review each element of the framework.

Drivers of Data Governance

Business Alignment

Business alignment is frequently cited as the most important factor in the success of data governance, and with good reason. But what does "business alignment" really mean? Does it mean we should include influential members of the business community in the data governance program? Does it mean we should interview a variety of decision makers and analysts, and address their most pressing needs? Does it mean we should make sure the work of data governance is focused on demonstrable business value? Of course the answer is yes to all of these questions, but we need to be more specific.

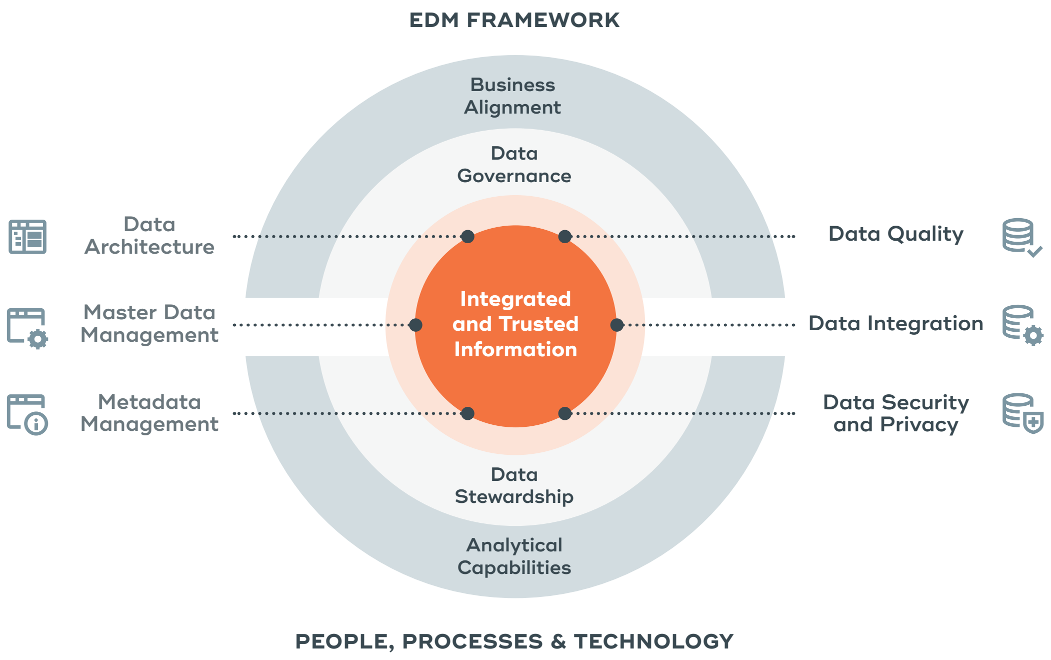


Figure 1. A Framework for Enterprise Data Management

Again, the purpose of data governance is to ensure data is in a condition to support valuable action. But which action? This leads us to what is by far the most important element for the success of data governance: alignment to funded business initiatives that are expected to achieve business value. Virtually every major business initiative – such as supply chain optimization, omni-channel marketing, or enterprise risk management – requires data for success. Delivering that data, just-in-time, just-enough, and in just the right condition to meet the needs of application and analytical use cases within those initiatives provides not only justification for data governance, but also provides a built-in scoping mechanism. Every act of data management – every data element deployed, every data quality issue resolved, every data definition elucidated – should be in service to a near-term need within a funded business initiative.

In addition to aligning with funded company initiatives, data governance should address data issues that affect business operations, such as incorrect commission payments due to misallocated order values or excessive false positives in fraud detection models due to misclassified charges. Addressing data problems like these can be as simple as keeping a list of issues that are directly impacting operations and resolving them in priority order. Here, data governance leaders must be cautious.

Issues that may hinder business initiatives are automatically important because they could disrupt the vetted value of funded efforts that are expected to succeed. Prioritizing these issues is relatively straightforward. However, when dealing with operational issues outside the scope of any funded initiative, data governance must take more direct responsibility to assess and prioritize the issues based on business impact. A good way to foster this prioritization is to establish a fixed and visible budget specifically to address operational issues. Resolution of any issues may of course be proposed for funding on their own, but the general issue resolution budget should typically be used for issues that are too small to justify going through the process of a funding request yet are collectively worth addressing.

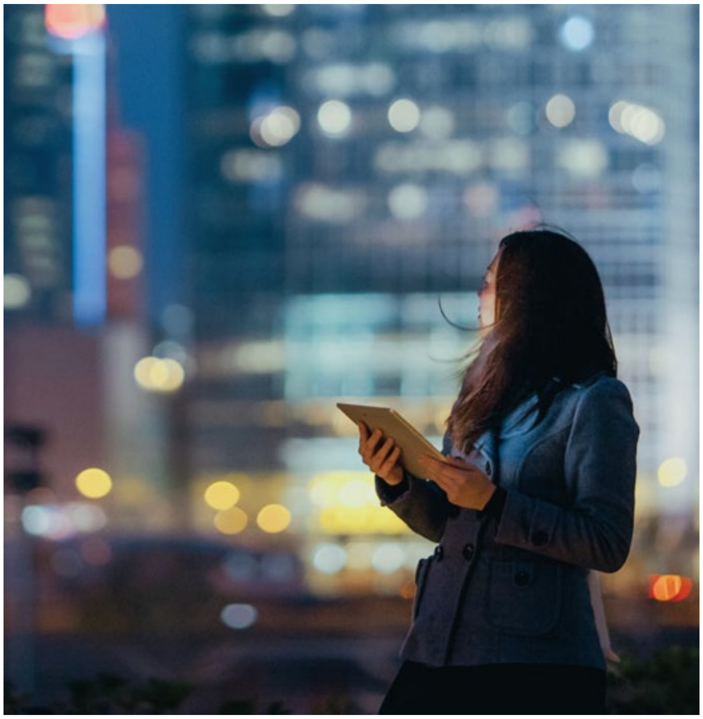
Dealing with operational issues is important, but it must not take the place of or distract from the essential strategy, which is supporting funding business initiatives.

Analytic Capabilities

Data by itself has no value. It only has value when it's used to provoke meaningful action. Making data actionable requires the help of analytic capabilities including skills, methods, and tools to support human-directed analytics and analytic functions embedded within applications.

Analytic capabilities today suffer from the same issues as data governance, although in the case of analytics, the "wild west" approach is much more common than over-centralization. Because of this, many organizations struggle to provide even the most basic and widely needed reporting reliably and consistently.

Much like data governance, to remedy this situation, analytic capabilities should be enabled through the middle ground between centralized and distributed responsibilities. A central analytics team, led by a chief analytics officer or a combined chief data and



analytics officer role, should provide expertise in the use of analytic methods and tools, and facilitate planning and reuse of analytic and reporting objects. The central team then has a dotted or straight-line relationship to analysts throughout the organization who directly serve the needs of each business area.

IT is responsible for providing a range of tools for analysts and for supporting application development teams with embedded analytic capabilities for production applications.

Understanding the needs and challenges of the analyst and developer communities provides useful input for the priorities of data governance. They may find some data challenging to navigate, or they may encounter quality issues. They may also identify opportunities to provide value through new and innovative sources of data. But here again, priority must be given to requirements that can be traced to how they will benefit targeted business initiatives.

In addition to supporting funded business initiatives with the shared data they require, guidance provided by data governance should allow for highly decentralized self-service. The self-service capability should have easy access to authorized analytic and data resources. It should also enable self-service data provisioning for experimentation, ad-hoc analysis, and prototype development – without letting “self-service” become just a cover for the “wild west” pitfall.

Data Management Capabilities and the Role of Data Governance, Data Stewardship, and IT

We’ve established that data governance works best when it supports business initiatives and operations, and provides the underlying data for the human and machine analytics needed for the success of those initiatives. But what does data governance *do* to support business initiatives? What does it mean to ensure the data is “ready” to support the initiatives?

At the executive level, data governance, working with its cross-functional constituency, provides the political will, funding, and coordination to deliver data to meet immediate business needs while simultaneously contributing to coherent, shared, data resources. It’s through the execution arm – data stewardship – that data content experts from the business cooperate with technology experts from IT to deploy data in a way that will meet in-scope business needs.

In the following sections, we’ll briefly describe each data management practice and discuss the role of data governance, data stewardship, and IT within each practice.

Data Quality Management

Data quality management is about making sure that data is “fit for purpose.” Data quality can never be perfect across the enterprise; therefore, the point of data quality management is not simply to make sure that data is “good,” but that it is in the condition necessary to enable business initiatives and business operations. Incorrect inventory balances are a problem to the extent that they, for example, affect the ability to automatically replenish inventory to cover predicted sales.

The two most important practices within data quality management are data profiling and data quality monitoring.

Data profiling casts a wide net across data and fishes out any quality issues that may exist in the data, looking specifically for those issues that could affect in-scope use cases. Data profiling is typically performed within data delivery projects and periodically as needed for investigation of in-scope issues.

Data quality monitoring is more systematic. It typically runs in production on a regular basis, evaluating known data quality concerns and providing quality metrics and details to people and systems that can improve quality. While data profiling looks for data issues, data quality monitoring keeps an eye on issues that require ongoing attention in production.

Returning to the automated inventory replenishment example, data profiling determines that there’s a data quality problem with inventory balances that could affect an application project. Data quality monitoring uses statistical and other rule-based techniques to alert the right people if the problem crops up in production.

The role of data governance,in its oversight function, for data quality management is to:

- Ensure that data quality is focused on in-scope business initiatives and operational issues
- Prevent data quality from become an unfocused "fix all the data" program

The role of data stewardship, in its hands-on, day-to-day, and project-based function for data quality management, is to:

- Ensure the assigned data domain has the quality needed for specific, targeted application and analytic use cases
- Interpret data profiling results by finding any issues that may affect in-scope use cases
- Help determine technical and non-technical root causes of relevant data quality issues
- Help identify an appropriate fix for the identified issues and propose system modifications, business process changes, training, etc.
- Identify data quality rules to be implemented in production for ongoing monitoring of identified issues

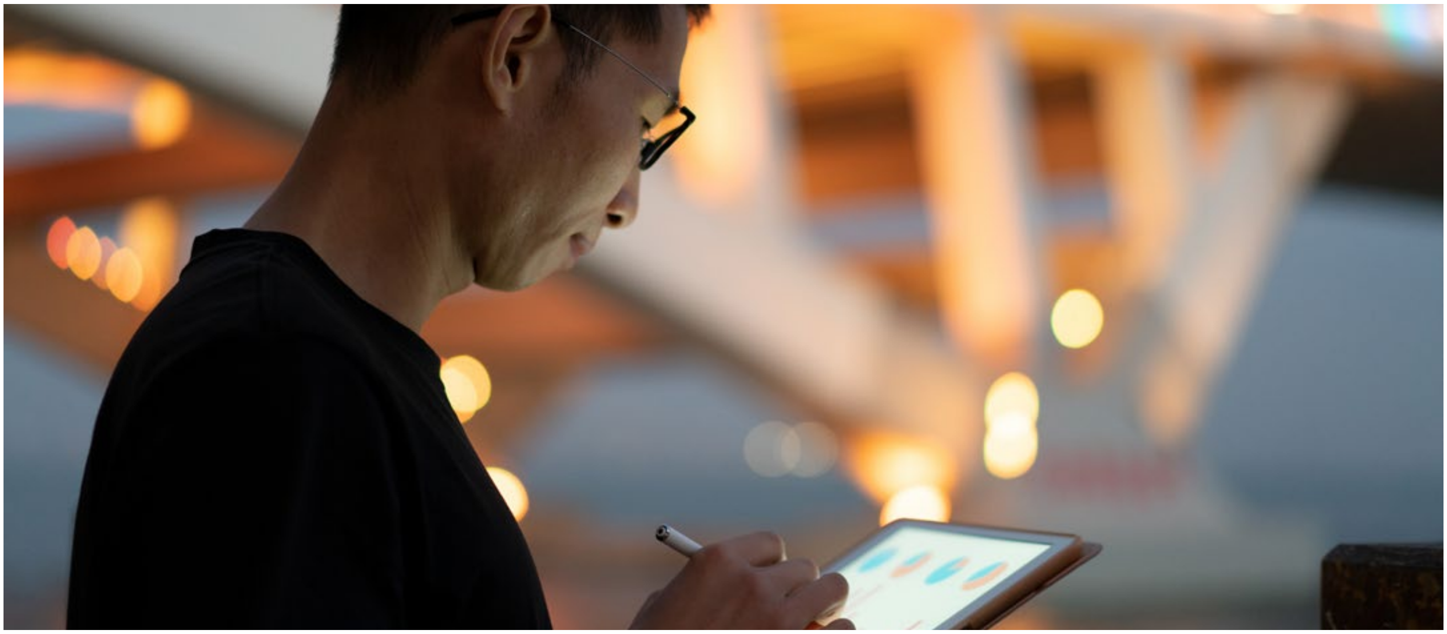
The role of IT for data quality management is to:

- Implement data profiling and data quality monitoring tools
- Run data profiling processes to assist data stewards in looking for data quality issues that could affect targeted use cases
- Translate data quality rules into production processes to monitor selected data quality issues on an ongoing basis

Data Integration

Enterprise data lives in many different data stores using a variety of technologies, data structures, and naming conventions. To make use of data across data domains, it's necessary to integrate the data, whether it's physically integrated into one data store or logically integrated by creating easily navigable links, or most commonly, a combination of both.

For example, for a pharmaceutical company to fully understand the value of a prescription medication, data must be collected regarding the chemical components,



prescription rates, clinical efficacy, adverse reactions, patient characteristics, and other elements. Because these data domains are typically stored in different systems, and even within different institutions, processes must be established to acquire and connect the data for easy access by applications and analysts.

Every enterprise has several core data domains that are needed repeatedly by a variety of otherwise disparate application and analytic use cases. Therefore, integrating data is not just important for individual use cases that require data from several sources, it's also important for positioning the data to be reused and extended as needed. This avoids unnecessary and redundant work to integrate the same data repeatedly.

The role of data governance for data integration is to:

- Ensure that data integration is focused on integrating data just-in-time and just-enough for targeted use cases
- Promote the deployment of core, enterprise data so it's easily reusable and extensible rather than splintered and deployed redundantly for various projects

The role of data stewardship for data integration is to:

- Help determine the authoritative sources of data to be integration for in-scope targets
- Help develop rules to translate data into a form that is usable by multiple applications and end users
- Promote consistency of data elements so data can be semantically linked across domains

The role of IT for data integration is to:

- Implement data integration tools, many of which have similar or overlapping functionality such as extract, transform, and load (ETL), data engineering, data preparation, data workflow, and data pipeline tools
- Map data elements from data sources to structures within shared repositories and implement transformation rules and logic
- Elicit requirements from applications and analytic use cases to meet immediate needs, limit integration of each project to needed elements, and position processes to be extended for additional elements as needed for future use cases

Data Security and Privacy

Data should of course be accessible only to people and systems authorized to use that data. Failure to adequately protect data can result in fines, damaged reputation, or weakened competitive position. Several state, national, and international regulations require organizations to secure financial, personal, and other data, but these regulations usually establish only general guidelines.

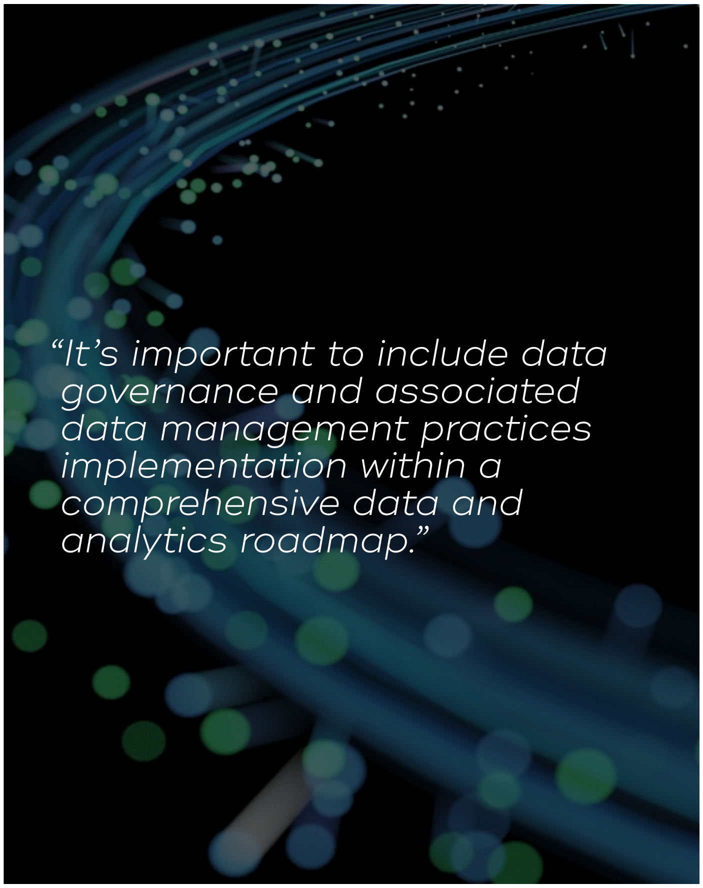
Each company must institute policies to direct how regulations are to be implemented and how data is to be protected, even beyond regulatory requirements. For example, although the European Union's Global Data Protection Requirement (GDPR) applies only to European citizens, organizations must decide to what degree they will offer the same protections to everyone, considering the cost of implementation or segregating treatment based on nationality, public expectations, and potential risk of reputation. These are business decisions, not legal decisions.

An organization's legal department can only interpret regulations and make recommendations. The information security team likewise interprets regulations to recommend specific procedures and technical mechanisms for compliance. But that's not enough. Both functions must work in partnership with business executives – that is, data owners – across the organization to make policy and compliance decisions for specific data domains, considering risk and cost.

As an important role within data governance, data owners assigned to one or more data domains – such as customer data, employee data, financial data, and so on – must ultimately be responsible for these decisions. With these decisions in place, data stewards, working with IT, implement the policy choices in automated solutions and semi-manual request processes.

The role of data governance for data security and privacy is to:

- Establish data owners for sensitive data domains (and extend the role to participate in other data management practices within assigned domains if these responsibilities are not in place already)
- Work with information security and the legal department to interpret regulations and develop internal policies, codifying how the company intends to implement the regulations
- Work with information security to develop a data classification scheme (e.g., classifications such as internal only, confidential, personally identifiable, and public) and the technical and non-technical expectations for each level
- Establish company policy for security and privacy of assigned data domains beyond regulatory compliance



"It's important to include data governance and associated data management practices implementation within a comprehensive data and analytics roadmap."

The role of data stewardship for data security and privacy is to:

- Propose the assignment of classification levels to data domains and elements, independent of systems
- Establish and document roles to be authorized for automatic approval and access to data domains and elements
- Work with information security and other areas to develop procedures for special access requests

The role of IT for data security and privacy is to:

- Work with information security to establish the technical implications for each classification level
- Implement technical capabilities required for security (e.g., encryption, single sign-on)
- Test security capabilities and regularly probe for vulnerabilities

Metadata Management

The most common definition of metadata describes it as "data about data." While this is true, there is no stark line of distinction separating data from metadata. For example, in photography, information that describes a photograph – such as the date it was taken, location, camera settings, and so on – are referred to as "metadata." But an analyst writing a report documenting the inventory of photographs within a museum or a periodical may legitimately refer to such attributes as "data."

In any case, metadata helps end users, application developers, and systems understand the content of data. Metadata falls into three somewhat overlapping categories:

1. **Business metadata** describes data in business terms, including any definitions, formulas used, and synonyms. Business metadata typically requires a human to document; hence it requires discipline. Modern data catalogs offer "democratic" access to data and associated metadata. This allows anyone to update or suggest definitions, rate the quality of sources, and access associated reports and queries when leveraging and evaluating data
2. **Technical metadata** describes the system-level characteristics of the data such as data type, size, physical display characteristics, and statistics describing the data for use by database optimizers. The data dictionary typically included within a database management system (DBMS) is a type of repository for technical metadata. All data management tools leverage technical metadata in one form or another.
3. **Operational metadata** includes any descriptive information on processes that affect the data. For example, information about a data pipeline that updates data, including general information about the process along with information about process execution, such as date, runtimes, error rates, etc., are considered operational metadata.

The role of data governance for metadata management is to:

- Determine the focus for metadata management efforts (e.g., business, technical, operational, or a combination) and the expected value
- Ensure the focus of the implementation directly supports the approved value for targeted use cases and constituents (e.g., impact analysis for IT, analytic support for end users, data navigation for application developers)

The role of data stewardship for metadata management is to:

- Help determine the role of metadata in near-term application and analytic use cases
- Help determine appropriate sources of metadata and the priority of elements within each source to support use cases
- Document business definitions and respond to recommended updates and enhancements from the wider community
- Help establish communication vehicles and mechanisms (e.g., training, data catalog) to educate end users on available data and support collaboration and learning

Master Data Management

Data can generally be divided into two categories. The first is transaction data. This includes data derived from accumulated transactions like snapshots such as inventory levels and account balances. The second category is master data. Master data describes real-world entities and concepts such as suppliers, retail locations, equipment, customers, employees, and so on. Think of transaction data as data involving numbers and math applied to those numbers, while master data provides ways to organize the numbers; for example, sales (transaction data) by customer (master data).

Master data requires special attention for several reasons:

- The identical real-world entity can be represented in multiple source systems, such as one person with a bank account, mortgage, and multiple credit ratings, all managed by separate applications
- Master data entities are typically organized in hierarchies, such as employees within job roles, departments, and business units
- Master data entities can change their attributes over time, such as a customer changing marital status from single to married

The role of data governance for master data management is to:

- Ensure ownership of master data coordinated across business areas that share some responsibility for master data domains
- Ensure that master data initiatives, including domain-specific initiatives such as "Customer 360" or "Customer Data Platform," are aligned to business initiatives that will leverage the data, thus driving the scope of master data delivery and reconciliation
- Facilitate resolution of business consistency issues that cannot be resolved by master data management directly; for example, if separate departments within a service organization offer similar but inconsistent services, no master data management process or tool can fully reconcile this overlap

The role of data stewardship for master data management is to:

- Work with IT to develop roles and workflows for assigned master data domains (additions, updates, deletions, hierarchy management) across business areas, such as marketing and logistics, with each contributing information regarding different attributes of the same product
- Determine the most authoritative sources for specific data elements within each master data domain, especially when the same attributes are stored in multiple systems with inconsistent values
- Help determine whether master data for assigned data domains should be rationalized at the source (source of record) or rationalized by cross referencing disparate sources in a shard data resource (source of reference)
- For the source of reference approach, maintain cross reference links to correlate original sources that remain un-reconciled

The role of IT for master data management is to:

- Implement master data management tools and link them to other relevant data management capabilities, such as data integration; for example, to leverage master data cross-references in a data pipeline
- Implement workflows within tools to support master data management additions, changes, and deletions, along with management of master data hierarchies
- Advise data stewards and data owners on the pros and cons of a source of record versus source of reference strategy, considering targeted use cases (Analytic use cases tend to benefit from a source of reference approach, whereas transactional business processes tend to require a deeper consolidation at the sources)
- Identify and develop or acquire mechanisms to manage reference data, which is a subset of master data dealing with simple codes and descriptions, that does not have adequate controls or operational applications for management of the data

Data Architecture

Data architecture is an expansive topic, but in summary, it's about organizing data so it can be effectively captured, linked, and accessed. For example, data architecture determines the kinds of shared data resources that will be needed (e.g., data lakes, data warehouses, data mesh domains) and the interrelations of these resources along with guidance on the structures within them, such as internal data modeling approaches and external access mechanisms.

The enterprise data architecture function, working within the broader enterprise architecture team, is typically responsible for establishing a vision and plan for shared data resources along with the structures and processes to organize data within and among the resources.

The role of data governance for data architecture is to:

- Understand the role of major data repositories, access mechanisms, and in partnership with enterprise architecture, work across departments to ensure they are used when appropriate, especially for highly reusable data
- Working with enterprise architecture, approve exceptions for specific applications and use cases when warranted

The role of data stewardship for data architecture is to:

- Participate in sessions to define the structure and relationships of data from a business perspective (i.e. logical data modeling)
- Help control scope of data requirements, including data modeling, by considering the business needs of in-scope use cases

The role of IT for data architecture is to:

- Implement tools to support data architecture, including enterprise modeling tools and logical and physical data modeling/database design tools
- Establish architectural vision and roles for shared data resources, including centralized and distributed resources and the relationships among them
- Establish standards and guidelines within the various data layers, such as data modeling and access mechanisms and approaches
- Facilitate logical data modeling sessions to uncover business needs and translate models into optimal physical structures for extensibility, scalability, and performance
- Establish or acquire enterprise modeling frameworks to be used as scaffolding to be detailed just-in-time and just-enough to support in-scope use cases

Planning and Implementing Data Governance as Part of a Comprehensive Data and Analytics Roadmap

To ensure that data governance and the associated data management capabilities are positioned to add strategic value, and not just to "mature" capabilities, it's important to include data governance and associated data management practices implementation within a comprehensive data and analytics roadmap. A comprehensive roadmap contains five tracks, as shown in Figure 2.

Business Initiatives. This track depicts the business initiatives that have been selected as the drivers

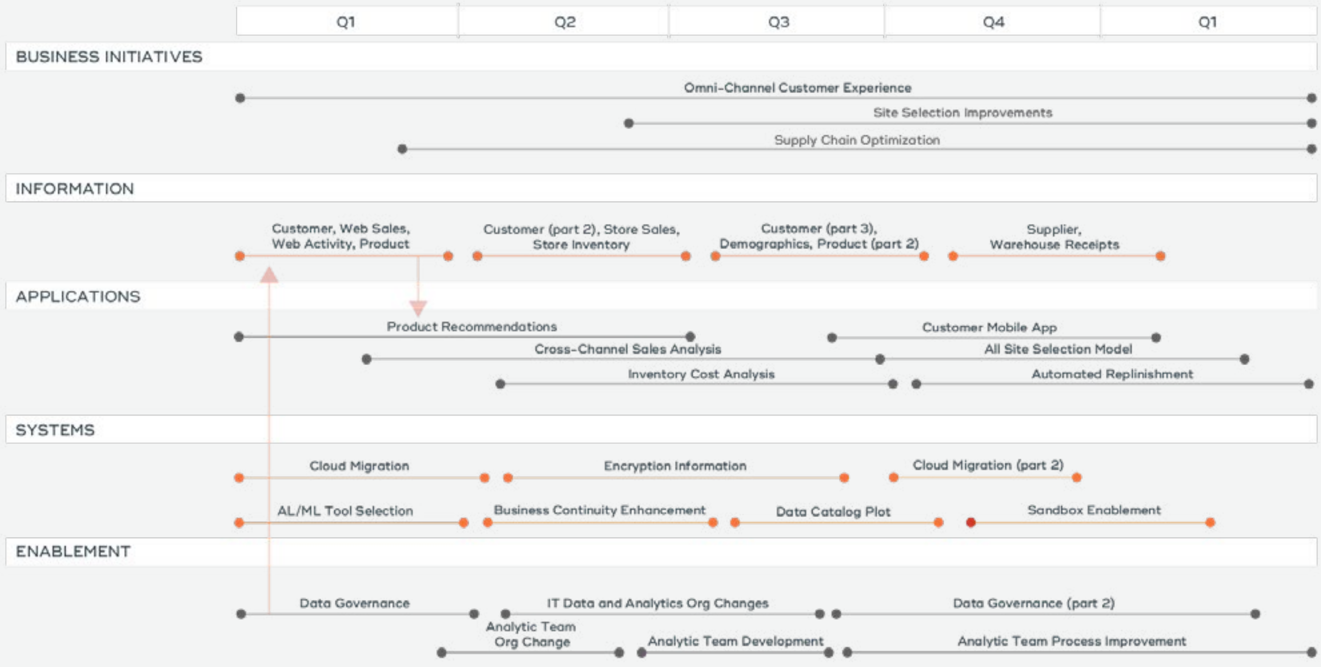
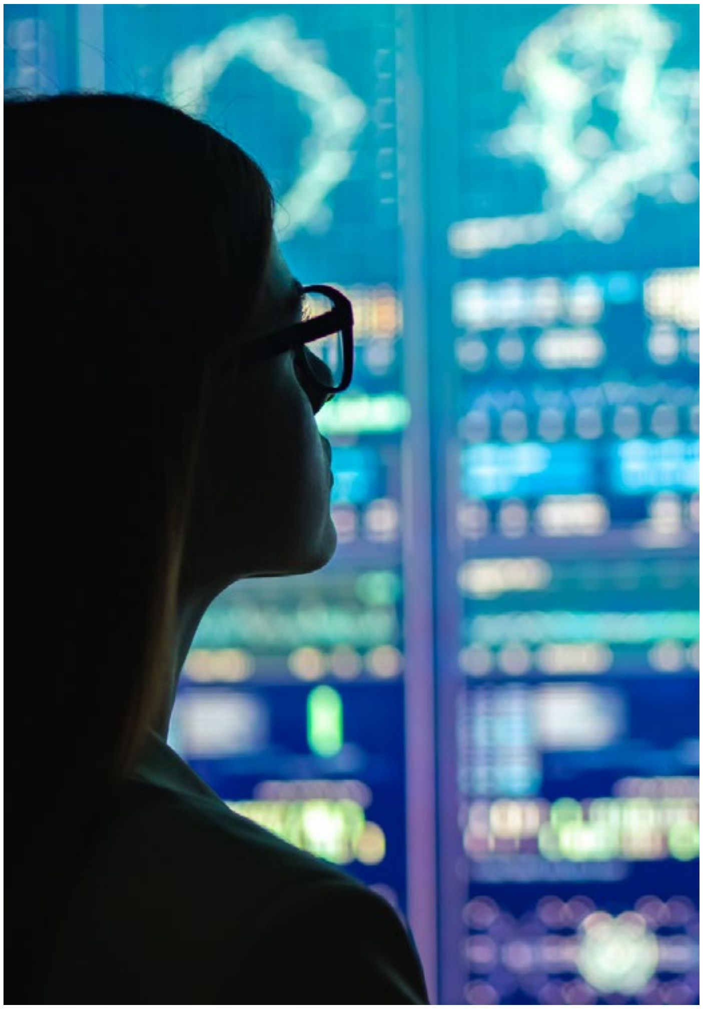

Q1
Q2
Q3
Q4
Q1
BUSINESS INITIATIVES
Omni-Channel Customer Experience
Site Selection Improvements
Supply Chain Optimization
INFORMATION
Customer, Web Sales, Web Activity, Product
Customer (part 2), Store Sales, Store Inventory
Customer (part 3), Demographics, Product (part 2)
Supplier, Warehouse Receipts
APPLICATIONS
Product Recommendations
Customer Mobile App
Cross-Channel Sales Analysis
All Site Selection Model
Inventory Cost Analysis
Automated Replinishment
SYSTEMS
Cloud Migration
Encryption Information
Cloud Migration (part 2)
AL/ML Tool Selection
Business Continuity Enhancement
Data Catalog Plot
Sandbox Enablement
ENABLEMENT
Data Governance
IT Data and Analytics Org Changes
Data Governance (part 2)
Analytic Team Org Change
Analytic Team Development
Analytic Team Process Improvement

Figure 2. Data and Analytics Roadmap. Effective data governance is implemented just-in-time and just-enough to ready in-scope data for in-scope applications and analytic use cases.

for the rest of the roadmap. These initiatives are typically sponsored outside of the data and analytics organization. The CDO or a similar role should sponsor data initiatives that support the initiatives of other CXOs, not compete with them for resources.

Here we are looking for business initiatives that have been vetted for value so we can propose how to contribute to that value.

Positioning data governance as a "load bearing wall" underneath vetted strategic initiatives provides a much needed sense of urgency to data governance. This not only justifies the existence of the program but also sharpens the scope so every action is prioritized based on the near-term impact it will have on the applications and analytic use cases within selected initiatives.



Information. This track depicts the iterative rollout of shared data resources in direct support of selected business initiatives. These efforts should be sponsored by the CDO, who should clearly articulate exactly how these projects do two things:

1. Deliver the data needed for supported business initiatives – just-enough, just-in-time, and in just the right condition
2. Contribute to coherent, shared enterprise data

Application and Analytic Use Cases. This track depicts the projects that build solutions to directly leverage data for business value. These should include application projects that are already within the scope of identified business initiatives. Again, these applications should typically be sponsored outside of the data and analytics program. If you carefully examine the plans for just about any business initiative, you'll usually find plans for data-intensive applications, either packaged or custom, such as recommendation engines, clinical diagnostic tools, or social media sentiment analyzers within customer service applications. In addition to applications, analytic use cases are often planned within initiatives. However, very often, easily predictable reporting and analytic needs are given short shrift in the budgeting process. If these gaps are found, the CDO should advise other CXOs on the missing pieces so they are planned appropriately.

Once identified, projects in this track provide a scoping mechanism for the other tracks. Each data element delivered in the information track should have an identified need in the Application track. If this dependency relationship is not in place, either data is being deployed without a good rationale or an application project is proliferating data by sourcing it without help from the data and analytics organization. Either situation must be remedied by linking data delivery plans to application or analytic use cases.

Systems/Infrastructure. This track depicts the work needed to provision infrastructure, whether in the cloud, on-premises, on the edge, or some combination thereof. Projects in this track deploy compute capability, storage mechanisms, backup and recovery, business continuity, and so on. As with the other tracks on the roadmap, the idea is to deploy only what you need, only when you need it – in direct support of business initiatives and the applications and data required for their success.

Enabling Capabilities. Finally, this is the track where data governance and associated data management practices live, along with other organizational and process capabilities. Notice the first project in this track is generically labelled "Data Governance." Although this is vague, its position on the roadmap indicates the real focus. One purpose, for example, is to ensure that customer data deployed in the information track is "fit for purpose" to support the Product Recommendations project in the Application track. Notice how this dependency relationship adds a sense of urgency and purpose to data governance

If we were to dig deeper into the scope of the data governance project, we'd find small elements of some or all data management practices being leveraged and matured as needed to make sure the data is ready for the targeted business use. For example, the scope of this project could include:

- Assigning data stewards to the data domains within the first Information track project (customer data, etc.)
- Profiling customer data to identify data quality issues that could hinder success of the Product Recommendations project
- Implementing or leveraging a master data management tool and process to reconcile overlapping sources of customer data – to the extent necessary for the Product Recommendations project
- Integrating transaction data from sales channels that will contribute to and leverage the Product Recommendations application
- Establishing or leveraging a data catalog annotating the data needed to monitor the impact of the Project Recommendations application
- Examining and interpreting regulations that will become important due to new uses of customer data in the Product Recommendations engine – such as documenting the use of machine learning models as specified in the European Union's GDPR rules

In this way, the implementation of data governance and every participant in the program align to actions that directly support the most important initiatives of the enterprise. Data governance thus becomes strategic and important, by definition, rather than the proverbial solution in search of a problem.

While deciding on actions to improve the organization, it's important to think about sustainment. The best way to sustain – and continuously improve – the program is to link data governance and other elements of data strategy enablement to the larger operating model.

For example, you should:

- Ensure the CDO is included in the company strategic planning process to offer ideas on how to leverage data and initiate data initiatives in direct support of strategic business initiatives as they emerge
- Ensure that the enterprise architecture function has an enterprise data architecture capability that examines not only data and analytics technologies that are leveraged within projects, but also promotes the appropriate use and expansion of shared data resources and data content
- Link data management activities and roles directly into the solution delivery methodology, especially for projects within the information track of the roadmap

These elements become part of the machinery of the organization, keeping the roadmap up to date, and continuously aligning data governance with the initiatives relying upon it for trustworthy data.

About the Author

Kevin Lewis is a Director of Data and Architecture Strategy with Teradata. Kevin shares best practices across all major industries, helping clients transform and modernize data and analytics programs including organization, process, and architecture. These best practices advocate for strategies that deliver value quickly while simultaneously contributing to a coherent ecosystem with every project.

About Teradata

Teradata is the connected multi-cloud data platform for enterprise analytics, solving data challenges from start to scale. Only Teradata gives you the flexibility to handle the massive and mixed data workloads of the future, today. Our open approach embraces the modern ecosystem to create a seamless experience for ingestion, exploration, development, and operationalization. Teradata's experts and partners around the world can show you how to drive business outcomes and unlock unlimited value by turning data into your greatest asset. Learn more at **Teradata.com**.

17095 Via Del Campo, San Diego, CA 92127   Teradata.com

The Teradata logo is a trademark, and Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

© 2021 Teradata Corporation   All Rights Reserved.   Produced in U.S.A.   07.21