



**Modern Data Governance:**  
Strategies for Data Policies that Stick

By David Loshin and Abie Reifer

January 2018

## About the Author

---

**David Loshin** is a globally recognized thought leader and expert consultant in the areas of analytics, big data, information utilization, and business intelligence, and author of over 50 white papers and 9 books on information management and high performance computing.

**Abie Reifer** has over 25 years of experience in technology strategy, steering global organizational decisions and adoption of emerging technology in areas such as telecommunications, mobile applications, Health IT, customer-centricity, and government agency consulting.

## About Eckerson Group

---

Eckerson Group is a research and consulting firm that helps business and analytics leaders use data and technology to drive better insights and actions. Through its reports and advisory services, the firm helps companies maximize their investments in data and analytics. Its researchers and consultants each have more than 25 years of experience in the field and are uniquely qualified to help business and technical leaders succeed with business intelligence, analytics, data management, data governance, performance management, and data science.



## About This Report

---

The research for this report is made possible by [Collibra](#).

## Executive Summary

---

*Recognizing data governance as a priority, more organizations are hiring chief data officers and forming data governance councils to define data policies to promote data quality and usability. But implementing data governance can be complex. Programs that focus only on organizational structure and operating models or, conversely, attempt to survey data structure and semantics from the bottom up may need additional support to develop the socialization, championship, and resources required to ensure sustainability.*

*This report lays a foundation for a modernized approach to data governance that is business-focused. We first examine the types of data policies that must be considered within the scope of a data governance program. This is followed by an analysis of the weaknesses of the conventional approach. We then consider the specification of data policies, standards, and rules and how those become the focus of a “data governance reboot.” We also offer guidance on maintaining a business focus as the key motivating factor for defining data policies and standards that “stick.” Finally, the report summarizes the technologies that support successful data governance programs.*

### **Covered in this report:**

- Objectives of a data governance program
- Challenges of the conventional approach to data governance
- Data policies, standards, and rules
- A business-oriented approach to operationalization
- The technologies that support the business-oriented approach

## Data Governance Ascending

---

Today's organizational data environments are complex. Organizational teams must manage numerous systems and platforms for transaction processing, operational processing, business intelligence, reporting, and analytics. Some organizations cope with this growing complexity by implementing practices to organize, manage, govern, and facilitate their use of data assets across business function and technological boundaries.

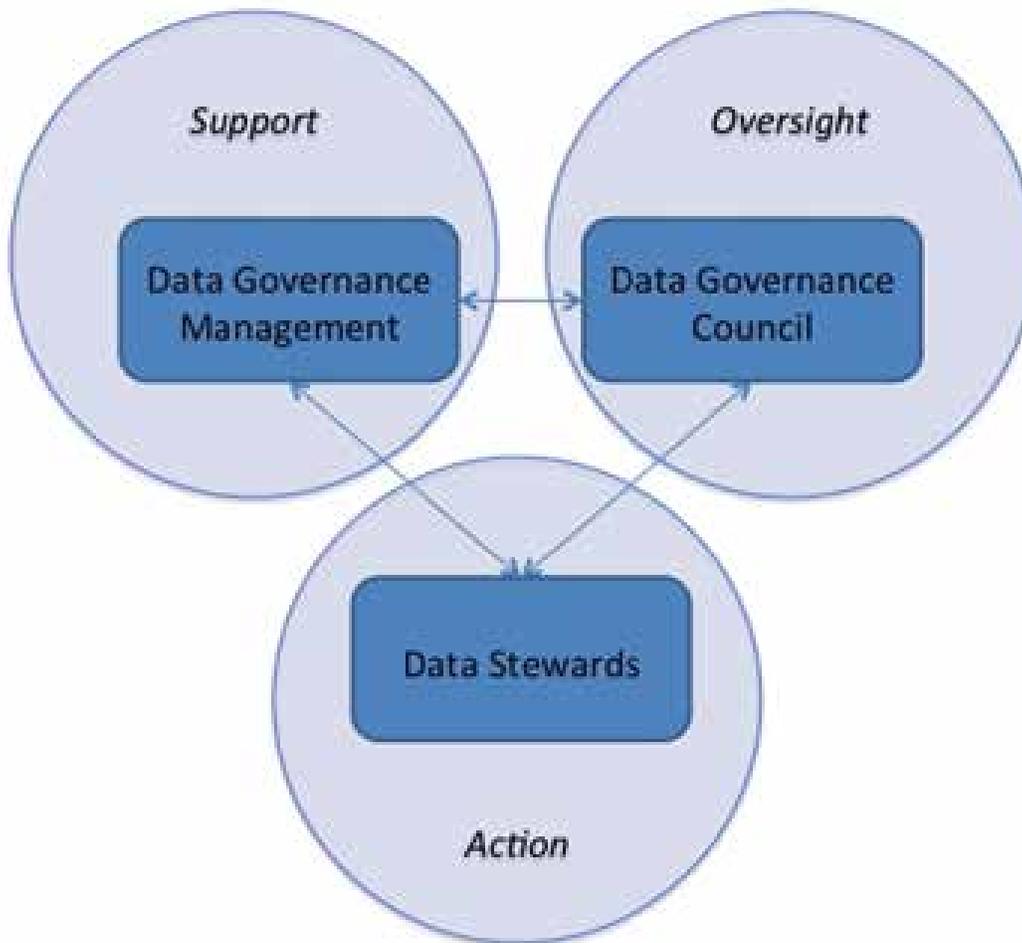
Data governance is a key part of this strategy: specifying policies overseeing corporate data; instituting processes and procedures guaranteeing that data policies, standards, and business rules are observed; and operationalizing those processes in a trustworthy and auditable manner. Astute corporate stakeholders prioritize data governance by hiring chief data officers (CDOs) to establish the corporate data governance program, acquire the right technologies to support it, and hire the right team to put data governance into production.

### Objectives of a Data Governance Program

What does data governance entail, and how does an organized program go beyond traditional data management processes and procedures? An organization's data management activities (such as archiving, backup and recovery, system security, and data disposition) may already adhere to established policies. However, a data governance program is meant to establish an ongoing practice to coordinate the definition of policies and standards and also enforce their compliance. A data governance program encompasses the organizational structure (such as a data governance council and affiliated data stewards and analysts), an operating model within which those parties work together, and defined procedures that guide their activities. Overall, the objectives of instituting a data governance program are threefold:

1. To define and agree to policies, standards, and rules that govern all aspects of the data lifecycle;
2. To develop the procedures for operationalizing compliance with policies and standards and put those procedures into production; and
3. To continuously monitor compliance and take action when policies are not observed.

**Figure 1. The operating model guides the activities in a data governance program.**



## Approaches to Implementing a Data Governance Program

It is a daunting task to implement a data governance program in a methodical and disciplined way with the right level of detail. The complexity grows for organizations that maintain and process large amounts of data. Organizations desiring to oversee data management and tackle their data issues tend to follow one of two possible approaches to implement a data governance program:

**Top-Down.** Recognizing the need for enterprise-wide change management, many businesses take an organizational approach: Hire a CDO, create a data governance council, schedule meetings, and task a team of data stewards to enforce agreed-to data policies.

**Bottom-Up.** Conversely, some businesses employ automated data profiling tools to survey existing databases and collect structural metadata, which become the foundation for standards and rules.

## Challenges in Traditional Data Governance

Both approaches present challenges. The top-down approach is often seen as bureaucratic and restrictive, imposing additional work on staff members without necessarily demonstrating tangible value. The largely mechanized bottom-up approach produces data element summaries, but data professionals still must invest effort to analyze these summaries to isolate and document the discovered standards and rules.

Furthermore, some data governance programs establish processes and procedures that overwhelm an organization which has the effect of undermining the program's long-term effectiveness. And managers specify tactical milestones that reflect program implementation, but struggle to define program targets that are tied to specific business requirements and measurable business goals and objectives.

Without measurable business goals and objectives, those involved in the process often don't fully comprehend the issues that they need to deal with. When this happens, data stewards cannot properly prioritize issues, solutions are delayed, and the absence of business metrics makes it difficult to detect any measurable positive impact of data governance on the business.

## Data Governance Needs a Reboot

Data stakeholders need to understand two key things:

1. The connection between business objectives and data usability, and
2. How to create and operationalize data policies that yield quantified value to business executives.

**In this report.** We lay a foundation for a modernized approach to data governance that is business-focused. We will first examine the types of data policies that must be considered within the scope of a data governance program. An analysis of the weaknesses of the conventional approach follows. Then we consider the specification of data policies, standards, and rules and how those become the focus of a "data governance reboot." And we offer guidance on retaining a business focus as the key motivator in creating data policies and standards that "stick." Finally, the report summarizes the types of technologies that support successful data governance programs.

# Data Policies: A Primer

---

***Data Policy: A documented set of guidelines for ensuring the proper handling and management of an organization's digital assets.<sup>1</sup>***

Expectations are high at the launch of the data governance program. Data governance council members and the data stewards (the individuals tasked with overseeing operational compliance as well as addressing data issues) take on their roles with enthusiasm. But enthusiasm eventually turns to frustration as the data governance council members and stewards struggle to define relevant data policies that are easily put into practice.

There is a need for practical data policies whose compliance can be measured and monitored so that remedial actions can be taken. Yet the conventional approach focuses more on *the process for policy definition and approval than on the content and practicality of compliance.*

Consider the following guidelines to develop good data policies:

**Associate data policies with their business context.** This will amplify the perception of their importance, thereby motivating compliance.

**Align data policies with an organized measurement framework.** Data governance teams should define metrics for each data policy, so the team can measure its impact on business strategies or other desired outcomes. A measurement framework provides the procedures and tools to assess and quantify the degree to which the data sets comply with defined standards. This will simplify the data stewards' efforts to harmonize implementation, monitoring, and remediation.

**Identify the scope of data policies.** Scope defines both the organizational and technical context of a data policy. Organizationally, consider whether each data policy applies to the entire organization, a business unit, or members of a particular workgroup. Technically, determine whether it applies to an entire data model, specific tables, or individual data elements.

## Associate Data Policies with Business Context Objectives

Data governance practitioners must understand and internalize fundamental business drivers if they want to gain executive level sponsorship. Articulating the relationship between high quality data and achieving business goals helps executives understand the importance of data governance and promote the program throughout the enterprise. How can a data governance team link what it does to business goals and objectives?

<sup>1</sup>Adapted from <http://searchcompliance.techtarget.com/definition/Data-governance-policy>

Here are a few ways:

**Regulatory compliance:** Data governance professionals may not completely understand the broad and complex data implications of governmental regulations. There are two facets to the data dependencies associated with regulatory compliance. The first is translating the wording in the regulation into data definitions, data standards, and business rules that must be applied to the data. This straightforward process involves finding critical passages in the text and mapping them to logical and physical data elements used by business applications. The second is instituting methods to augment developed applications with controls that monitor and measure conformance to the business rules to demonstrate compliance.

**Conformance to industry standards:** Companies across industries proactively form cooperative groups to specify standards and document generally accepted practices (GAPs) of their respective businesses. Similar to government regulations, these standards and practice documents include data definitions, data standards, and business rules.

**Improving operations:** There are core processes associated with all businesses, and the absence of data oversight diminishes operational efficiency. Identifying data dependencies and building defining data controls providing oversight can lead to all sorts of business improvement, such as increased revenue, lowered costs, shorter product time-to-market, reduced rework, improved vendor relationships, and even higher customer satisfaction.

**Business renovation initiatives:** As organizations begin to explore how innovative technologies (such as big data, Internet of Things, or business analytics) influence corporate-wide business improvements, they must understand that careful data administration and oversight contribute to successful technology adoption.

## Understand the Types of Data Policies

Data policies govern aspects of all phases of the data lifecycle, from requirements assessment through modeling, acquisition, storage and management, integration, protection, quality, and disposition. Data policies can be organized around operational functions, such as the following:

- **Data architecture management:** These policies will guide the structural specification of data models, the methods used for devising data models, standards for selecting and implementing database technologies, standards for alternative storage frameworks (such as object representations like XML or JSON, or other NoSQL data representations), the platforms and methods used for storage, the technology used for transmitting data, and the methods for sharing data.

- **Observing reference data standards:** Observing reference data standards limits variability and interpretation of data values and data elements. Standards guide reference data models, advise developers about conformance to those models, provide standard representational formats for data types and their values, enumerate reference data sets (such as “states of the United States” or “claim diagnostic codes”), and formalize naming conventions. Data standardization rules direct how non-conforming data values are transformed into the standard representation.
- **Data protection, handling, and security:** The policies guide the classification of sensitive data. This includes guidelines for specifying levels of sensitivity, documenting the formats in which that information can be managed (such as electronic, paper, microfiche), determining what protection are necessary, directing the collection of only what is needed, specifying the access rights for different organizational roles, designating the methods for physical security and protection (such as physical storage requirements), designating the methods for electronic security and protection (such as encryption and masking), and providing standard language for disclosure and data use agreements.
- **Data entity management:** These policies govern entities that are critical to the operations and management of the business, such as location, product, customer, employee, and vendor. Entity management policies specify reference models for each entity type, mappings from reference models to physical representations (including relational, object, and NoSQL models), centralization of entity repositories, and guidelines for reuse of shared entity data among the different business functions. Data entity policies will also include rules for entity uniqueness and identity resolution. These policies are also meant to guarantee cross-application coherence and synchronization so that all data consumers see a consistent view of shared data.
- **Data provenance:** Data provenance policies ensure that sources of data critical for compliance (e.g., clinical trials) can be traced back to authoritative sources. These policies facilitate data reuse while guarding against improper use, misinterpretation, and noncompliance with data use agreements. They help data stewards establish reliable sources for information by documenting how data flows from acquisition through all points of use. In addition, these policies direct capturing source definitions, documenting semantics, review of usage patterns, and logging consumers to document compliance with data use agreements.
- **Data storage and retention:** These policies are designed to control and manage the accumulation, retention, archival, and disposition of data within the organization. These policies specify the types of technologies used for data storage (such as magnetic tape, conventional disk storage, storage area networks, or solid-state disks), and rules about assigning data to particular storage formats. In addition, when there are external requirements for data retention and

disposition, these policies govern retention durations, time frames for archiving (that is, moving data from near-line to off-line storage), the requirements and methods of disposition, and guidelines for preventing unauthorized recovery of deleted data.

- **Data Integration:** These are policies that guide how data sets from separate sources are combined and transformed for downstream use. These policies may be linked with the policies for data handling and protection, particularly when combining data sets results in exposing sensitive information that was not available in any of the original sources.
- **Data quality control:** These policies are intended to formally describe data quality expectations and specify the processes by which data values, instances, and sets are validated against those expectations, the methods for measuring conformance to data quality rules, the ways that data validity is monitored and reported, and the processes for assessing criticality, prioritization of discovered issues, root cause analysis, and remediation.

## Policies, Standards, Rules

There is a discrete hierarchy associated with business-oriented data policies that simplify the data steward's job of operationalization. This hierarchy differentiates policies, standards, and rules.

- A **policy** is a high-level description of one or more statements intended to influence behavior at a global level. The policy describes expected outcomes.
- A **standard** provides a framework for what is and is not acceptable. Defining standards establishes the means for validating conformance to the expectations.
- A **rule** directs or constrains actions to ensure that when standards are met, the policy is observed. In the best scenario, implementing the rules will enforce compliance with the standards, thereby demonstrating observance of the policy.

For example, a data policy might state that a customer's personally identifiable information (PII) will be exposed only to those in the organization who need to see it. One defined standard would enumerate those data elements deemed to contain PII, while another standard would describe different roles and corresponding access rights to view PII. Some corresponding rules would include instructions to project owners to insist on encrypting the data values in data elements deemed to contain PII as well as instructions for decrypting the data in response to a request from an individual with the appropriate authorization.

## Improvements to Conventional Data Governance

---

Conventional data governance approaches depend on three pillars:

- An **organizational structure** that distinguishes different classes of participants in the data governance program: An executive level (providing sponsorship and championing the program), a program level (for administration, coordination, and communication), a business level (for a data governance council representing the business functions), a tactical and operational level (consisting of data stewards, metadata analysts, data quality practitioners, etc., who actually execute the data governance procedures).
- An **operating model** that describes the different roles, their responsibilities, and their interactions associated with defining, debating, approving, and deploying data policies.
- **Technologies** that support the data governance program, including metadata repositories, business rules engines, data quality suites, and dashboards for reporting and monitoring, among others.

These are all necessary components of a data governance program, yet the following challenges can hinder implementation:

- **The Org Chart Challenge:** You need an organizational structure to manage the data governance resources. Program managers spend a lot of time formulating the different organization levels, tweaking job descriptions, defining responsibilities, and identifying candidates from across the organization to fill the roles. But creating and filling that org chart is just the means – but not the end – to establishing a data governance capability. At the beginning, many candidates hesitate to commit to participating until a critical mass of players accumulates. And as the program gets off the ground, normal waves of corporate turnover can have grave impacts, especially when one of the program’s champions seeks greener pastures.
- **Roles without responsibilities:** Sometimes individuals are tapped to fill roles even though role descriptions and responsibilities are not completed. It is unfortunate when new data stewards lack a proper understanding of their expectations, or when data governance council members agree to policies before any plans are in place for enacting those policies. Additionally, some organizations bring in newly hired data stewards (with little corporate knowledge) instead of developing and cultivating them from within. This leads to additional delays while the new employees acclimate to their new environment and gain sufficient understanding of the business.

- **Inoperable Operationalization:** Just because a data policy has been approved does not mean it is automatically put into practice. Different types of policies, standards, and rules have different methods of implementation, and all of them require buy-in from a range of potential participants. This means convincing people to commit to process renovation as well as make changes to the system development lifecycle to incorporate embedded data controls, generation of alerts, and data compliance reporting. Yet in many situations, there is no practical plan for engagement, socialization, and application design that allows integrating data policy compliance directly into the business applications.
- **Metadata busy work:** An important part of data governance and stewardship is developing business glossaries and capturing metadata for critical data elements. Yet the availability of automated technology to accumulate statistical information about data sets and store it in a metadata repository is often taken as a license to perform a broad, undirected scan of all the data assets to document the metadata. While capturing metadata is valuable, these automated processes require human input to review and qualify what has been discovered and to agree on business term definitions for the glossary – activities that are often overlooked.
- **Change mismanagement:** The need for data governance indicates that existing practices and processes are inadequate to demonstrate compliance. A data governance council can define, approve, and implement compliance to data policies, but in practice, nothing is going to happen until the critical mass of participants are engaged, empowered, and trained to modify noncompliant behaviors. Unfortunately, most conventional data governance program blueprints lack guidance for motivating change management.
- **Yet another meeting:** Conventional data governance programs will tend to schedule lots of meetings: data governance council meetings, data steward meetings, meetings to review data issues, etc. Meetings are beneficial when there are agendas to be followed and when the purpose of the meeting is to review completed work and to identify specific action items to track. But when untrained data stewards are expected to meet to review data issues they are ill-prepared to fix, their meetings will not yield data management improvements. Data governance council meetings must be brief enough to maintain the focus of those senior managers who are contributing their time, or they will quickly lose interest.
- **Technology addiction:** Metadata scans are just the tip of the iceberg. Metadata repositories, data profiling tools, analytics tools, business rule systems, policy repositories, and business glossary tools all contribute to the data governance program, but do not constitute the data governance “solution.” Over-reliance on technology can distract from the real need for effective process improvement.

These issues can overwhelm and derail the data governance program. And the appearance of these challenges can indicate a more insidious problem: many data governance programs are disconnected from the core business processes with which they should be *tightly coupled*.

## The Data Governance Reboot

---

These challenges are rooted in the dependence on the two aforementioned implementation styles. The top-down style establishes the high-level organization structure that issues the directives to the company, but does not provide the means for implementation. The bottom-up style leverages technology to assess and scan data assets, but is limited in directing compliance with policies. Organizations that rely on only one or the other style are often stymied in creating an effective program by three common unmet needs:

- The need for direct business engagement, to establish the criticality of instituting data governance and the significance and accountability of the different roles involved.
- The need for long-term program management that considers organizational needs (such as funding, staffing, and technology resources) at the pre-launch, start-up, establishment, and maintenance stages of the data governance program lifecycle, as well as understanding how the various roles change over time.
- The need for quantifiable measures of qualitative progress that are tied to business objectives, not technical milestones.

### Addressing the Shortcomings of the Traditional Model

Making data governance successful and sustainable means addressing the shortcomings of the conventional model. In essence, a data governance “reboot” defines a modernized data governance strategy that embraces the organization’s business goals. It directly engages the business, provides vision across the program lifecycle, and specifies qualitative metrics of progress. You can reduce the risk of diminished effectiveness or even outright failure by undertaking a program that links data policy compliance to meeting business goals and creating corporate value.

### Link the Top-Down and Bottom-Up Approaches

The key is to combine the best practices of the top-down and bottom-up styles by augmenting them with “middle-out” processes that connect business-oriented actions to the results of the top-down directives or outcomes of the bottom-up techniques. As an example, a top-down approach includes

soliciting information from the business leaders about fundamental corporate objectives and their affiliated business policies. A bottom-up approach will look at all the data sets and survey all the data elements and try to identify the core data domains that are touched by multiple business processes. Combine these with a “middle-out” process that determines how business-oriented data policies that are directly linked to the business policies impact the touch points for data domains and their data elements.

This is the fundamental theme behind the data governance reboot: find the intersection between the results of the bottom-up processes and the directives issued from the top-down processes. You can then integrate data controls and data directives into the application development process to implement continuous measurement, monitoring, and notifications when noncompliant data is found. These techniques effectively operationalize data policy compliance.

## Implementing Business-Oriented Data Governance

---

Refine business policies into data policies by identifying the data dependencies that are directly connected to business goals. This implies a refinement process designed to tease out the implied standards and rules inherent in the business policy. Transform high-level data policies into collections of standards and business rules. Data standards establish the framework for data value acceptability, while rules measure compliance with those standards. When melded with the right technology, standards and rules drive operationalization of data governance in a way that yields measurable and reportable results.

Defining standards and rules allows you to retool processes to add monitoring and auditing of compliance to data rules, and consequently observance of the corresponding business policies. Staff members will be motivated to take responsibility for complying with data policies when they are recognized as adding value (so that their relevance cannot be ignored).

### Defining Policies, Standards, and Rules

Take the following steps to define data policies, standards, and rules that are adaptable to the mechanics of compliance:

#### *Clarify Business Objectives and Identify Business Policies*

Differentiate between ill-defined desires for “better data” and precise targets for measurable, data-driven value. Engage the organization’s business leaders and discuss corporate goals. Determine the drivers of business policies (such as regulatory compliance, improved productivity, or increased sales margins, for example). Identify business policies with desirable outcomes that you can measure. Work with the business leaders to prioritize the business policies based on anticipated benefits.

### *Identify Data Dependencies*

Many if not all of the desirable outcomes will be influenced by the quality of the information. Find the discrete links between a business policy and particular data sets and precise data elements. This achieves two goals:

1. You are able to document the lineage tracking data concepts to their manifestations in the actual data sets used by different business processes. This allows you to link data use to specific business goals.
2. It helps you find the application touch points in the different business applications that are likely candidate locations for inserting data controls to validate conformance to data rules.

### *Define Data Policies*

Business policies specify principles, guidelines, and courses of action from the business side. The data analyst translates those business principles and guidelines into data policies and standards and the courses of action into rules associated with the types described earlier.

### *Specify Methods of Compliance*

Compliance means conforming to standards and obeying the rules. Some of these rules will require manual methods for oversight and review. However, many will be discrete assertions about data values (such as “the value must be one of the 2-character abbreviations for States of the United States”) and about relationships between data values (such as “the completion date must be later than the start date”). These types of data rules are candidates for automated validation using data profiling and assessment tools. Review the different standards and rules, determine whether they can be automatically validated, and determine the best methods for using tools to monitor and report data standard and rule compliance.

### *Develop Business Metrics*

A business metric matches a quantifiable measurement of data compliance to a score that conveys business relevance. Well-defined standards and business rules should be easily linked to their corresponding business contexts and objectives. For example, consider specifying a standard to compute the value of a data element representing a concept defined in a financial regulation. A measurement counts the number of times that the data element’s value is properly computed. A corresponding metric relates the risk of regulatory penalties to the percentage of times that the data element’s value can (or cannot) be validated using defined data rules.

### *Identify Business Process Touch Points*

You establish the lineage for the critical data elements that are central to data policy compliance as a byproduct of analyzing the data dependencies. Using the mapping to the business application touch points, the data steward can determine where data instances are created, updated, or just read. These

business process touch points are the appropriate candidates for introducing measurements that can feed your relevant metrics.

### *Configure Data Controls*

Data controls implement validation measurements. A data control either validates data that has been provided to the system, or ensures that data quality and usability rules are observed when the data is presented to a data consumer. These controls can be configured as reusable data services or embedded directly into application code.

## Operationalization Is a Repeatable Process

Defining data standards and rules is essentially a repeatable process. For each set of business policies, walk through this process to define executable data controls that provide quantifiable measurements that can be combined to report against defined business-relevant metrics. And fortunately, this repeatable process can begin small by selecting a relatively straightforward set of business policies and defining the corresponding data standards and rules. As those are moved into production, the scope can be expanded to a different set of business policies. Once these pieces are in place, a standard implementation plan can be executed to deploy the controls and connect the metrics to a data governance dashboard that reports on the state of compliance to business and data policies.

## Technology Support

---

The conventional approach to data governance is challenged by the limitation of trying to balance top-down directives with bottom-up data analysis. The program manager tries to establish specific technical governance roles while simultaneously acquire technical tools for the new team to use for stewardship activities. The business-oriented approach bridges the gap between the organization chart and the use of technology by creating a holistic model that focuses on balancing the practical governance implementation with observing overall business policies.

This holistic, “middle-out” approach provides a roadmap for leveraging technology to develop data controls and reporting whose implementations are integrated directly into the system development lifecycle. It refocuses attention on achieving business goals by adopting tools to perform the right supporting functions.

### Tools Enable, But Don't Drive the Program

Instead of letting tool acquisition drive the data governance process, it *enables* the data governance program. Careful determination of technology needs sets stakeholder expectations and communicates that business policy compliance is a byproduct of assurance of data quality and usability.

With that said, organizations that embrace this approach will need appropriate tools for their particular business situations. Through briefings and reviews, the report authors have examined the essential offerings of data governance tool vendors as well as how their development and sales approaches align with data governance modernization. We were pleased to encounter a growing consensus around the need for tools and technology that support the data governance process instead of an attempt to drive the data governance program.

## Data Governance Technologies

A number of tools perform different functions to guide a business-oriented data governance approach. The types of tools and capabilities that organizations should consider include, but are not limited to, the following types of tools.

### *Data Discovery and Profiling*

The quality of an organization's data remains an essential requirement for governance activities. Traditional data governance expects data stewards to "cleanse" data before those data sets are made available. Broader definitions of quality and usability by different communities of data consumers complicate the cleansing process because different users may apply different rules and data standards to measure quality. Defining business-oriented policies, standards, and rules presents further challenges, as these may need to be applied in different ways in different contexts.

That is where data profiling and discovery tools come in. These tools are used in three ways: (1) assessing data to identify potential anomalies and inherent metadata, (2) evaluating the usability and sensitivity of the content, and (3) using data profiling to proactively monitor conformance to defined data rules.

You use discovery and profiling tools to collect statistics on data, validate the data elements' formats and data types, and quickly identify potential anomalies requiring data steward attention. Catching issues early in the process prevents the propagation of noncompliant data. Data discovery tools can help identify data fields that contain sensitive data and should be tagged as PII or protected health information (PHI) in case data protection policies need to be applied and those values need to be masked, encrypted, or redacted.

But recognize that aimlessly scanning the data using a data profiling tool will not improve data usability. Rather, data stewards must have a good understanding of the business goals, the business policies, and correspondingly, the dependent data. Staff members who understand business policy are more likely to recognize data anomalies, differentiate potential issues based on business priority, and recognize potential policy violations. That contextual knowledge is a prerequisite to effective use of data profiling

and discovery tools to prevent accidentally overlooking issues that might negatively impact the business.

### *Business Glossary*

An increased need for integrating data from many sources exposes semantic inconsistencies. To observe data policies, there must also be standards for the meanings of commonly used business terms. This raises the importance of documenting and harmonizing business definitions. A business glossary provides the environment for capturing the authoritative definitions of business terms and data element concepts.

The recent generations of business glossary tools incorporate collaboration around workflows where business representatives can share their definitions and other significant aspects, expectations, and rules associated with the business term or data element. The collaborative environment facilitates discussion, agreement, and approval of definitions among those with domain expertise. The business glossary also highlights when the same term has multiple meanings in different business contexts, or when similar terms have the same definition. Data stewards take advantage of this capability to harmonize meanings and solidify the understanding of data. This is especially critical when the volume of data exceeds the ability to manually capture and track definitions.

Several products incorporate an initial set of terms within the glossary that are suited for particular business domains. Tool users can access these reference domains, which are commonly based on data hierarchies developed by standard bodies, regulatory agencies, or other industry groups for the relevant business domain.

### *Data Standards Management*

As we have discussed, data standards are critical as the foundation for compliance with implied and/or stated business policies. Data standards are often managed within metadata tools, but are also linked with some of the other types of data governance tools for assessment, lineage, business glossary, and validation. Standards help create efficiencies within the organization as well as through the industry ecosystem. They simplify data exchange and help in scaling cooperation among organizations by integrating with an increasing number of partners and client systems.

Some industry associations establish their own sets of business-specific data standards (such as the facets of parts and their corresponding measurements). In addition, industries such as banking, finance, health care, and pharmaceuticals have data standards that are defined by regulatory agencies. Because of the tight coupling of data standards to specific industries, some products are industry-specific, and are continually updated with the externally-defined standards.

### *Data Lineage and Provenance*

Data lineage tools allow you to track the origin of data elements along the data flow, including the locations and types of transformation that have been performed. Data lineage provides a historical record of where and how data were brought into the organization and offers insight into the cross-function data dependencies. Data lineage enables the data analyst to perform impact analysis to determine where applications need to be reviewed should some governing business policy be modified.

Data provenance is of particular importance in industries with compliance or regulatory requirements. In regulated industries such as pharmaceuticals, being able to track information about drug manufacturing or subjects who are enrolled in a clinical trial, or demonstrating data lineage to auditing agencies are critical requirements. Other regulated industries such as banking and finance have similar requirements, including the U.S. Federal Reserve's Comprehensive Capital Analysis and Review (CCAR) and the European Commission's General Protection Data Regulation (GDPR).

Trying to implement an approach that covers lineage of all data being managed by the organization can become overwhelming for those tasked with lineage activities. The "middle-out" approach allows data stewards to adapt data lineage and provenance capabilities to their needs. It enables them to identify those domains that are most important to the business and lets them focus on incrementally tackling those areas. Iterative refinement allows the stewards to work through remaining areas of significance in priority order to meet the necessary business and regulatory requirements. And when tools provide data stewards with a visual representation of data element lineage, they can better assess the best locations for instituting data controls.

### *Data Catalog*

As more organizations begin to build and use data lakes (general repositories for all corporate data assets), the potential for disarray grows. Increased data volumes and variety complicate the tasks of capturing and tracking data. This is particularly true in data-driven environments in which new types of data and data sources appear with little or no notice to staff.

As we have suggested, data profiling helps in assessing new data sources. The next step is organizing and analyzing the profiling results and documenting what can be learned so that newly-acquired data assets are accessible, usable, and compliant with existing data policies.

Data catalog tools can be used to help an organization organize, inventory, categorize, and keep track of data assets. These tools allow analysts to find what they need, discover what's new and interesting, and capture knowledge. A data catalog captures all the information necessary for enabling use, including structural metadata, semantic metadata, storage format information, location, data access roles and rights, and the data standards and rules that are applied to the data set. Catalogs often provide a means for defining data categories and having assigning data elements to one or numerous categories. This also provides a way for users to identify related or associated data elements.

Several vendors blend the capabilities of their discovery and catalog tools to include machine learning algorithms that more quickly identify and catalogue incoming data. The results are submitted to data experts who monitor and validate the findings and facilitate inclusion in the data catalog.

### *Business Rule Management*

Most businesses define a set of metrics associated with key performance indicators (KPIs) that measure value and success. And as we have suggested, a number of data governance tool vendors have acknowledged the link between business goals and data usability by providing a means for managing business-oriented data rules and metrics. These tools provide an interface for users to define rules that can easily be tied to individual atomic data elements associated with corporate KPIs. This helps ensure that the organization is meeting business objectives as well as regulatory or compliance requirements.

Data business rules can be viewed as a way for enforcing a level of quality and completeness that may affect multiple metrics and processes. Business rule management tools can help senior business leaders to better articulate corporate objectives for data use. By providing a repository for these rules, the business rule management tool supports their communication across the enterprise. This better enables rule operationalization and encourages prioritization of the practical implementation of data governance within the company.

### *Requirements Management*

When an organization operates in an environment with regulatory or compliance requirements, there are typically only a small number of experts that completely understand the regulations' details. Noncompliance poses the risk of financial penalties, brand erosion, and negative impacts to the customer experience. Therefore, capturing and sharing data requirements and subjecting them to version control improves the ability to socialize data governance and spur its adoption.

A number of data governance tools provide the ability for capturing, managing, and sharing requirements. This expands their visibility beyond the experts and allows a larger cohort of staff members to participate in ensuring the requirements are met. The tools often also include workflow capabilities and security protections so that only those with appropriate permission can modify the requirements. Collaborative features allow you to share requirements for review with others. When finalized, the tool allows you to publish the requirements for a broader audience to see and use. Different types of notifications and alerts (such as texting or emails) are directly incorporated to inform requirements "subscribers" of updates.

### *Integration with Desktop Productivity Tools*

Some products provide plugins that allow users to access information captured in their governance platform directly from their desktop productivity tools, such as Microsoft Office. For example, rather than requiring users to open their governance tool to reference information, services are configured to let them

access the information directly within a drop-down menu added to an email client (e.g., Outlook) or a word processing tool (such as Word). Users can quickly access information located within the governance platform for inclusion in documents and communications with team members.

## Considerations

---

This report is intended to raise awareness of the need for reconfiguring the organization's approach to data governance. We have presented ideas for renovating the ways that data policies and standards are defined, as well as reframed data governance as a business-driven activity. We suggest that the group implementing data governance gain a strong understanding of the meaning of data governance, quality, control, and usability from a business perspective to improve how data governance is implemented from an operational perspective. The processes and supporting tools can then help members of the organization tackle what is otherwise an overwhelming task, and through a more thoughtful process achieve the organization's business goals.



Need help with your business analytics or data management and governance strategy?  
Want to learn about the latest business analytics and big data tools and trends?  
Check out **Eckerson Group** research and consulting services.