Implement Your Data and Analytics Governance Through 5 Pragmatic Steps

Published 6 July 2020 - ID G00729295 - 18 min read

By Analysts Guido De Simoni, Saul Judah

Initiatives: Data Management Solutions **and 1 more**

Data and analytics governance has become more challenging. Progress is hindered by new requirements, new technology capabilities and a lack of maturity in the discipline. Data and analytics leaders should consider five key steps to implement data and analytics governance for business value.

Additional Perspectives

- Summary Translation: Implement Your Data and Analytics Governance Through 5 Pragmatic Steps
  (25 August 2020)

Overview

Key Challenges

- The diversity and pace of business demand for data and analytics far exceeds the ability to meet it through existing governance capabilities.
- Coordinating the many different types, sources, users and uses of data assets across an organization can be a complex operation requiring diverse implementation approaches and styles for data and analytics governance. Fifty-five percent of respondents in a recent Research Circle Survey on data and analytics governance said the lack of a standardized approach to governance blocked its governance objectives.
- To align data and analytics governance programs and succeed with data and analytics programs, emerging technologies must be evaluated, requiring a significant effort in assessing how and in what to invest.

Recommendations

D&A leaders addressing the implementation of data and analytics governance should:

- Drive targeted stakeholders' engagement, from governance strategy to scoped execution, by leveraging agreed upon corporate goals.

- Define with effectiveness the measurable activities required for deployment, and the roles involved, by linking the business scenario to key data management implementation steps.
- Expose the data rules that must be applied when executing on the data and analytics strategy for business value by leveraging essential policies to enable key governance decisions for continuous improvement.
- Optimize technology investment to support data and analytics governance execution by selecting data management critical capabilities (that is, capabilities available across data management programs like data quality, master data management, metadata management and data integration).
- Explore, test and validate how modern solutions can operationalize and automate data and analytics governance implementation by deploying augmented data management options for future enhancement.

Introduction

Gartner defines *data and analytics governance* as the specification of decision rights and an accountability framework to ensure the appropriate behavior in the valuation, creation, consumption and control of data and analytics. It is very much a multidimensional discipline where people, data, process and technology contribute to outputs, outcomes and business engagement with evolving capabilities and focus (see "7 Must-Have Foundations for Modern Data and Analytics Governance").

According to Gartner's 2019 Data & Analytics Governance Survey, 80% of the participants believe data and analytics governance is important in enabling business outcomes. Yet, from the same survey, we know that four in 10 participating organizations do not assess, monitor or measure data governance. With data and analytics governance in such poor shape in most organizations, it is unsurprising that its decisions are implemented in a disconnected manner (see "The State of Data and Analytics Governance Is Worse Than You Think").

The most common data and analytics governance model is based on the command-and-control approach most associated with compliance and regulation initiatives. This is a serious impediment to the ability of organizations aiming to achieve digital business success and business value. Historically, governance has been IT-led and command-and-control oriented. Today, with increasing numbers of data creators and consumers, governance must shift to being about enablement and appropriate levels of more granular governance.

Business value driven by data and analytics is created within business functions as well as through formalized data and analytics teams. Because governance practices in business functions are localized, fragmented and inconsistent, they limit the value realized and scalability of investment.

While command and control limits data and analytics leaders, adaptive governance enables them. Adaptive governance requires new practices as the new organizational capability that determines the governance styles and mechanisms that will deliver required business outcomes in a business context (see "Data and Analytics Leaders Must Use Adaptive Governance to Succeed in Digital Business").

Additionally, modern data and analytics use cases need a portfolio of capabilities driven by their constantly changing requirements, which will not be fulfilled well enough through point-based tools. Digital transformation is a driver of this convergence. Organizations want to have automated, synchronized, integrated, cost-effective and efficient solutions with a central design, yet a distributed deployment.

The need for such aggregation of capabilities or platforms is driven by the growing recognition that the work of data and analytics governance is different than the work of data management. Although the capabilities that serve both are similar, the context in which those same capabilities are used differs between governance and management (see "Modern Data and Analytics Requirements Demand a Convergence of Data Management Capabilities").

Since most organizations are still in the process of strategizing and deploying data and analytics governance across the organization, data and analytics leaders must:

- Use a list of agreed upon corporate goals to scope target scenarios
- Link the business scenario to key data management steps
- Leverage essential policies to drive key decisions
- Select technology based on critical data management capabilities
- Deploy augmented data management options

Analysis

Use Agreed Upon Corporate Goals to Scope Target Business Scenarios

A successful data and analytics governance initiative starts when organizations leverage a well-crafted data and analytics strategy that reflects the goals of the corporate strategy. This requires identifying the strategic business outcomes and their relative priorities (see "How to Craft a Modern, Actionable Data and Analytics Strategy That Delivers Business Outcomes"). The organizational business strategy should be used as a primary source to understand the goals and direction the organization expects to take, the market drivers, and the regulatory landscape.

Often, the priorities are expressed as critical business programs that are either in-flight or planned, and these have assigned executive stakeholder leaders. This implies that, before proceeding, it is critical to share high-level findings to the key business leaders. Ensure you have expressed the problem statement in the right business language, correctly understood business priorities and made valid assumptions.

Effective stakeholder engagement requires data and analytics leaders to deconstruct the business problem and decisions in the underlying data and supporting analytics (see "Tool: Enable Data Literacy Through Stakeholder Analysis and Linking to Business Outcomes"). The full definition of the data analytical requirements must articulate:

- An understanding of the data and analytics outputs that will be provided as part of the decision process
- That the required datasets previously identified are the foundation for leveraging specific scenarios for executing on data and analytics governance

Further reading:

"How to Optimize Business Value From Data and Analytics Investments ... Finally"

"Tool: A Living Library of Real-World Data and Analytics Use Cases"

Link the Business Scenario to Key Data Management Steps

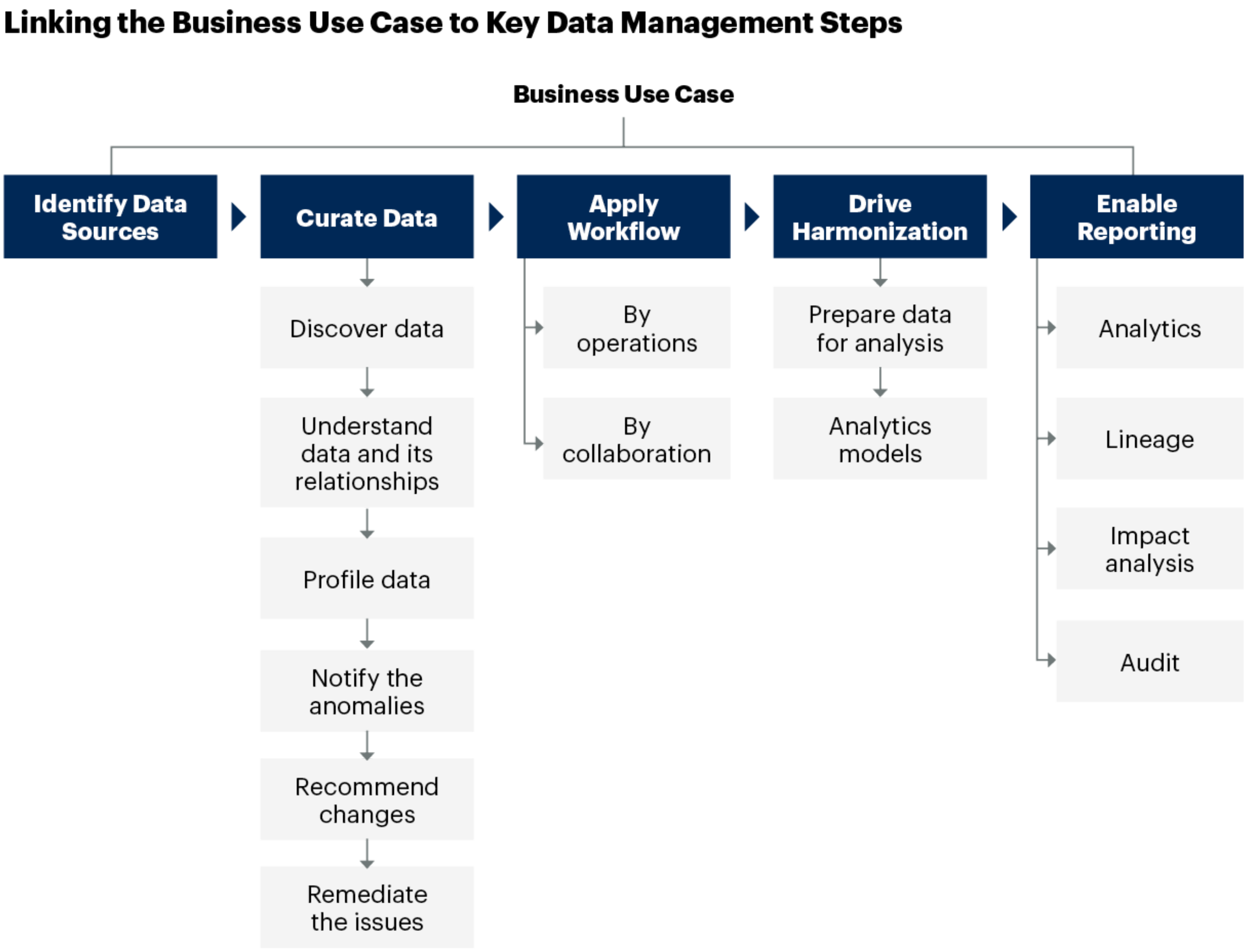
You need a model to be able to describe how various data assets relate for a specific business purpose. Without a model, you cannot discover the relationship between assets, without which the data is of limited use. This is about moving from a use case scenario to the semantics of it.

Semantics enable:

- Better understanding of data — You can share a more complete understanding of data and its structure among both people and applications (see "Data Modeling to Support End-to-End Data Architectures").
- Reuse of knowledge — Relationships identified by a group of users can be reused by other groups or expanded for additional use cases that embed the previous identified relationships (see "How to Use Semantics to Drive the Business Value of Your Data").
- The underlying information supply chain — By mapping the information supply chain, data and analytics leaders can more explicitly communicate the utility and value of data and accountability for it (see "Applied Infonomics: Use a Modern Data Catalog to Measure, Manage and Monetize Information Supply Chains").

Figure 1 shows how the implementation of data and analytics governance leverages semantics in order to optimize the context of work associated with a scenario and the use of technology to operationalize it.

Figure 1: Linking the Business Use Case to Key Data Management Steps



Source: Gartner
729295_C

Implementing data and analytics governance rules through data management activity requires the following activities:

- **Identify data sources:** Begin your data discovery journey by identifying which data sources are needed to achieve the goals of the use cases prioritized and modeled (semantics). Roles involve data owner and subject matter expert (SME).
- **Curate data:** The primary purpose of this step is to classify and curate to:
  - Discover data assets owned by organizations by exposing the metadata.
  - Understand data assets to establish provenance and identify relationships that provide a complete picture of data.

  - Profile data to better understand its structure. This involves gathering "technical metadata" and profiling the statistics on the data (useful in flushing out any anomalies and data quality issues).
  - Recommend changes at the data and metadata levels.

- **Apply workflow:** The operations part of the workflow monitors how the data pipelines and data jobs run and, if they fail, notifies the appropriate parties. The cross-team collaboration workflow is required for several use cases — for example, regulatory compliance — to control which partners have access to what data by partner type. Roles involved are business process manager and data steward.
- **Drive harmonization:** This step starts the process of enabling departments to perform analytics on the curated corporate data assets. Roles involved are data engineer and data scientist.
- **Enable reporting:** This step enables decision making in relation to specific requirements. For example, new regulations, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), require lineage to be traced from data source to data consumption. Impact analysis helps data outliers to be identified to minimize data quality problems in the downstream data pipeline processes. Roles involved are business analyst, data consumer and decision maker (including the data steward).

There might be different feedback loops between the steps. For example anomalies identified to minimize data quality problems can drive changes in the source data schema (schema drift) or changes in the model.

Further reading:

"Create a Chief Data Officer Dashboard to Measure Your Business Impact"

"5 Steps to Build a Business Case for Data and Analytics Governance That Even Humans Will Understand"

"5 Steps to Build a Business Case for Continuous Data Quality Assurance"

Leverage Essential Policies to Drive Key Decisions for Continuous Improvement

A foundational aspect of data and analytics governance is specifying the essential principles, policies and practices that drive continuous improvement when the organization executes on the data and analytics strategy. The application of policies might be situational; however, when focused on a specific scenario, policies are harmonized and enforced across the scenario. This leverages potential operationalization and automation of policy execution by modern tools and solutions.

The policies that are most applicable to you depend on the business driver or outcomes you are aiming to achieve. For example, policies pertaining to quality are common when the driver is related to improving business outcomes, such as improving revenue through upselling or cross-selling customers.

Security-related policies are common when the outcome desired is compliance for legal regulations. Data and analytics leaders should consider how the following policy types apply to their organization:

- **Privacy:** Various information assets, particularly those holding personally identifiable information, must be treated with care to avoid exposing sensitive data. Policies for information privacy specify requirements for anonymizing such data. These policies often mandate adherence to legal requirements, such as the U.K.'s Data Protection Act 1998, or regulatory requirements, such GDPR and CCPA (see "Why Privacy Is an Opportunity to Drive Data Value").

- Security: Access rights to information assets are crucial for minimizing risk. Information security policies focus on who and what can access information. Clear mandates for segregation of duties and the principle of least privilege are included in such policies (see "Use Infonomics to Quantify Data Monetization Risks and Establish a Data Security Budget").

- **Quality:** Data quality is a key concern for organizations, and poor-quality data creates significant risk and challenges. This policy type specifies the required levels of "fitness" (for example, the validity, completeness and accuracy) for the information to have optimal risk and value to the enterprise. When dealing with data at scale, other criteria of evaluation apply, like triangulation, likeability and rating (see "How to Overcome the Top Four Data Quality Practice Challenges").

- **Availability:** There are various uses of data, spanning operations, applications, analytics and compliance. As such, there is a wide range of requirements related to the availability of data. One of the most critical elements of availability is timeliness.

- **Retention:** Data assets can lose value over time and even become a risk to the enterprise after reaching the end of their useful or legally required life span. Retention policies, which support governance across the information supply chain. specify when data assets must be archived and how, how long they must be retained, and when they will be disposed of (see "There Are Many Options for Data Retention After Application Decommissioning").

- **Ethics:** Ethics policies specify what things the organization will do and, more importantly, will not do with information to prevent violating the trust or privacy of customers or other stakeholders. This comprises sign-off procedures for the use of analytical results; an information use code of conduct; and mandatory use of technologies, such as dynamic masking for anonymizing data. In addition, artificial intelligence (AI) ethics is an emerging area in overall digital ethics (see "AI Ethics: Use 5 Common Guidelines as Your Starting Point").

Every policy could be applied across the data management steps analyzed in this document in relation to one or multiple scenarios. Scoping based on a critical scenario enables policy setting, policy enforcement and policy execution to be tested and validated for data and analytics governance effectiveness.

Further reading:

"Top 10 Legal Concepts That Data and Analytics Leaders Must Know to Drive Business Value"

"Smart Data Sharing — Five Insights to Get It Right"

"Adopt SMART Information Principles for Effective Data and Analytics Governance"

"Generally Accepted Information Principles for Improved Information Asset Management"

Select Data Management Critical Capabilities to Optimize Technology Investment

The contribution of data management capabilities to data and analytics governance is dependent upon the style, approach and scope of use cases (see "Data and Analytics Leaders Must Use Adaptive Governance to Succeed in Digital Business"). Here, we refer to the capabilities that align to the key data management steps and that support the defined policies.

Consider:

- Data management steps and capabilities
- Essential policies and capabilities

Data Management Steps and Capabilities

When the focus is on **data curation,** data and analytics leaders should consider investments in data catalogs/metadata management solutions. Data catalogs have now become the foundational components for self-service analytics and data governance and compliance, as well as an entry point for modern data architectures such as data lakes. These solutions use various approaches to curate and classify the data, including using built-in functions, machine learning (ML) to autoclassify, and manual tagging (ontology provided by business users). Approaches also include leveraging social collaboration to enrich metadata through data asset certification, ratings and reviews, and published ontologies such as the Financial Industry Business Ontology (FIBO).

In addition, the metadata is stored in a database that may be relational or a graph database. Graph databases are becoming more popular because they store the relationships and can be used to graphically display connections.

Conversely, **data lineage** is usually visualized via a visual map that displays the journey of the data from the source databases to its final analytical use or disposal.

To choose optimal lineage capabilities:

- Determine which products are involved in the data pipeline. Many lineage options from traditional RDBMS or ETL vendors may not support competitive products.
- Choose products that provide both summary and detailed end-to-end lineage and allow you to zoom in and zoom out to the level of detail required.
- Understand the level of customization within products.

Essential Policies and Capabilities

Lists of capabilities organized according to categories of work associated with policy setting, policy enforcement and policy execution are available across various data and analytics tools and technologies. They can be used for operationalizing various aspects of the work of data and analytics governance regarding policy management. We summarize our examples in Table 1.

Table 1: Policies and Industry Practices, Solutions and Capabilities

Policy	*To Address*	*Industry Practices/Solutions/Capabilities*
Privacy	■ Violation of privacy laws and consequential penalties ■ Exposure of customers to identity theft ■ Jeopardization of customer trust	■ Proactive identification and encryption of personally identifiable information ■ Data curation and masking
Security	■ Risk of unauthorized data access ■ Data breach ■ Risk of violating industry regulations	■ Access control list ■ Data encryption ■ Firewall rules ■ Strong password policy ■ Control physical access to data centers

Policy ↓	*To Address* ↓	*Industry Practices/Solutions/Capabilities* ↓
Quality (Truth)	■ Questionable data ■ Misleading data leading to poor decision making	■ Data discovery ■ Data profiling ■ Data quality ■ Data-cleansing services ■ Master data management ■ Metadata management
Quality (Trust)	■ Patterns of repeatable interactions/transactions ■ Paid or free data ■ Geographical jurisdiction of data ■ Comparability of sources ■ Data directionality ■ Change of data when used ■ Likeability of data ■ External contributions to data	■ Data discovery ■ Data analytics ■ Enterprise metadata management
Availability	■ Impacts on poor completion of tasks ■ Encouragement of “bad” data silos ■ Jeopardization of analysis and decision making ■ Absence of relevant data leading to inappropriate business decisions ■ Lack of sharing of data and collaboration within the enterprise in support of the decision-making process	■ Tools and applications that provide consistent access to data ■ Data discovery

Policy	*To Address*	*Industry Practices/Solutions/Capabilities*
Retention	■ Extension of disaster recovery time (exceeding service-level agreement [SLA]) ■ Loss of data ■ Loss of revenue ■ Jeopardization of customer trust	■ High-availability database clusters ■ Continuous data replication ■ Mirroring ■ Data replicas in geographically distinct data centers
Ethics	■ Violation of trust of customers ■ Violation of privacy of customers	■ Dynamic masking for anonymizing data

Source: Gartner (July 2020)

When different policies apply to the optimization of investment, they must follow a proper assessment at the technical and financial levels. For more details on the operationalization of policy management, see "The Role of Technology in Data and Analytics Governance."

Further reading:

"Data Modeling to Support End-to-End Data Architectures"

"Toolkit: How to Optimize Business Value From Data and Analytics Investments ... Finally"

Deploy Augmented Data Management Solutions to Accelerate and Enhance Governance

Augmented data management refers to the application of AI and ML for optimization and improved operations of data management tasks. AI and ML are applied — based on the existing usage data — to tune operations and to optimize configuration, security and performance (see "Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration"). They are also applied to create, manage and apply policy rules within the different products, such as metadata management, master data management, data integration, data quality and database management systems (see "Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders").

With distributed data and analytics solutions, the governance requirements also become distributed and dynamic. Traditionally, many data management systems have combined governance and management in their design and deployment. This wasn't a mistake; it was a compromise. But that compromise is no longer necessary. It is now possible to have new models that enable an innovative response.

Emergent governance is the new concept that should be leveraged with augmented data management options (see "Future of Data Management, 2019 Edition"). When data is authored, it has different management requirements than when it is reused in a second, third or further use case. Initial data authoring processes always capture data in a "revealed" fashion; the application or process that captures the data "knows" everything about the data. The business process that is represented in that authoring process already has governance rules.

For example, a system that authors security clearance for admission to an import duty facility "knows" that the identity of cleared personnel should not be made readily available to casual data users. Because the business process is about securing a facility, the access rights to that facility are also secure. When the nature of the business process for a data authoring system is known, then the data governance requirements are also known. When data is requested for reuse, this creates a second half of a data contract.

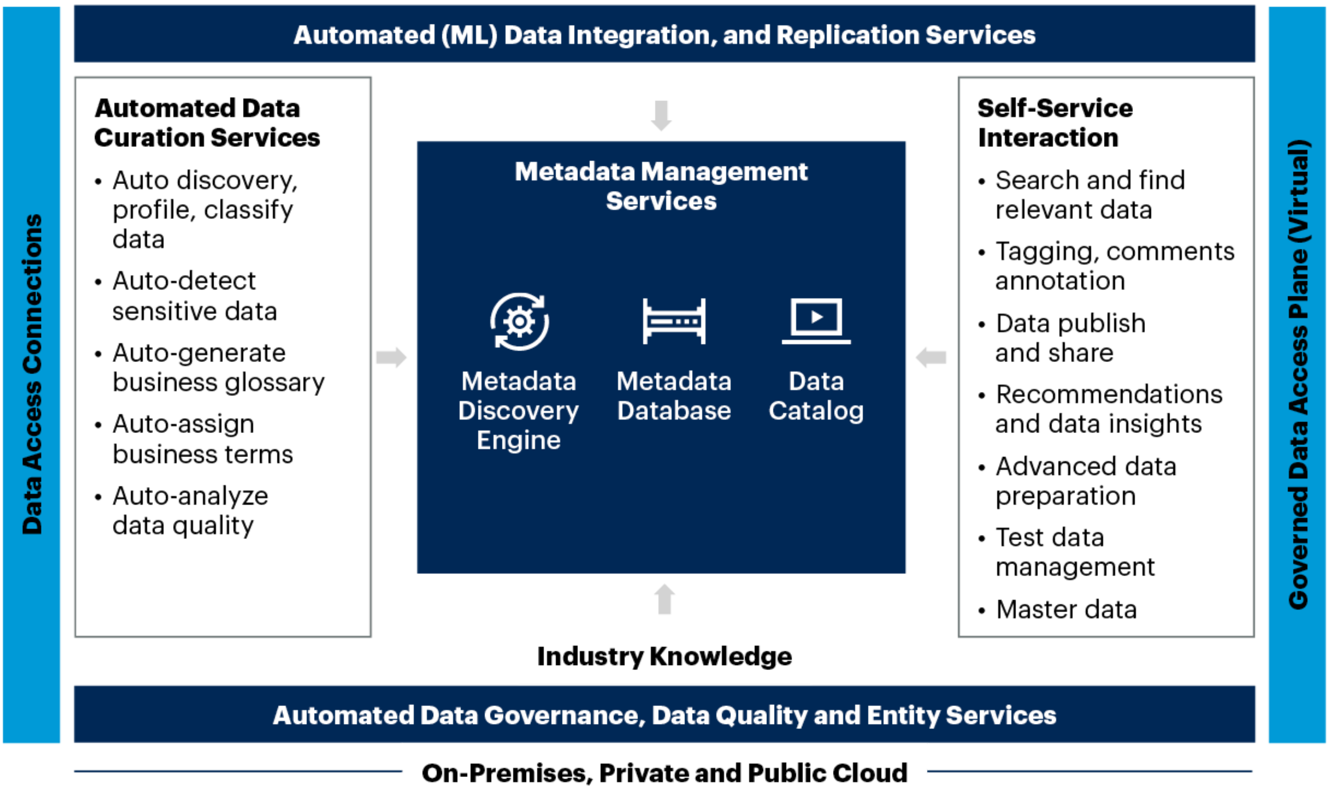
That reuse also has an associated business process that already has governance capabilities. By evaluating the use case against the authoring case, it is possible to introduce artificial intelligence to evaluate requests for data reuse. By tracking all such requests for reuse in metadata, a governance model emerges that has case examples that include situational analysis and can start to create "training data" for ML.

The emergent governance system or platform can then apply lessons learned to new data that is encountered. Emergent governance can determine the data and analytics governance rules without ever having a human look at the data. In cases when the decision is not binary — those "gray area" situations — humans with proper governance compliance can create training cases and further improve the model.

Figure 2 shows an example of such a platform.

Figure 2: Example of Data and Analytics Governance Platform

Example of Data and Analytics Governance Platform



Source: Gartner
729295_C

Organizations interested in the concept of emergent governance should test and validate the capabilities of such platforms. We are starting to see examples of platforms oriented to data and analytics governance from vendors such as:·

- Alex Solutions
- Collibra
- Global Data Excellence
- IBM
- Informatica (packaging separate products but tightly integrated and leveraging CLAIRE)

Further reading:

"Top 10 Trends in Data and Analytics, 2020"

Evidence

The analysis in this document is based on information from several sources, including:

- Conversations about data and analytics governance with users of Gartner's client inquiry service.
- Interactive briefings during which vendors provided Gartner with updates on their strategies, market positioning, recent key developments and product roadmaps.
- Data findings from Gartner's 2019 Data & Analytics Governance Survey. This survey was conducted online from 31 October through 22 November 2019 with 55 ITL Gartner Research Circle members — a Gartner-managed panel. Forty-four percent of respondents were located in North America, 33% were from the EMEA region, 15% were from APAC and 9% were from Latin America. All respondents were required to advise the team making decisions related to data and analytics governance, be a member of the team, or be a decision maker or team leader.

Recommended by the Authors

Modern Data and Analytics Requirements Demand a Convergence of Data Management Capabilities

Magic Quadrant for Metadata Management Solutions

Recommended For You

Summary Translation: Implement Your Data and Analytics Governance Through 5 Pragmatic Steps

Modern Data and Analytics Requirements Demand a Convergence of Data Management Capabilities

How Augmented Data Management Capabilities Are Impacting MDM and Data Governance

Data Engineering Is Critical to Driving Data and Analytics Success

Toolkit: Map Your Data Management Landscape With the Data and Analytics Infrastructure Model

Supporting Initiatives

Data Management Solutions

Track

Data and Analytics Strategies

Track