The SAS® Data Governance Framework: A Blueprint for Success

SsaS

THE POWER TO KNOW.

Contents

As a concept, data governance has been around for decades. By the 1980s, the computing boom led to technology designed to tackle things like data quality and metadata management, often on a departmental basis to support database marketing or data warehousing efforts. While pockets of "data governance" emerged – and sometimes grew as a grassroots effort – it was rarely a hot topic in the IT community.

By the early 2000s, data governance started to get more attention. The collapse of companies like Enron led the US federal government to establish rules to improve the accuracy and reliability of corporate information. Data governance was a key component of these efforts. For example, the Sarbanes-Oxley Act required executives to know – and be personally responsible for – the data driving their businesses. In the face of such industry drivers, data governance matured rapidly. Technology that grew out of data quality or business process management tools began to offer automated creation and management of business rules at the data level.

To keep pace with technologies, trends and new regulations, data governance continues to evolve. Consider, for example, the demands placed on governance by streaming data analytics, the internet of things, Hadoop and requirements of regulations like Current Expected Credit Loss (CECL) or the General Data Protection Regulation (GDPR).

The majority of C-level executives today understand why they need to manage data as a valuable corporate asset. And to ensure proper oversight, executives in emerging roles – like the chief data officer or the data protection officer – have appeared in many boardrooms. But despite the high adoption of data governance as a formal set of practices, there are many examples of organizations that are struggling to overcome failed attempts or tune ineffective organizations.

Consider this quote from an executive at an integrated health care provider:

> *"Jim is the fourth data governance director from the corporate office in the past six years. I hope this time it sticks."*

Or this quote from a risk manager in a regional bank:

> *"Our first attempt at data governance kicked off with great fanfare a few years ago, then fizzled. Now there is quite a bit of skepticism this time around."*

There are many reasons why data governance fails[1] – or at least, underperforms. But the cause is often one of these:

- Data governance is not well-defined.
- Business executives and managers consider data "an IT issue, " and business units and IT don't work together.
- The organization's unique culture isn't taken into account.
- Organizational structures are too fragmented, the culture doesn't support centralized decision making, or decision-making processes are not well understood or designed.

---

[1] Read more by downloading a SAS Best Practices white paper, *10 Mistakes to Avoid When Launching Your Data Governance Program*.

- Data governance is viewed as an academic exercise, or is treated like a finite project.
- Existing steering committees with knowledge and clout are overlooked.
- Execution is lacking; data is not managed in a structured, tactical and repeatable way.
- The return on investment (ROI) isn't clear, and it's hard to link data governance activities to business value.
- Key resources are already overloaded and can't take on governance activities.

So, no matter how much you need data governance, there are a variety of reasons it may not work. In this paper, we will highlight the SAS Data Governance framework, which is designed to provide the depth, breadth and flexibility necessary to overcome common data governance failure points.

Framing Data Governance

Starting data governance initiatives can seem a bit daunting. You're establishing strategies and policies for data assets. And, you're committing the organization to treat data as a corporate asset, on par with its buildings, its supply chain, its employees or its intellectual property.

However, as Jill Dyché and Evan Levy have noted, data governance is a combination of strategy and execution. It's an approach that requires one to be both holistic and pragmatic:

- Holistic. All aspects of data usage and maintenance are taken into account in establishing the vision.
- Pragmatic. Political challenges and cross-departmental struggles are part of the equation. So, the tactical deployment must be delivered in phases to provide quick "wins" and avert organizational fatigue from a larger, more monolithic exercise.

To accomplish this, data governance must touch all internal and external IT systems and establish decision-making mechanisms that transcend organizational silos. And, it must provide accountability for data quality at the enterprise level. The SAS Data Governance framework illustrates a comprehensive framework for data governance that includes all the components needed to achieve a holistic, pragmatic data governance approach.

The framework presented here is a way to avoid data dysfunction via a coordinated and well-planned governance initiative. These initiatives require two elements related to the creation and management of data:
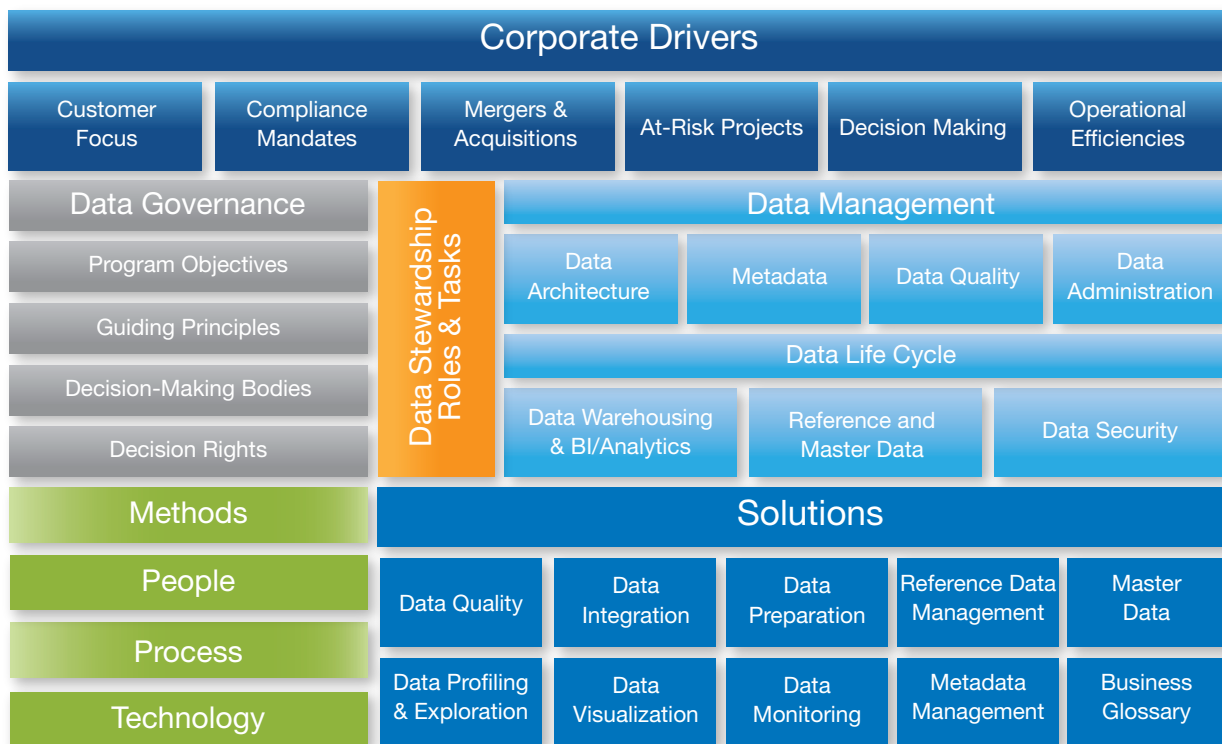
- The business inputs to data strategy decisions via a policy development process.
- The technology levers needed to monitor production data based on the policies.

Collectively, data governance artifacts (policies, guiding principles and operating procedures) give notice to all stakeholders and let them know, "We value our data as an asset in this organization, and this is how we manage it."

Definition of Data Governance

"The organizing framework for establishing strategy, objectives and policies for corporate data."

Jill Dyché and Evan Levy
Customer Data Integration: Reaching a Single Version of the Truth

Corporate Drivers					
Customer Focus	Compliance Mandates	Mergers & Acquisitions	At-Risk Projects	Decision Making	Operational Efficiencies

Data Governance	Data Stewardship Roles & Tasks	Data Management			
Program Objectives		Data Architecture	Metadata	Data Quality	Data Administration
Guiding Principles		Data Life Cycle			
Decision-Making Bodies		Data Warehousing & BI/Analytics	Reference and Master Data		Data Security
Decision Rights					

Methods	Solutions				
People	Data Quality	Data Integration	Data Preparation	Reference Data Management	Master Data
Process	Data Profiling & Exploration	Data Visualization	Data Monitoring	Metadata Management	Business Glossary
Technology					

The top portion of the framework – Corporate Drivers – deals with more strategic aspects of governance, including the corporate drivers and strategies that point to the need for data governance. The Data Governance and Methods sections refer to the organizing framework for developing and monitoring the policies that drive data management outcomes such as data quality, definition, architecture and security.

The Data Management, Solutions and Data Stewardship sections focus on the tactical execution of the governance policies, including the day-to-day processes required to proactively manage data and the technology required to execute those processes.

While the framework can be implemented incrementally, there are significant benefits in establishing a strategy to deploy additional capabilities as the organization matures and the business needs require new components. It's important to develop a strategy that can address short-term needs while establishing a more long-term governance capability.

On the bright side: Organizations never start from zero. Groups exist in your organization that have varying levels of governance maturity. As you develop your long-term data governance plan, the framework can help you understand how the individual components can be used as a part of the whole, helping you achieve a sustainable program for data governance.

Corporate Drivers

Many companies that recognize a need for data governance find it challenging to get broad consensus and participation across business units. Why? It's often difficult to tie the results to a business initiative or demonstrate ROI.

As much as possible, it's important to tie data governance activities (and investments) to corporate drivers. This will allow you to more rapidly link data governance "wins" with key business goals. Consider these contrasting, but very real examples of data governance program launches:

Regional Bank

A regional financial services company launched an initial data governance program with great fanfare and a surprising amount of business-unit support. It identified data stewards, acquired data profiling tools and decided to tackle data quality problems in its startup efforts. Profiling revealed the data elements with the largest amount of issues, and the teams went to trace data, identify root causes and find solutions.

There was only one problem. The fields identified as the most problematic (after months of mapping and tracing) were phone number and seasonal addresses, neither of which had any strategic value to the executives. The program failed to win incremental support, and executives turned their time and attention to problems that tied more closely to their business strategies and drivers.

Global Bank

This bank, while a larger institution than the regional firm, faced increasingly complex compliance mandates around Basel III and risk data aggregation principles. As a result, data governance became a sanctioned set of practices within the bank's multipronged compliance strategy.

Launching data governance with this more focused approach, the bank focused on data required for risk data aggregation. The data steward teams consolidated all siloed data quality efforts into a single area, identified the key data owners for this data, and built a program designed to illustrate how they were managing information to the necessary regulatory bodies. This program demonstrated its value and received increasing executive support.

As these two banks found, companies get more traction if the governance initiative links to a specific strategic initiative or business challenge. Not only does this allow governance activity and investment to follow corporate objectives (and make clear the ROI for such activity), but it also eliminates the "academic exercise" label that is sometimes applied to data activity. When aligning data governance with corporate drivers is not possible, bottom-up approaches can be championed to drive progress via prototyping smaller projects with clear wins to build momentum.

Data Governance: Putting It Together

Obstacles and challenges related to organizational culture, decision-making culture and staffing limitations are also factors that have high potential to take your data governance program off the rails.

The levers for managing these issues are the program objectives, decision-making bodies and decision rights outlined in this part of the framework. These are the planning tools that enable data governance to be implemented in a way that fits the culture and staffing. Taken together, the program objectives, guiding principles, and the roles and responsibilities make up the data governance charter that needs approval by senior leadership as part of the initial launch.

If key resources are overloaded, there must be a clear set of stakeholders, key performance indicators (KPIs) and ideally some measure of ROI to obtain funding for resources needed to launch and sustain governance. Also, it must be clearly delineated which activities will be done and by whom. How is this accomplished? It is all about planning today based on a future vision of a mature data management process and developing a road map to get there.

Program Objectives

Like any enterprise program, data governance needs to have identified objectives (again, aligned to corporate objectives) that can be used to measure against. These are large-scale efforts to modify/improve key business processes, and since those processes will both consume and deliver data to others, it is critical to have policy guidance that defines the linkage back to data governance. Potential linkages are:

- Including data stewards in planning and work teams.
- Identifying data risks and mitigation steps.
- Setting standards for metadata capture for both programs and applications.

Guiding Principles

C-level executives often refer to corporate strategy and business drivers when determining which initiatives to fund. Participants can refer back to their data governance guiding principles when a difficult question on program direction comes up. These principles illustrate how data governance supports the company's culture, structure and business goals. Some guiding principles include:

- Data will be managed as a shared asset to maximize business value and reduce risk.
- Data governance policies and decisions will be clearly communicated and transparent.
- The data governance program will be scaled based on the size of the business unit.

Decision-Making Bodies

A common component of successful data governance is getting the right business stakeholders involved in decisions about data and how it is managed. The decisions should reflect the needs of both the individual business units and the enterprise.

Many organizations create a data governance council, which usually includes leaders from business and IT. The data governance operating procedures created by this council can facilitate data-related decisions while balancing the needs of the business units.

Data governance constituents include:

• Enterprise data governance office.

• Steering committee.

• Data governance council.

• Data steward team.

> Chief data steward.
> Business data stewards.
> Technical data stewards or data custodians.
> Working groups.

• Architecture team.

• Data requirements manager.

• Metadata manager.

• Data quality manager.

• Security and access manager.

• Business constituents.

Many of these roles exist in some form. Effective data governance requires these individuals and groups to become entrenched in the decision-making process around data rules and processes. Once the data governance role is part of a people's jobs, they are more likely to make better decisions about the role of data – and how it applies to the corporate mission.

Decision Rights

After designating the decision-making bodies, the next step is to define roles for the identified data governance activities. A good tool for this is the RACI approach.

RACI stands for R = responsible (does the work);
A = accountable (ensures work is done/approved);
C = consulted (provides input); and I = informed (notified, not active participant). Identifying a person's position on the RACI continuum helps figure out who's doing what – and how.

As an example, you could use RACI for any of the following activities:

• Approve policies and procedures.

• Develop policies and procedures.

• Monitor compliance.

• Identify data issues and proposed remediation.

• Establish data quality service-level agreements (SLAs).

Data Stewardship

The definition of stewardship is "an ethic that embodies the responsible planning and management of resources." In the realm of data management, data stewards are the keepers of the data throughout the organization. A data steward serves as the conduit between data governance policymaking bodies, like a data governance council, and the data management activity that implements data policies.

Data stewards take direction from a data governance council and are responsible for reconciling conflicting definitions, defining valid value domains, reporting on quality metrics, and determining usage de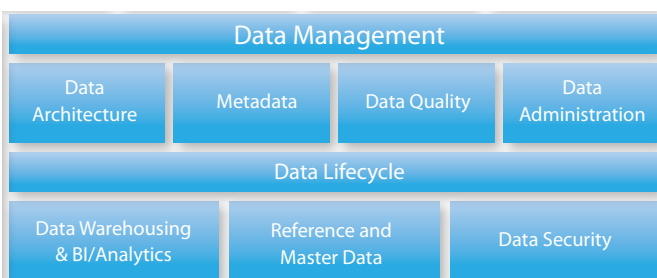tails for other business organizations. Organizationally, data stewards generally sit on the business side, but they have the ability to speak the language of IT.
(A related, but more technical role, is the data custodian who sits on the IT side. Data custodians work with data stewards to make sure applications enforce data quality or security policies – and create data monitoring capabilities that are fit for the purpose.)

Data stewards can be organized in a number of ways: by business unit, top-level data domain (such as customer, product, etc.), function, system, business process or project. The key to success in any data stewardship organization is granting authority to the stewards to oversee data (within their domain). Without this authority, you lose a linchpin between business and IT – and the entire governance apparatus can fall into disarray.

Data Management

Data management is the set of functions designed to implement the policies created by data governance. These functions have both business and IT components, so it is vital that the overall program be designed holistically. Data management functions include data quality, metadata, architecture, administration, data warehousing and analytics, reference data, master data management and other factors.

Data Management			
Data Architecture	Metadata	Data Quality	Data Administration
Data Lifecycle			
Data Warehousing & BI/Analytics	Reference and Master Data	Data Security	

Data Stewardship Roles

Data stewards are the go-to data experts within their respective organizations, serving as the points of contact for data definitions, usability, questions and access requests. A good data steward will focus on:

- Creating clear and unambiguous definitions of data.

- Defining a range of acceptable values, such as data type and length.

- Enforcing the policies set by a data governance council or any other oversight board.

- Monitoring data quality and starting root-cause investigations when problems arise.

- Participating in the definition and revision of data policy.

- Understanding the usage of data in the business units.

- Reporting metrics and issues to the data governance council.

While data management is fairly broad, not all of these disciplines must be included in the first phases of a governance program. Some programs focus more on business definitions (metadata) initially, while others may emphasize a single view of the customer (master data). Here's how a data governance strategy affects your data management program:

Data Architecture

Data architecture encompasses the conceptual, logical and physical models that define a data environment. Standards, rules and policies delineate how data is captured and stored, integrated, processed and consumed throughout the enterprise. A comprehensive data architecture defines the people, processes and technology used in the management of data throughout the life cycle.

The standardization of policies and procedures in the data architecture prevents duplication of effort and reduces complexity caused by multivariate implementations of similar operations. Examples of data architecture artifacts include entity-relationship diagrams, data flows, policy documents and system architecture diagrams.

Metadata

Metadata management includes maintaining information about enterprise data such as its description, lineage, usage, relationships and ownership. There are three distinct types of metadata:

- Business. The functional definition of data elements and entities and their relationships.
- Technical. The physical implementation of business data definitions in database systems and the rules applied in moving this data from system to system.
- Operational or process. The record of data creation and movement within the architecture.

Effective data governance requires a way to capture, manage and publish metadata information. A metadata management system provides a business glossary, lineage traceability and reusable information for business and data analysis. An automated technology far outperforms documents and spreadsheets – the traditional form of metadata management – because it's almost impossible to reuse definitions or trace lineage across a variety of shared documents.

Data Quality

Data quality includes standards and procedures on the quality of data and how it is monitored, cleansed and enriched. Traditional data quality includes standardization, address validation and geocoding, among other efforts.

In a data governance program, automated tools cleanse and enrich data in both batch and real-time modes. Data quality technology is used in a standalone fashion and integrated with transactional systems for ultimate flexibility. The definition of rules for data quality and data integrity should be managed in the business realm, but the actual execution of these rules should be managed by the IT group.

Data Administration

Data administration includes setting standards, policies and procedures for managing day-to-day operations within the data architecture, including batch schedules and windows, monitoring procedures, notifications and archival/disposal.

In a data governance program, the IT organization is primarily responsible for setting and managing these policies and procedures, consulting with the business for reasonability. The data administration process can also include SLAs for performance.

Data Warehousing, Business Intelligence and Analytics

Data warehousing, business intelligence (BI) and analytics have evolved into a separate data management system. Unlike transactional systems, these initiatives give business units a way to process vast amounts of information and perform more advanced analytics. With a data warehouse feeding BI and analytics efforts, you can get more insight from past events and forecast future events, providing better insight for more effective management decisions and strategy.

Tools and techniques for this area include data movement tools, including ETL (extract, transform and load); ELT (extract, load and transform); data federation methods; sophisticated security; and data reporting and visualization tools. With data governance in place, systems have the right data available to perform more accurate analysis – and get more value from BI and analytics programs.

Master Data

Some data elements, like customer or product, are vital to the operation and analysis of any business – and are common across most internal and external systems. This group of data elements is referred to as master data. Over the years, master data management (MDM) has been an attempt to integrate this data across the enterprise, with varying degrees of success – and price tags.

Today's MDM approach focuses on consolidating, matching and standardizing this data across transactional systems. With this pool of master data, you get higher quality information and coordinated data across all constituent systems.

MDM is inherently a very data governance-dependent effort. The first steps of creating a single view of the customer, for example, require different parts of the business to agree on the definition of the word "customer." MDM also requires IT and business to understand how the customer is represented across the systems – and what elements to view as common master data. Starting an MDM approach without data governance is a common reason why initial MDM investments fail to perform for many organizations.

Reference Data

Reference data is used by all transactional and business intelligence systems to provide lookup values to applications for storage efficiency. This data should be managed to ensure consistency across all platforms, utilizing a specialized reference data management system, an MDM system or external linked data sources – or a combination of these methods.

Data governance provides the ability to create better reference data, as business groups come to agreement on the terms that support business activities. The data governance council, working with a network of data stewards, provides a way to catalog and capture those terms and assign ownership. Once a common set of state codes is used, for example, reports that previously used different versions of state codes (which may change across departments) will now yield more aligned and accurate results.

Data Security

Data security includes policies and procedures to determine the level of access allowed for both source-level data and analytics products within the organization. The best practice for applying these policies is to develop a role-based model where access rights are granted to roles and groups, and individuals are assigned to one or more roles or groups. Special care should be taken for regulatory requirements and privacy concerns. Ensuring the right people have access to the right data is key to effective governance.

Data security documentation includes policies, procedures, RACI charts, matrices and other relevant information. The business organization is responsible for defining the levels of access required for various roles and groups, while the IT and security organizations are responsible for implementing the requirements in the system architecture.

Data Life Cycle

Data should be managed from the point it enters your organization until it is archived – or disposed of when it is no longer useful. The business side of the organization is responsible for defining the sources, movement, standardization/enrichment, uses and archival/disposal requirements for the various types of data used by the enterprise. The IT organization is responsible for implementing these requirements.

Methods

| People |
| Process |
| Technology |

The execution of a data governance program involves three main areas: people, process and technology. All three are necessary to properly execute the charter developed by the data governance council.

People

An organized structure of people who have the proper skills is essential for the success of the data governance program. This structure is conceived and documented as part of the creation of the data governance charter:

- *The data governance council* contains senior staff familiar with both the operations and strategic direction of the organization. They determine the high-level policies of the program and approve the procedures developed to carry out those policies.

- *Stakeholders* are members of the business and IT management teams who have a direct connection to the program. So, they are the most invested in the success of the program. They provide feedback to the council and get regular updates on the progress of the program.

- *Stewards* are the subject-matter experts responsible for executing the policies enacted by the data governance council. They are responsible for the quality of the data in the organization, helping maximize its value.

- *Data producers or consumers* either create data through an application or use data to drive decisions as part of a business process.

These groups need the authority to create policies and procedures that drive the program's success. In return, they are directly accountable for them. It's also important that these duties are not viewed as secondary, but as an integral part of their job descriptions.

Process

The second major area of program execution is process. These policies are unique to each enterprise, but there are certain themes that are common to all successful programs:

- *Measurement.* Measurement allows organizations to maintain control over data governance processes. Measurement also demonstrates the program's effectiveness to both users and management, since these programs are often viewed as an overhead activity not directly related to profit generation. A data governance program is a program of continuous improvement, so effective measurement is a basic component of any successful program.

- *Communication.* Complete, concise communication is vital to the success of any large program, and data governance is no exception. Any communications framework should focus on reusability, broad acceptance and effectiveness. Along with this, effective training on all facets of the communication infrastructure can help integrate a strong communications effort throughout the program.

Technology

Modern organizations are complex entities, and their data architecture reflects this. While it's possible to administer a data governance program using documents, spreadsheets and database-embedded data quality validation routines, this method is labor-intensive and difficult to manage.

Current best practices include automated tools to perform tasks such as standardization, data quality, data profiling and monitoring. This allows the organization to scale to the largest volumes of data processing, maintaining the same rules and processes for any application.

Solutions

Solutions				
Data Quality	Data Integration	Data Preparation	Reference Data Management	Master Data
Data Profiling & Exploration	Data Visualization	Data Monitoring	Metadata Management	Business Glossary

The solutions section of the framework describes the types of applications used to implement and automate the various activities of a data governance program or data management initiative. These tools include:

- **Data quality**. Establish and enforce the rules and policies to make sure that data meets the business definitions and rules established by the data governance program.

- **Data integration**. Combine data from multiple data sources to provide a more unified view of the data. This includes the movement, processing and management of that data for use by multiple systems, applications, tools and users. Data federation or data virtualization is another form of data integration where users are provided access to masked, cached and blended views of their data without regard to where it's physically located.

- **Data Preparation**. Provide self-service access to business users so they can blend, cleanse and transform data without knowing how to code or burdening IT.

- **Reference data management**. Create and maintain a set of commonly used data that can be referenced by other applications, databases or business processes. Reference data can include things like state codes or medical codes, and once completed, this data can be stored as lists of values, code tables (lists of text/value pairs) or hierarchies.

- **Master data**. Consolidate, standardize and match common data elements, like customers or products, to achieve a more consistent view of these entities across the organization.

- **Data profiling and exploration**. Analyze existing data and potential new sources of data to determine its content, helping you understand what to expose and what quality or continuity issues exist.

- **Data visualization**. Produce graphical representations of data in various forms, including dashboards and analytical structures.

- **Data monitoring**. Detect data quality issues within the data through ongoing enforcement of rules.

- **Metadata management**. Acquire and store information about data, helping describe details about application data assets throughout the organization and the lineage of that data.

- **Business glossary**. Build a repository of definitions and business rules for data systems from a business perspective.

Summary

As organizations continue to become more data-driven, their success will ultimately hinge on the ability to maintain and use a coherent view of this data. Better data – and a clearer view of what this data means – can drive insight and help you make better decisions every day.

The components of the SAS Data Governance framework provide a comprehensive structure for enterprises of all sizes to establish and sustain an effective data governance program. With these elements, you can provide trusted, timely, high-quality data to all users in the organization. And create the data that powers a more effective, efficient and responsive organization.

Learn More

Find out more about SAS Data Governance by visiting sas.com/datagovernance.