# Table of Contents

- **Introduction** .................................................. 1
- **Data Governance – Use Case and Challenges** ............ 1
- **Collaboration – the Key to Success** .......................... 3
- **Managing Reference Data Across the Enterprise** ........... 7
- **Conclusion** .................................................. 12
- **Recommended Reading** ........................................ 13
- **Contact Information** .......................................... 13
- **For More Information** ........................................ 13
**Introduction**

In recent years, the concept and discipline of data governance has grown in importance as organizations are forced to comply with industry or governmental regulations, cut costs to improve margins and use data-driven initiatives to increase revenue. As a result, there is often an executive-level mandate to create a data governance center of excellence (COE) that includes the people (line-of-business managers, data stewards, information technology developers and so on), defined business processes and technology that are required to govern data at an enterprise level.

The goal of a data governance COE can vary. In many cases, the goal is to provide better visibility into a corporation’s data assets to drive better and quicker business decisions. Or it can be to comply with regulatory requirements such as Sarbanes-Oxley, HIPAA or Solvency II. Or it can be to simply improve the efficiency and operations of data management at an enterprise level.

Although the goal of data governance initiatives is often easy to define and understand, organizations struggle to implement the programs. The inability to collaborate or share the business requirements with technology providers, the lack of controls for managing reference data across an enterprise, and, ultimately, the difficulty in measuring the success of these programs often lead to a lack of ongoing support and commitment at an organizational level.

This paper explores the challenges organizations have today in implementing a data governance program via an actual business case. It highlights SAS technology that can help you solve many of those challenges.

**Data Governance – Use Case and Challenges**

The retail industry provides an excellent use case for exploring data governance best practices. In the retail industry, a common driver for data governance programs is to enable better integration of new or updated data, which enables more strategic and real-time marketing campaigns. In many cases, often due to mergers and acquisitions, retailers are forced to consolidate loyalty card programs, brands and sales territories. The inability to implement these changes efficiently can prevent long-term value from being realized from merger and acquisition activity.

This paper explores a fictitious regional retail sporting goods chain (which is based on real-world customer scenarios) named D&G Sporting Goods. D&G is looking to expand its footprint from the Midwestern US to the East Coast with the acquisition of a regional golf equipment and instructional retailer. As part of this acquisition, D&G is committed to implementing a data governance program to:

- Improve customer experience via improved searches and social media integration.
- Increase efficiency when bringing new data sources into the supply and marketing chain.
- Establish enterprise-level data ownership and stewardship programs to drive better analytics to regional chains.
As part of the data governance program, D&G created a data governance COE team, chaired by its chief intelligence officer (CIO), and included membership from the line-of-business managers (golf, camping, exercise and others), the vice president of analytics and business intelligence, directors of information technology, and representation from business managers in the various regional stores.

The initial task of the COE team was to survey the team members to determine existing issues that could serve as a foundation for requirements for the data governance program. Here are the key issues that the members of the COE team highlighted:

- Point-of-sale (POS) dashboards are great, but data quality is still the Achilles’ heel.
- Communication of changes to data (hierarchies and corrections) aren’t handled uniformly and, therefore, can’t be deployed quickly and efficiently.
- The company needs standard methods to detect, assess, fix and communicate multiple types of data quality issues. It needs to be able to easily integrate new data sources.
- Data ownership needs to be addressed across brands and business areas.
- New strategic and marketing initiatives, such as golf training centers, need new sources of clean and reliable data.

Traditionally, each of these challenges might have been managed by a particular line of business or by a specific application used in the company. For example, data quality practices might already exist in the customer relationship management (CRM) system, but they are not consistent across enterprise-resource planning, digital marketing or accounting systems. Furthermore, determining the ownership of data and the rules for governing it can be complicated, as various departments in an organization might view their needs independently.

Most IT and business executives view the principles of data governance applied across their enterprise as critical to their success, yet few data governance implementations span more than a single department in scope. Data governance programs are often driven by individual groups (or only the IT organization) instead of collaborative groups, thereby limiting the programs’ scope and effectiveness.

As a result, organizations are now looking for technology solutions that:

- Enable better collaboration between business areas and IT.
- Manage reference data across the enterprise.
- Provide capabilities to measure and monitor the ongoing success of enterprise data governance programs.
Collaboration – the Key to Success

Although data governance programs are typically composed of cross-functional teams whose members come from various organizations in an enterprise, there are several key roles that play an important part of any data governance initiative. These roles are:

The **business data steward** is the primary touch point for all data issues in a subject area, and accountable for quality and usage of data within his or her subject area. This person’s responsibility is to:

- Define data quality metrics and thresholds for the subject area.
- Ensure compliance with governance policies and processes within the subject area.
- Identify business metadata to be collected for the subject area.
- Oversee appropriate business use of data in the subject area.
- Create data audit guidelines for data updates and new data sources.
- Work with the data architect to define data relationships.

The **data architect** defines, models, designs and maintains the data based on business and data requirements. This person’s responsibility is to:

- Define source data extraction standards.
- Provide data modeling expertise.
- Create, maintain and support enforcement of data modeling and naming standards.
- Maintain reference data architecture.

The **data quality lead** ensures that data conforms to business requirements and maintains the processes and automation necessary for data correction. This person’s responsibility is to:

- Perform root cause and source-data error analysis.
- Perform production data quality monitoring and data remediation.
- Design data quality improvement projects.
- Recommend data quality threshold-level changes.
- Run regular quality inspections of data and create data quality improvement projects for data not conforming to established standards.

There will likely be additional key individuals in any successful data governance program, but the roles they assume are typically a mixture of those above – although they might come from different areas of the business. Because data governance requires a number of users and teams to implement the initiative, collaboration between the various team members becomes critical to overall success. To meet this need, many organizations use internal wikis, data-modeling tools, or various office and document management products to exchange information with members.
Although these tools and products provide a good starting point, they give minimal direction. They are often hard to manage as the information needs of the team increase. What is really needed is for each of the users to be able to create and share information with each other in a way that is specifically customized toward data governance best practices. For example, the business data steward wants to define business terminology, requirements and other details in business terms. A data architect wants to view the information that the business data steward provides to understand and implement the rules required by the actual data. Creating rules that can be reused and have stored metadata is a bonus for the data architect, because it allows similar rules to be built and applied across multiple physical data stores. It provides for lineage and impact analysis so that if future changes are made, it is clear which physical systems and downstream reports need to be updated. Finally, the data quality lead needs to view both the business terminology and the rules created by the data architect to be able to know how to interpret and fix data quality issues when they occur.

Figure 1 illustrates the type of information that users in each role create and need to understand to be successful in a data governance initiative.

The business data network (BDN) is a Web-based data governance application designed specifically to support enterprise data governance initiatives and meet the needs of users who work in the roles defined above. It provides a collaborative environment for collecting, documenting and sharing business, operational and technical information. The BDN enables business user, data architect and business data steward collaboration. Business users can use the BDN to define and document information in business terms; business data stewards can leverage the business terms to ensure data alignment; and data architects can implement jobs and rules to manage and master the data.
The BDN also enables users to define a repository of business terms and their associated attributes and relationships. Business terms include key information, such as:

- Name and description.
- Source systems.
- Owner (IT and business).
- Related processes (data quality services, data workflows and applications).

The example in Figure 2 shows how D&G Sporting Goods could use the BDN to design, manage and deploy its data governance practice. D&G can create terms such as CUSTOMER, SUPPLIER, ACCOUNT and others that describe the data it uses to manage its business.

Figure 2: Business data network can be used to design, store and deploy governance practices

Figure 3 illustrates an example term representing customer information. The CUSTOMER term contains an example of how the term “customer” might be represented in the BDN. It includes information about requirements needed to create a complete customer record, pointers to documentation such as data models that store customer information, rules that describe what the customer record should look like, physical source systems that contain customer data, related terms that leverage customer information, the importance of this term to the overall system and other details.
Related objects – such as physical tables and source systems that store customer information, sales reports, analysis, models and other relevant customer data – can be linked to the term. Collaboration between who is responsible for the term and who consumes the data is supported on the Notes and Contacts tab. Notification of changes or updates to the term can be sent via email to all interested parties. Full versioning and history of changes to the CUSTOMER term is supported and displayed on the History tab. Some of these features are illustrated in Figure 4.

Content to populate the business data network can be imported from a variety of sources. The import format is XML-based, and the client is shipped with a number of examples.
Managing Reference Data Across the Enterprise

Most data governance initiatives require organizations to manage reference data at an enterprise level. For D&G Sporting Goods, this requirement is critical because of its recent acquisition. Customer, sales and human resource data from the acquired company must quickly be reconciled with existing systems. Providing a consistent view of the customer for the loyalty programs, sales territories for sales operations and changes to sales hierarchies is key when it comes to taking advantage of the revenue growth promised by the acquisition.

Having the ability to centrally manage this common data and provide interfaces for the various lines of business to import, update and publish this reconciled data to operational systems is a major component to any data governance program. It enables faster integration of new data sources and provides a consistent communication and publication mechanism for the entire organization.

The Reference Data Manager is a Web-based stewardship client offered as part of the overall SAS Data Management suite. It can help you centrally manage, version and publish reference data across an organization. It manages lists, lookup tables and hierarchies, and can be published to relational database tables, flat files, XML documents and other formats. It is tightly integrated with the business data network, so a business term such as SALES TERRITORY might be related to a hierarchy managed in Reference Data Manager (see Figure 5, below).

Figure 5: The Reference Data Manager can help centrally manage, version and publish reference data across the organization
Data Governance Monitoring, Remediation and Lineage

One of the ongoing criticisms of data governance as a discipline is the struggle organizations have in determining the overall value and ongoing effectiveness of the program. In some cases, this is because of the inability to capture the key performance indicators (KPIs, also called dimensions), which are the measurements of the overall success of the program. More commonly, the issue is that the KPIs for measuring success are documented but aren’t accessible or easily viewed in a dashboard that can show the state of overall quality of the data managed by the data governance program.

The BDN has integrated monitoring and dashboard capabilities that enable the business data steward to monitor the KPIs associated with the data governance program. The dashboard can be customized to help identify and remediate data issues with visuals describing the status of quality problems and how they trend over time. The dashboard also enables you to drill down to get more detail about specific KPI dimensions, such as accuracy and integrity. Dimensions and thresholds are fully user-configurable. From ongoing, integrated monitoring you can better react to potential errors and fix problems more quickly. You can drill down to see specific problem records. An example of the dashboard is shown in Figure 6.

Figure 6: Integrated monitoring with a highly visual dashboard that is drillable to specific problem records
Specific data errors in problem records can be interrogated from the dashboard, using the data monitor as shown in Figure 7.

Figure 7: The dashboard can drill to specific error details as part of integrated data monitoring

In the case of D&G Sporting Goods, one of the charters of the data governance program is to improve the quality of information that is feeding the POS dashboards. The ideal method provides a common and consistent application for subject matter experts to use to assess, remediate and fix data quality issues not only for the POS dashboards, but for other operational systems as well.

SAS Data Management includes a data remediation that is designed for subject matter experts to address data quality issues. Invalid data records (or sets of records) can be programmatically directed to a remediation queue. Once in the queue, the remediation interface supports the corresponding workflow. You are notified when something needs your attention. The remediation interface supports alerts and notifications, status tracking, filtering and workflow queue management. An example of the remediation queue is shown in Figure 8.
Figure 8: Data remediation enables subject matter experts to be notified and address data quality issues in a managed workflow.

From the remediation queue, you can view the status of individual records (or sets of records). You can see what the data problems are, view individual records, assign others to work on the issue, assign a priority to the issue and correct the data. All these activities are managed based on individual authorizations, which an administrator manages and controls. You have the ability to notify others and collaborate on tasks. And an overview of all activities assigned to an individual describes the latest status of activity within the remediation queue, as illustrated in Figure 9.
Another key deliverable of data governance is to audit and track lineage of business and technical metadata throughout the enterprise. This is important for heavily regulated industries where an auditor must be able to track the entire data life cycle, including decision makers and subject matter experts who have ownership of the data. In the case of D&G Sporting Goods, it is important to determine how data that is feeding the enterprise data warehouse could affect business analytics and decision service programs for loyalty card programs.

The business data network also includes a lineage and impact analysis designed to assess interrelationships among data elements. It displays how all the components of the D&G enterprise fit together. Lineage shows data dependencies, related terms, physical objects associated to items, source systems, generated reports built from the terms and additionally depicts relationships to third party objects such as Excel spreadsheets or documentation. An example is shown in Figure 10.
Conclusion

Data governance combines the disciplines of data quality, data management, data policy management, business process management and risk management into a methodology that helps manage critical data assets throughout an enterprise. SAS offers a fully integrated suite of technologies that ensure efficient implementation and effective data governance initiatives. This paper described several key considerations and how they can be addressed with best practices. Business data networks, reference data management, data federation, quality and master data management are all aspects of an integrated data governance. Having a single integrated, configurable software platform to address all of these key considerations reduces total cost of ownership and breeds timely delivery of value to the organization. SAS Data Management is a cohesive suite of capabilities that help IT organizations deliver on the promise of data governance.
Recommended Reading


Contact Information

Your comments and questions are valued and encouraged. Contact the authors at:

Scott Gidley
SAS Institute Inc.
Work Phone: 919-677-8000
Scott.Gidley@sas.com

Nancy Rausch
SAS Institute Inc.
Work Phone: 919-677-8000
Nancy.Rausch@sas.com
About SAS

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 65,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®. For more information on SAS® Business Analytics software and services, visit sas.com.