Paper 283-2014

# Managing the Data Governance Life Cycle

Scott Gidley, SAS Institute Inc., Cary, NC

## ABSTRACT

Data governance combines the disciplines of data quality, data management, data policy management, business process management, and risk management into a methodology that ensures that important data assets are formally managed throughout an enterprise. SAS® has developed a cohesive suite of technologies that can be used to implement efficient and effective data governance initiatives, thereby improving an enterprise's overall data management efficiency. This paper discusses data governance use cases and challenges and provides an example of how to manage the data governance life cycle to ensure success.

## INTRODUCTION

In recent years, the concept and discipline of data governance have grown in importance as organizations are forced to comply with industry or governmental regulations, cut costs to improve margins, or 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 can be 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 data assets 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 assimilation 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 the merger and acquisition activity.

This paper explores a fictitious regional retail sporting goods chain named D&G Sporting Goods. D&G is looking to expand its footprint from the midwestern United States to the East Coast via 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:

- The lack of a data registry that allows business and IT to map, understand, and create terms for describing data across the enterprise slows their ability to assimilate new data.

- Communication of changes to data (new sources, changes to delivery styles and types, etc.) aren't handled uniformly and therefore can't be deployed quickly and efficiently.
- The company needs standard methods to track and manage the relationships of data and data processes to understand the enterprise impact of any proposed changes.
- Data ownership and role-based approval processes need to be established to enable assimilating new or sensitive customer data.

Traditionally, each of these challenges might have been managed by a particular line of business or by a specific application. Data glossary technology might be used to model a data registry, but it lacks approval and workflow processes that enable collaboration with the business.

## COLLABORATION—THE KEY TO SUCCESS

Although data governance programs are typically comprised 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:

**Data Governance Council:** this entity is the primary decision-making authority for a data governance program.

- Approve changes to crosscutting processes or business terms.
- Manage security, roles, and access.

**Business Data Steward:** this person is the primary touchpoint for all data issues in a subject area. This person is accountable for quality and usage of data within his or her subject area. Primary responsibilities include:

- Define data quality metrics and thresholds for the subject area.
- Ensure compliance to 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.

**Data Architect:** this person is responsible for the definition, modeling, design, and maintenance of data based on business and data requirements. Primary responsibilities include:

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

**Data Quality Lead:** this person ensures that data conforms to business requirements and maintains the processes and automation necessary for data correction. Primary responsibilities include:

- Perform root cause and source data error analysis.
- Perform production data quality monitoring and data remediation functions.
- 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 can be additional key individuals in any successful data governance program, but the roles that they assume are typically a mixture of the above roles, 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 the overall success of the team.

The SAS Business Data Network (BDN) is a web-based data governance application designed specifically to support enterprise data governance initiatives and to meet the needs of the users who work in these roles. It provides a collaborative environment for collecting, documenting, and sharing business, operational, and technical information.

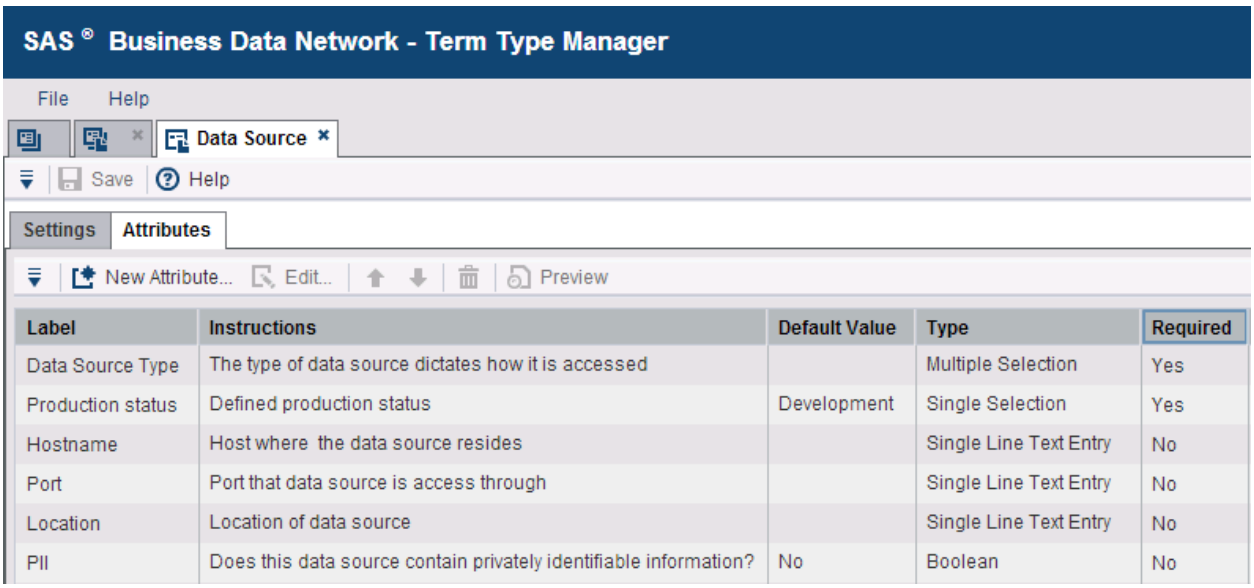
The following sections describe how D&G would use the BDN to govern the interactions, collaborations, and approvals required to assimilate data from a newly acquired company. Specific areas of focus include:

- Template-driven business terms that best describe the data and terms that are new to D&G processes.
- The use of workflow and business processes that ensure that new data elements are properly defined and approved based on the D&G data governance charter.
- Defining and visualizing relationships with existing D&G data elements, data quality tasks, and data integration jobs.
- Enabling security so that changes to any existing terms or processes are properly managed and secured.

## TEMPLATE-DRIVEN BUSINESS TERMS

At the beginning of any data governance initiative, the participants of the project must agree on what information and attributes should be captured to properly govern the required data. These attributes must be customizable because they might be different based on the type of data that you want to govern. For example, attributes for a data source might include legal notes on the usage, how often the data source is updated, how to access the data source, and who can access it. Attributes for a data element (such as a name or location) might include a link to the data source that it comes from, a data dictionary, an example of usage, and any data validations that are required to ensure accuracy.

The BDN provides *term templates*, which can be used by the data governance council to define standard templates for the various types of data elements that it wants to govern. Once a template is defined, it can be reused across the enterprise.




| Label | Instructions | Default Value | Type | Required |
|---|---|---|---|---|
| Data Source Type | The type of data source dictates how it is accessed | | Multiple Selection | Yes |
| Production status | Defined production status | Development | Single Selection | Yes |
| Hostname | Host where the data source resides | | Single Line Text Entry | No |
| Port | Port that data source is access through | | Single Line Text Entry | No |
| Location | Location of data source | | Single Line Text Entry | No |
| PII | Does this data source contain privately identifiable information? | No | Boolean | No |

**Figure 1. Defining Term Template Attributes**

Figure 1 illustrates an example Data Source template that contains attributes such as location, source type (database, external file, etc.), and a PII flag that indicates whether the source contains privately identifiable information. These attribute types can be free-format text, picklists, or on-off flags. Term template management enables default ownership to be assigned for a term and for workflows to be associated to help govern term creation, updates, and deletes.

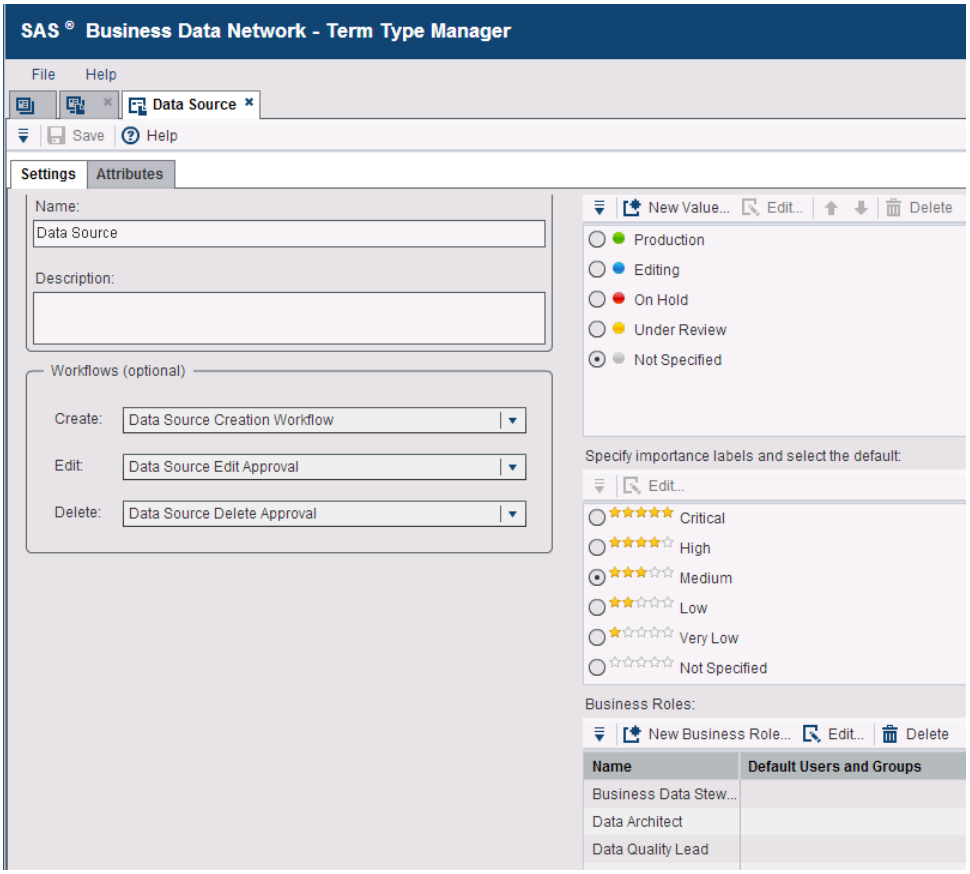



**Figure 2. Defining Term Template Attributes**

In figure 2, workflows can be assigned to the term template. These workflows help drive the life cycle of business terms within the BDN.

## BUSINESS TERM LIFE CYCLE MANAGEMENT

Once a term template has been created, a BDN user can assign that template to any new term that is created.

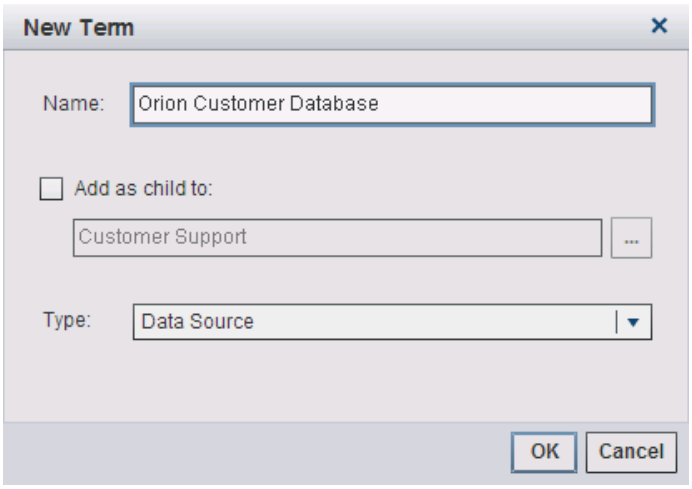



**Figure 3. Assigning a Term Template to a New Term**

Figure 3 highlights a new term that represents the data source at D&G's newly acquired company. It is assigned the

Data Source template.

Based on the term template workflow, once the data steward creates the term from the new data source from the acquired company, it is moved into a create state and directed to the data architect. The architect adds details to the term such as the database table and columns that it references.

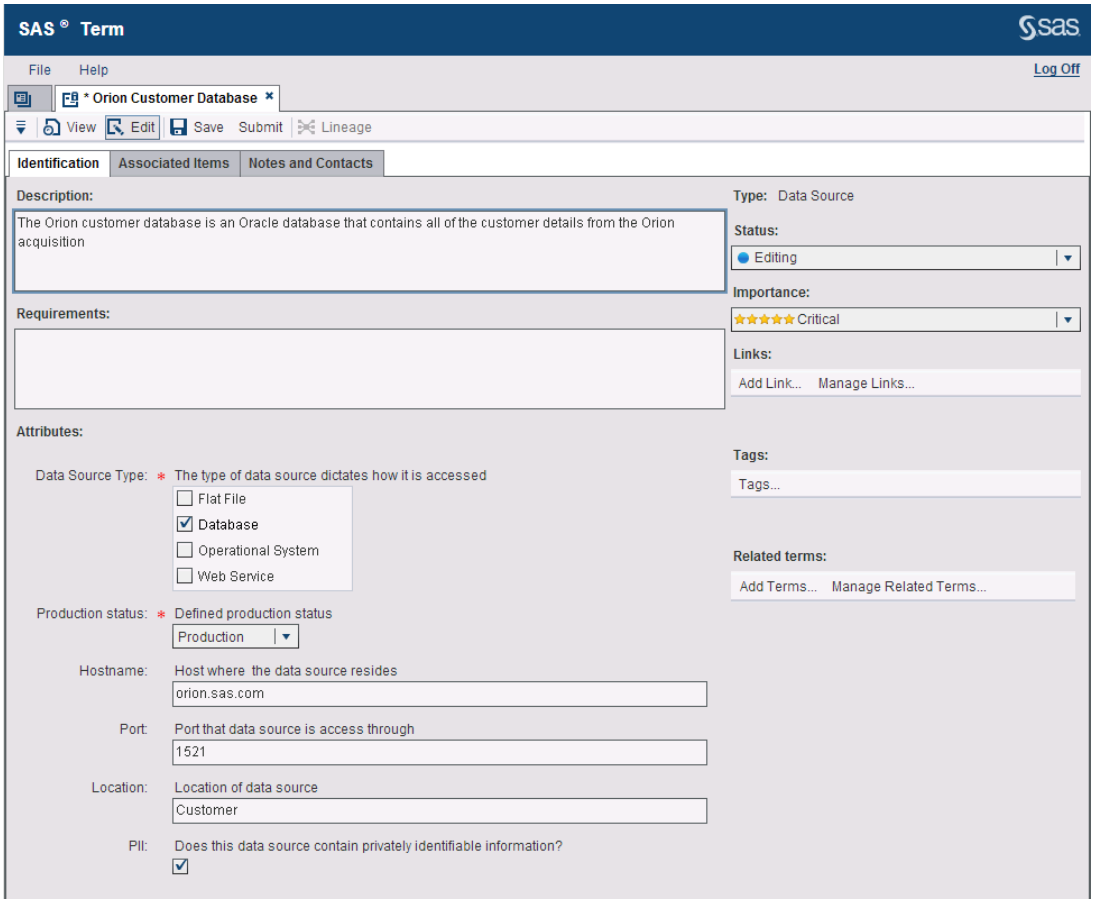



**Figure 4. Providing Detail for a New Term**

After it is saved, the term is routed to the ETL developer, who adds relationships to data quality, data validation, and data integration tasks associated with this new source of data. The completed term is sent to the data governance council for final approval. At that point, it is a production-level term.

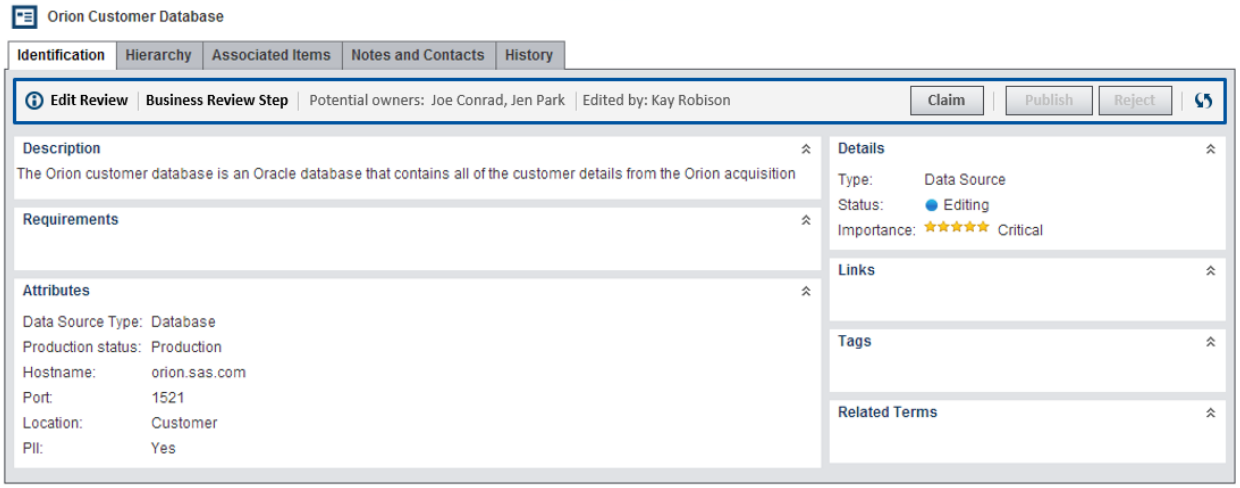


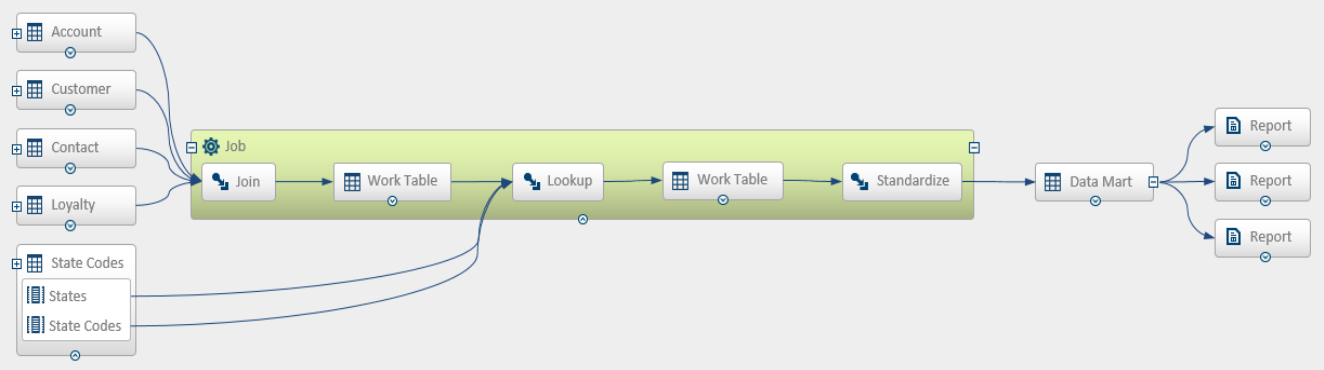

**Figure 5. Approval Process**

## RELATIONSHIP AND LINEAGE VISUALIZATION

A key aspect of any data governance deliverable is the ability to audit and track the lineage of business and technical metadata throughout an enterprise. This is critically important for heavily regulated industries in which an auditor must be able to track the entire data life cycle, including the 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 impact business analytics and decision service programs for loyalty card programs.

The BDN includes a lineage and impact analysis application that meets this need. It can display how all of the components of the D&G enterprise fit together. The application shows data dependencies, related terms, physical objects associated to items, source systems, generated reports built from the terms, and other object relationships. The application also shows relationships to third-party objects such as Excel spreadsheets or documentation.

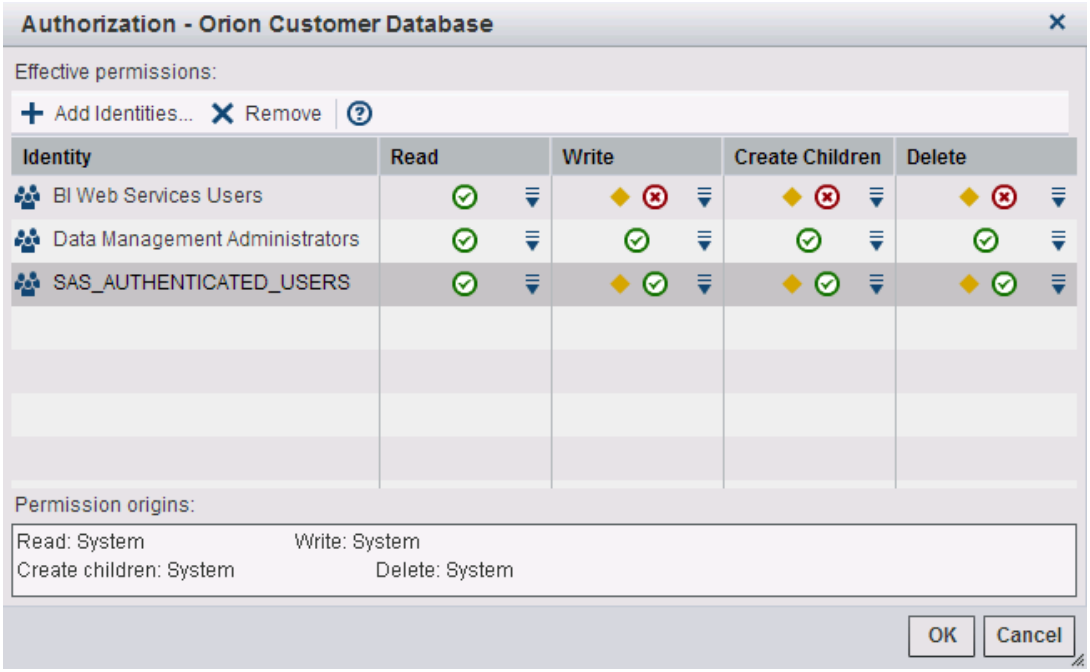
The following figure highlights the lineage view that illustrates the relationship of the customer and loyalty data from the acquired company with the existing data mart processes used by D&G.




## ROLE-BASED SECURITY

As data governance programs mature, the need to expand or contract the roles and permissions of the various program members increases. The BDN offers granular access control to terms, attributes, and relationships within the system.

In the case of D&G Sporting Goods, BI analysts needed to be able to view terms and their relationships, but they were not allowed to edit or delete terms. As the data governance program matured, it was agreed that data stewards would have full access rights to term template management.




This type of granularity and flexibility allows the BDN to meet the needs of most any secure environment.

## CONCLUSION

Data governance combines the disciplines of data quality, data management, data policy management, business process management, and risk management into a methodology that ensures important data assets are formally managed throughout an enterprise. SAS has developed a cohesive suite of technologies that can be used to implement efficient and effective data governance initiatives, thereby improving an enterprise's overall data management efficiency. This paper discusses data governance life cycle management. It explains how the SAS Business Data Network can be used to ensure that data governance initiatives remain successful, continue to deliver overall return on investment, and gain buy-in across the enterprise.

## CONTACT INFORMATION

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

Scott Gidley
SAS Institute Inc.
Cary, NC 27513
Work Phone: (919) 677-8000
Fax: (919) 677-4444
Scott.Gidley@sas.com
support.sas.com