

Paper 084-2013

Best Practices in Enterprise Data Governance

Scott Gidley and Nancy Rausch, 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 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 best practices. It explains where and how SAS capabilities (such as the Business Data Network, Reference Data Management, Data Federation, Data Quality, Data Management, and Master Data Management) 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.

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, 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 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 onboarding 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:

- 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.

Based on recent surveys, a vast majority of information technology (IT) and business executives view the principles of data governance applied across their enterprise as critical to their future success. Yet, less than 10% of data governance implementations span more than a single department in scope. More than two-thirds of data governance programs are not driven by collaborative groups. Instead, they are driven by individual groups in a company or driven by only the IT organization, thereby limiting their 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 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:

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 will likely 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. 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 provide minimal actual direction. They are often hard to manage as the information needs of the team increases. What is really needed is each of the users in the roles needs 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 so that he can understand and implement the rules required by the actual data. Creating rules that can be reused and that have stored metadata is an added bonus for the data architect because it allows similar rules to be built and applied efficiently across multiple physical data stores. It provides for lineage and impact analysis so that if changes are made in the future, 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.

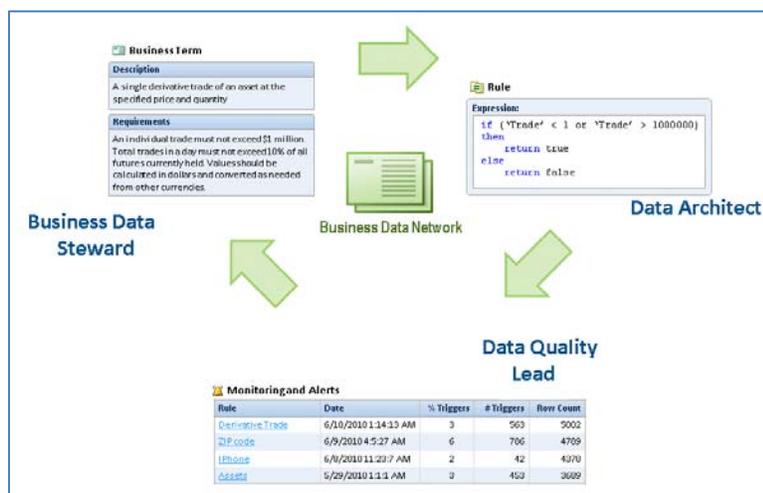


Figure 1: Collaboration among Key Data Governance Users

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 the users who work in the roles defined above. It provides a collaborative environment for collecting, documenting, and sharing business, operational, and technical information. The Business Data Network enables business user, data architect, and business data steward collaboration. Business users can use the the Business Data Network to define and document information in business terms, business data stewards can leverage the business terms provided by the business users to ensure data alignment, and data architects can implement jobs and rules to manage and master the data.

The Business Data Network enables users to define a repository of business terms and their associated attributes and relationships. The business term includes key information including:

- 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 Business Data Network 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 that it uses to manage its business.

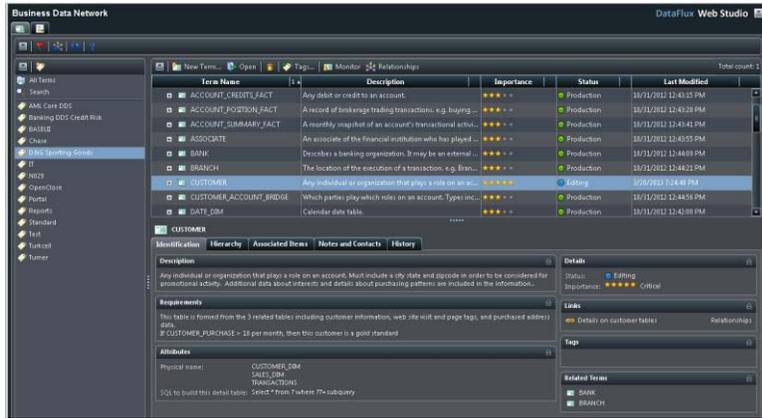


Figure 2: Business Data Network

Figure 3 illustrates an example term representing customer information. The CUSTOMER term contains information about how a customer record is stored in the D&G computer systems. 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.

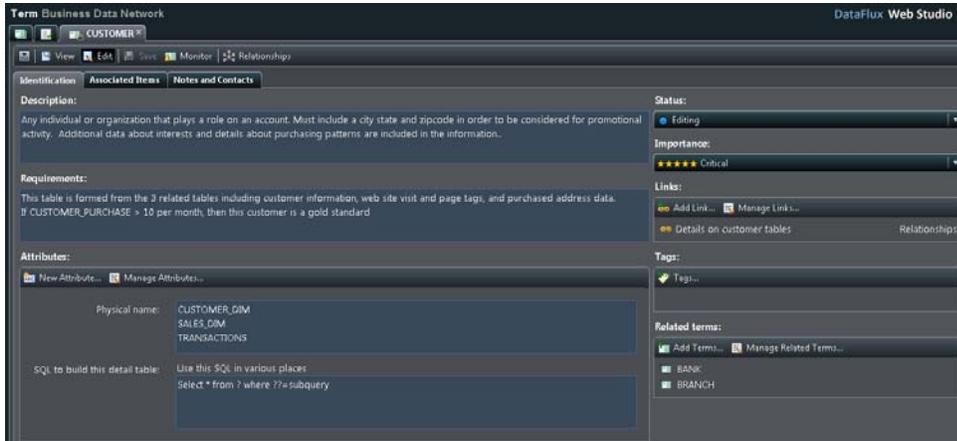


Figure 3: CUSTOMER Term in the Business Data Network

Related objects such as physical tables and source systems that store customer information, sales reports, analysis, models, and other information-related 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 e-mail 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.

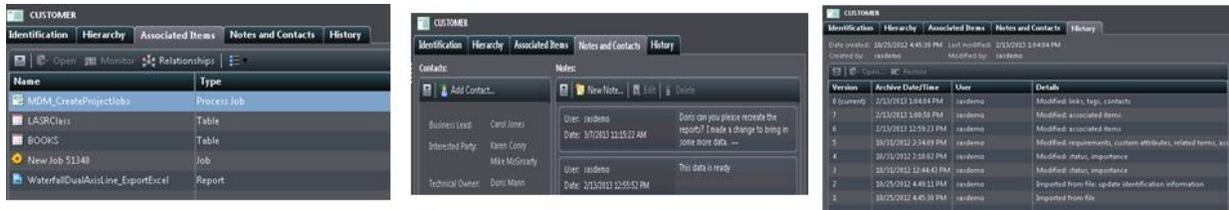


Figure 4: Additional Information Stored in the Business Data Network

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 have as a requirement to enable 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 critical to taking advantage of the revenue growth promised by the acquisition.

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

The Reference Data Manager technology is a web-based stewardship client offered as a part of the overall SAS® Data Management product 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 that a business term such as SALES TERRITORY might be related to a hierarchy managed in the Reference Data Manager. Figure 5 is an example of the Reference Data Manager.

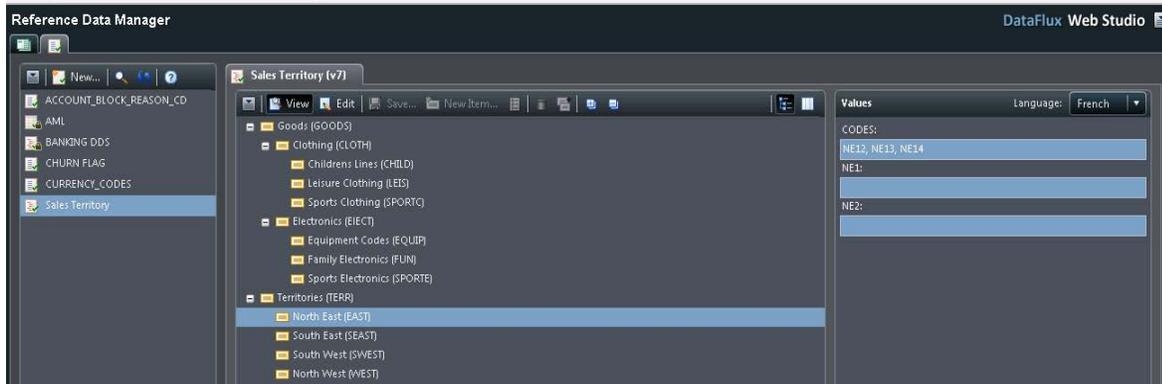


Figure 5: Reference Data Manager

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 properly 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 not able to be associated or easily viewed in a dashboard that can show the state of the overall data quality of the data in the data governance program.

The Business Data Network has integrated monitoring and dashboard capabilities that enable the business data steward to monitor the KPIs associated with a data governance program. The dashboard can be customized to help identify and remediate data issues. The data monitor and dashboard clients visualize quality problems and show how they are trending over time. The dashboard enables you to drill in to get more detail about specific KPI dimensions such as accuracy, integrity, and others. Dimensions and thresholds are fully user-configurable. Trending is displayed

so that you can see how your data is performing over time. This enables you to better react to potential errors and fix problems more quickly. You can drill in to see specific problem records. An example of the dashboard is shown in Figure 6.

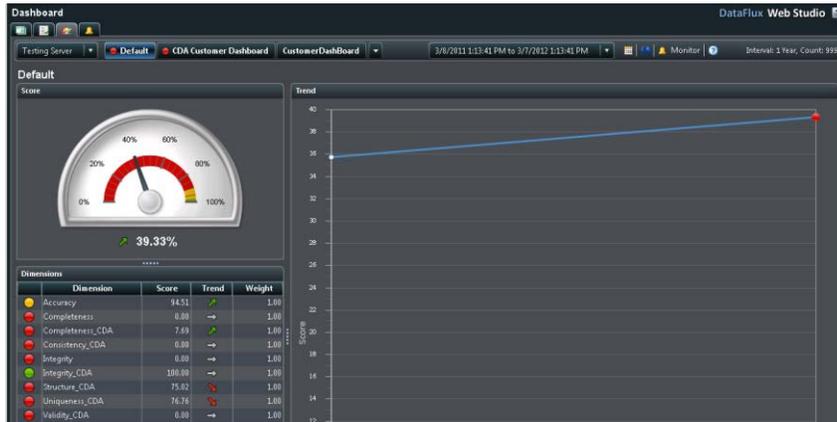


Figure 6: Dashboard

From the dashboard, you can drill in further to the data monitor application to see details about specific data errors. An example of the data monitor is shown in Figure 7.

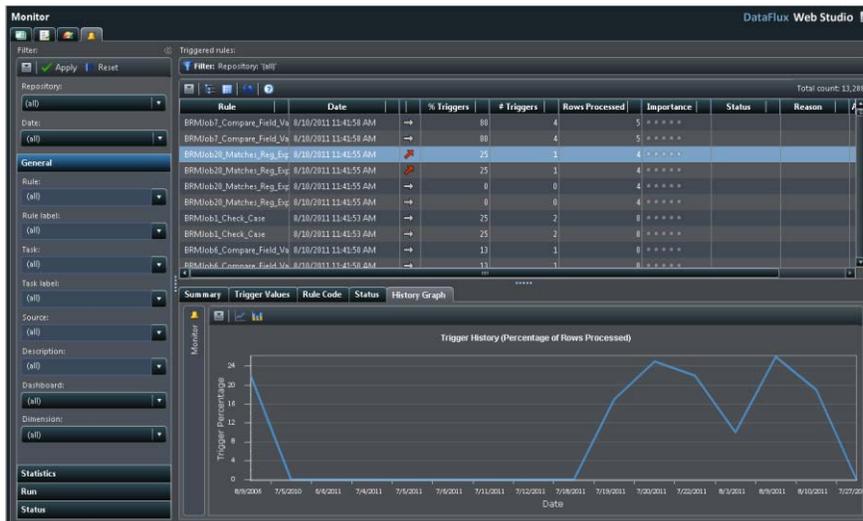


Figure 7: Data Monitor

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 would provide 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 has a new data remediation application that is designed to meet this need. Invalid data records or sets of records can be programmatically directed to a remediation queue. Once in the queue, the remediation interface supports 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 interface is shown in Figure 8.

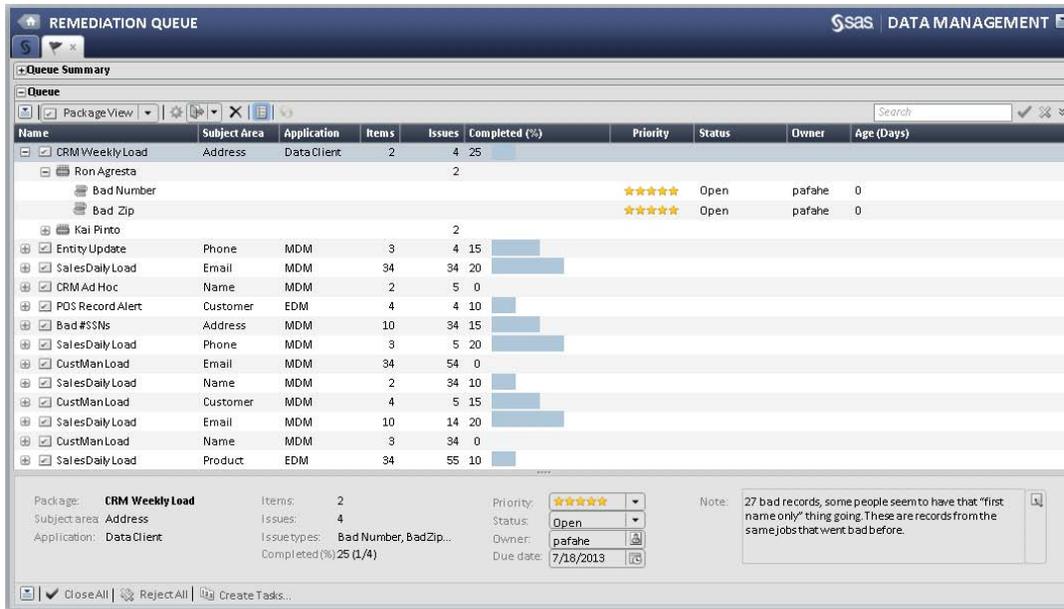


Figure 8: Data Remediation

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 (if you want to). All of these activities are managed based on your personal authorizations, and an administrator can manage data access for users. You have the ability to notify others and perform collaboration tasks. An overview window is available to show you the latest status of the remediation queue, as shown in Figure 9.

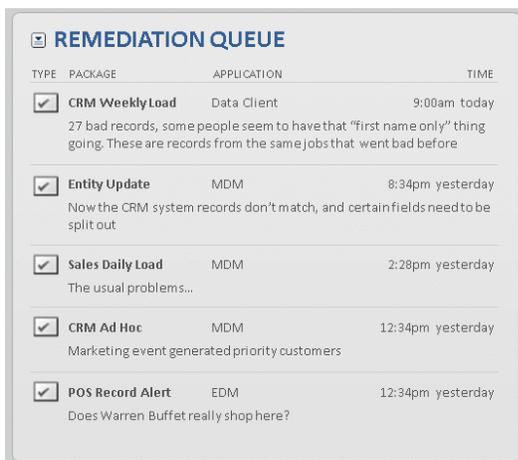


Figure 9: Remediation Queue Overview Window

Another key deliverable of a data governance initiative is the ability to audit and track lineage of business and technical metadata throughout an enterprise. This is critically important for heavily regulated industries where an auditor must be able to track the entire data lifecycle, 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 Business Data Network 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 lineage application shows data dependencies, related terms, physical objects associated to items, source systems, generated reports built from the terms, and other object relationships. The lineage application also shows relationships to third-party objects such as Excel spreadsheets or documentation. An example is shown in Figure 10.

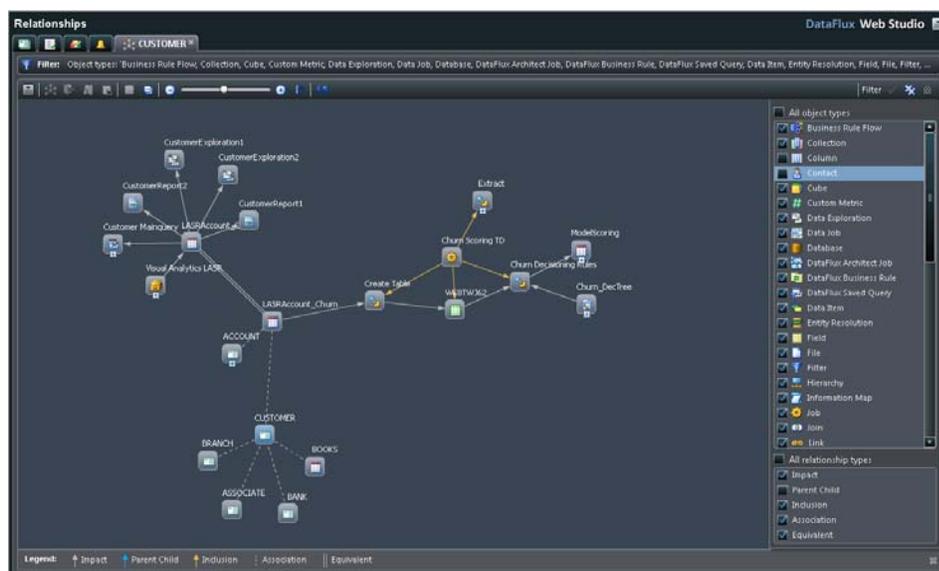


Figure 10: Impact Analysis View

ADDITIONAL FEATURES

There are several additional features in the SAS Data Management product suite that are worth noting for helping you successfully deploy your data governance initiatives. Master Data Management is a feature available in the SAS Data Management product suite that enables you to automate the process of selecting a best record. It has many new features that support data governance. For example, you can design your Master Data Management data model in the Business Data Network, and then export it directly into the Master Data Management application. Other features include workflow integration, reporting, and hierarchy management.

The Data Federation server can help you manage your data source systems. It provides full auditing and monitoring capabilities for all data managed through the Data Federation server. It provides many security features and single, point-of-control user-rights management including table, column, and row-level permissions.

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 discussed data governance best practices. It explained where and how SAS capabilities (such as the Business Data Network, Reference Data Management, Data Federation, Data Quality, Data Management, and Master Data Management) 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.

RECOMMENDED READING

- SAS Enterprise Data Management & Integration Discussion Forum. Available at http://communities.sas.com/community/sas_enterprise_data_management_integration.
- Rausch, Nancy, and Malcolm Alexander. 2013. "Best Practices in SAS® Data Management for BigData." *Proceedings of the SAS Global Forum 2013 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings13/077-2013.pdf>.
- Rausch, Nancy, and Malcolm Alexander. 2013. "What's New in SAS® Data Management." *Proceedings of the SAS Global Forum 2013 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings13/070-2013.pdf>.
- Hazejager, Wilbram, and Pat Herbert. 2011. "Innovations in Data Management: Introduction to DataFlux®Data Management Platform." *Proceedings of the SAS Global Forum 2011 Conference*, Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings11/141-2011.pdf>.

- Herbert, Pat. 2011. "Master Data Management, the Third Leg of the Data Management Stool: a.k.a. the DataFlux® qMDM Solution." *Proceedings of the SAS Global Forum 2011 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings11/146-2011.pdf>.
- Rausch, Nancy, Michael Ames, and Wilbram Hazejager. 2012. "What's New in SAS® Data Management." *Proceedings of the SAS Global Forum 2012 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings12/110-2012.pdf>.
- Rausch, Nancy, and Tim Stearn. 2011. "Best Practices in Data Integration: Advanced Data Management." *Proceedings of the SAS Global Forum 2011 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings11/137-2011.pdf>.

CONTACT INFORMATION

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

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

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.