Data Stewardship in Complex and Big Data Environments

A White Paper
# Table of Contents

1. **Introduction**

2. **What is Data Stewardship?**
   - What Are Data Stewards and Why Are They So Important?
   - What is the Purpose and Value of Data Stewardship?
   - Stewardship in Complex and Big Data Environments

3. **The Role of Technology in Data Stewardship**

5. **Data Stewardship Use Case**

9. **The iWay Information Asset Management Platform**
   - Data Quality Suite
   - Master Data Suite
   - Integration Suite

14. **Conclusion**

14. About Information Builders
Introduction

For data management professionals, data stewardship is a fundamental requirement for successful data governance. This sentiment stems from the notion that data stewardship is a necessity for implementing data governance disciplines.

Data stewardship is universally recognized as mission critical, and comes with distinctive sets of obstacles that test even best-of-breed organizations. Coordinating and managing data policy across multiple people, data domains, and interdisciplinary business processes can present a hearty challenge. Furthermore, accomplishing this in a multifaceted organization with compound data touch points, large data volumes, multiple data sources, conflicting business interests, various data consumers, and disparate technologies adds to the fury. These obstacles may test the patience and aptitude of any organization to move forward with data stewardship and governance.

As daunting as the task may appear, data stewardship remains achievable, even within the most dynamic data and business environments. Much of what is being accomplished is attributed to emerging technologies that automate the interoperable aspects of this discipline. In doing so, technology enables companies and agencies to overcome traditional multifarious data stewardship challenges to help them become data-driven organizations.

Before examining the role of technology in this practice, it is important to identify the concepts of data stewardship. Reviewing the operational and business intricacies of this practice can improve an organization’s understanding of the purpose and value that technical solutions bring to the overall initiative.
Data Stewardship

What is Data Stewardship?

Data stewardship is the orchestrated practice of monitoring, measuring, managing, and reporting on data integrity, data policies, and data performance within an organization. It typically consists of a group of individuals, known as data stewards, who channel their subject matter expertise in the data, business use of the data, and the technologies that facilitate the information.

A sub-domain of a broader data governance program, the data stewardship practice implements and monitors the policy set forth by the data governance organization.

What Are Data Stewards and Why Are They So Important?

Data stewards are essentially the custodians and guardians of the organization’s data assets at an operational level. Their individual expertise and unique visibility into the data is a vital component in managing these valuable resources across an enterprise, institution, or agency. Data stewards are “the eyes and ears” of the organization’s data and information assets; the more organized they become, the more benefits they yield to the organization.

Most data stewards do not hold full-time data management roles. However, our experience points out that an increasing number of organizations are on the path towards making data quality a dedicated part of their corporate structure. Steward positions are often filled by subject matter experts that dedicate a portion of their time to data-related activities, while fulfilling a specific business or technical responsibility. For example, a direct marketing campaign specialist has various responsibilities for helping identify, manage, and drive a meaningful marketing campaign. Although significant, only a portion of their time is devoted to managing data within a campaign list or database. Their insight into customer and product data, and how it drives business performance is an important contribution to the overall integrity of the information within the organization. The same holds true for analysts, report writers, data quality professionals, database administrators, data entry clerks, data integration architects, and more.

What is the Purpose and Value of Data Stewardship?

As a formalized network of data professionals, data stewardship ensures the integrity, visibility, usability, and security of the organization’s data. This association of individuals creates the framework used to shift organizations from a reactive, event-driven entity, to a proactive, data-driven enterprise. The process uncovers hidden opportunities and risk that are concealed in a normal course of business. Ultimately, these collective efforts improve the utilization and accuracy of data and information in a way that drives business performance and mitigates organizational risk.

Stewardship in Complex and Big Data Environments

Big data environments present challenges to data stewardship. The data and business processes are more distributed across systems and business entities, creating complexity. In practice, the operational framework for data stewardship remains the same. However, the technologies required to facilitate data stewardship and enable collaboration between data stewards must be capable of scaling these dynamic environments.
The Role of Technology in Data Stewardship

For successful data stewardship, organizations must look to technologies that can perform six mission-critical functions in an integrated fashion to promote data integrity, collaboration, workflow management, and accountability among participants. These six areas include: data integration, data profiling, data quality, data remediation for exceptions, document access, and collaboration/interoperability.

1 **Data Integration** – Data stewards must be able to access and share data from all data sources. For some entities, this may not be an immediate project-related requirement. However, gaining access to data and scaling it over time is critically important to long-term data steward and governance success. To achieve this, organizations need to adopt an integration platform that allows them to access data within all applications and systems in various real-time, near real-time, and batch scenarios.

2 **Data Profiling** – Data stewards require common tools that can help them to quickly analyze, measure, monitor, and report on data within specific data sources. Such technologies provide out-of-the-box functionality that can analyze data sources in minutes as opposed to the weeks and months it can take to get results from traditional SQL scripts. Configurable profile tools also allow for the inclusion of precise and reusable business rules specific to the entity. The use of a commercial profile tool reduces the time and resources necessary to profile data, reduces the education and training curve for technical and non-technical individuals, and provides a common reporting platform for data stewards across the organization.

3 **Data Quality** – Data stewards rely on data quality technologies to automate the correction, normalization, and linking of data based on the business rules they help develop and oversee. In complex data environments it becomes critical to standardize, match, and merge information, especially in diverse real-time and batch situations. Commercial-grade technologies of this nature will provide a higher level of automation and reduce manual intervention by data stewards.

4 **Remediation for Exceptions** – Data stewards are often confronted with data anomalies that require manual inspection, in some cases by multiple interdisciplinary business units. These circumstances are created whenever data results from automated data quality routines fall outside of predefined thresholds, or the results do not comply with data policy or business process. When this occurs, data must be fixed, and sometimes under conditions that require multiple review and security access levels.

It is common to find organizations that over time have established teams of people to review data. In less obvious or organized situations, the task is distributed across a number of people. In other cases, data that requires manual attention is often ignored, causing repetitive and ongoing problems across the enterprise.

Emerging technologies provide automated workflows for processing data problems that require manual inspection. These workflows can reduce the burn and churn that organizations frequently experience, and provide an audit trail in the process.
**Document Access** – During daily operations, data stewards need direct access to supporting documentation such as data standards, data policies, data dictionaries, data vocabularies, data governance charters, and other information. Advanced data stewardship technology allows users to access the documents they need to perform their data stewardship duties.

**Collaboration/Interoperability** – Data stewards must be able to effectively collaborate with each other and various work groups – because of the dynamic data and business landscape. It is important to have tools that promote the interoperability of data stewards across the organization.

Collectively, these technologies make the difference between a basic data stewardship program and a finely tuned, results-oriented program. They form the framework for enabling and scaling data stewardship and governance in highly complex and big data environments.
A large retailer needs to manage and coordinate the information requirements of multiple
distributed business units and stores across the regions of North America and Europe. The data
environment consists of several complementary – yet separate – applications and technologies,
multiple data owners, and numerous data warehouses and data marts. Information is highly
fragmented, analogous to the people and business processes. The challenge is further complicated
since there are few data standards and data validation checks-and-balances built into the systems.

The company recognizes that there are multiple business problems that flow from this set of
circumstances – adding cost to the bottom line and inhibiting revenue growth. Specifically, the
company has identified that poor data quality and interoperability between stores and business
units materially impacts marketing campaigns, inventory management, partner sales reporting,
customer satisfaction, and ultimately, overall sales.

Much of the problem is attributed to the rapid organic growth of the company, acquisitions, and
expanded partnerships. There are no signs that indicate a slowdown in these practices.

**Store Operations:** In addition to purchase transactions, each store collects loyalty membership
applications, works with local suppliers, tracks partner sales, and manages much of its own
inventories. Purchase transactions are recorded in a common point-of-sale (POS) system that
integrates with one financial and four inventory systems. Loyalty membership is manually
entered into either a home-grown or corporate marketing system from information supplied by
the customers at the store. Approximately 15 percent of stock is purchased directly from local
suppliers at the store’s discretion, with the remaining coming from corporate negotiated accounts.

The store merchandising manager is responsible for most of the inventory and supplier
relationships. Sales associates at the point of sale record loyalty membership applications and
partner sales information. Neither has quality checks built into the process.

**Marketing:** The marketing department works from multiple data marts for creating and maintain-
ing campaigns around customer loyalty programs. The reason for multiple instances is attributed
to poor data quality and the lack of trust in the data. When one data mart fails, the company builds
another one. Despite its best efforts, the department consistently deals with duplicate, inconsistent,
inaccurate, and missing customer data. This makes it more costly to create a marketing campaign
and erodes the effectiveness of it, preventing the company from achieving revenue targets.

For example, one popular campaign is to mail a 20 percent off coupon to customers during their
birthday month. This may increase the number of sign-ups for the customer loyalty program.
However, customers often sign up multiple times using alias names and e-mail addresses, which
has become counter-productive. Another campaign, though less effective, is to target loyalty
customers on the basis of their historical taste in products previously purchased. Due to untrust-
worthy data, the company cannot effectively match customer buying preferences with new or
existing products. This undermines the ability to market, up-sell, and cross-sell.
Partner Sales Reporting: For certain products, the sale of partner goods requires that additional customer information be collected at the time of sale. This data is used to facilitate warranties, rebates, recalls, and the reporting of partner sales for compensation purposes. Because of inconsistencies in how this data is collected, the sales administration office takes nearly four weeks each quarter to reconcile sales transactions with partners, and determine sales compensation for internal sales associates. The company is aware that current methods lead to underpayment, overpayment, and potential fraudulent activities.

Corporate Reporting and Analytics: The company has a team of analysts responsible for creating reports for upper management. This group pulls data from the financial, transactional, customer loyalty, and inventory systems for various reports. A data warehouse was created about five years ago to enhance this process.

The reporting team is consistently challenged with reconciling reports at the store, regional, and national level. This is attributed to significant inconsistencies between inventory systems, stock-keeping units (SKU), supplier product codes, area overlap, and duplicate customer information.

Inventory Control: While inventory systems are integrated, the SKUs, various supplier codes, and product descriptions often differ from store to store and inventory system to inventory system. This discrepancy results in costly overstocking and stock outs. Incorrect supplier purchases are also a common occurrence.

Merchandising: A small, but growing portion of the retailer’s business is conducted online, where customers can order items via the Internet. This part of the business is the responsibility of the merchandising department. The merchandising team is consistently challenged with posting the right product descriptions, item options, images, and pricing. In addition, the team needs to determine whether the goods are in stock, and at which locations.

The multiple inventory systems can easily lead to an error. It is not unusual for the wrong item to be shipped to the customer, or for it to be sold at the wrong price. When this happens, it creates havoc, rework, and unnecessary costs associated with shipping, returns, and billing – thus negatively impacting customer satisfaction.

The Solution: The initial solution for this retailer was to institute a corporate data governance practice with an immediate focus on data standards, data quality, and data stewardship. In conjunction with mobilizing the business process, the organization implemented tools for data profiling, data quality, and data stewardship workflow management.
A corporate data governance practice helps to drive measurable improvements in sales and operations.

With Information Builders’ Information Asset Management Platform, a corporate data governance practice was executed, with an immediate focus on data standards, data quality, and data stewardship. Customer data is validated in real time, product data is cleansed, and data among inventory systems is consistent and accurate, increasing the company’s revenue while eliminating unnecessary costs.

The technology allowed the company to formulate a collaborative data stewardship program between the following individuals and work groups:

- In-store merchandising team
- In-store sales associate team
- Corporate reporting team
- Sales administration reporting team
- Marketing campaign team
- Customer service representatives
The collaborative tools enabled the company to proactively catch and fix data problems up front, identify opportunities and risk, address and correct lingering inconsistencies, and share and fix common problems that impact multiple business operations.

The data quality technology allowed the organization to validate customer data in real time, as a transaction is entered at the store. It was also used to cleanse product data by cross-referencing, validating, and standardizing product information in both batch and real-time environments. This standardized data between inventory systems and significantly raised the level of consistency and accuracy of the information for downstream business use.

Common data profiling tools were implemented across work groups where consistent methods and standards were used to identify anomalies, trends, opportunities, and risk. It also led to consistent reporting across data domains and business units where data metrics could be applied and measured.

Data steward workflow management software was also implemented. The technology allowed the company to do the following: identify unique data problems at every step in the data lifecycle, enter the issue into an interactive workflow, and have the appropriate work group address the problem for resolution.

Collectively, the company witnessed a measurable improvement in sales and operations across the organization as they transformed from an event-driven organization to a data-driven company.
Whether the data environment is traditional or complex, Information Builders understands what your organization needs to manage information assets and enable successful data stewardship. Information Builders continues to evolve these three pillars of our iWay Information Asset Management Platform:

- Data Quality Suite
- Master Data Suite
- Integration Suite

This collection of advanced information management technologies positions organizations to unlock the door to successful data governance and data stewardship.

**Data Quality Suite**

The Data Quality Suite is powerful and flexible, enabling users to validate and handle data quality in a number of ways – even in real time as users are interacting with information online, or as new data is being generated during the course of transactions and events. This approach ensures that bad data is proactively kept out of the environment, before it can pollute other systems and sources. Data quality can also be managed in batch at pre-defined intervals to effectively eliminate pre-existing errors and inaccuracies.

**Data Quality Firewall**

With the Data Quality Suite, organizations can create a real-time data quality firewall that leverages fully customizable business rules to proactively stop bad information from entering systems. When used in conjunction with the iWay Master Data Management and Integration Suites, this approach can support such initiatives as master data management, business-to-business (B2B) integration, and application integration.
Data Profiling
As the first step in any data quality initiative, data profiling enables users to assess and understand data through basic analysis. The Data Quality Suite’s profiling capabilities include pre-built metrics and customizable business rules for monitoring data, developing distributable reports, and performing offline analytics in batch or real time. The suite can generate advanced, built-in, in-depth reports and profiling information that provide efficient and accurate tracking and comparisons of data quality improvements over time.

Data Quality Management
This state-of-the-art solution is designed to optimize information integrity throughout its lifecycle. It leverages comprehensive, fully customizable business rules and a portfolio of localized dictionaries, as well as advanced unification, validation, cleansing, and enrichment techniques to not only identify and correct bad data, but also to proactively prevent it from entering the environment in the first place.

Serving as the main hub for data quality within an organization, the Data Quality Suite delivers centralized management of business rules, information integrity, and data flows, while providing a single point of integration and management for data from external master data systems and other sources.

Data Stewardship
A web-based foundation for data governance allows business users and data stewards to manage configurable process-based remediation workflows to correct and resolve data quality issues quickly and efficiently.

Notifications can be sent to mobile devices when human intervention is required.

Data Steward Portal lets data contributors and data stewards to manage configurable process-based remediation workflows to efficiently correct and resolve data quality issues.
With ease of access for business users, data quality issues can be monitored, managed, resolved, and tracked in real time, with an end-to-end audit trail of all changes. The workflow orchestrator for remediation processes can be customized, and optionally allows remediation edits to be reincorporated back to the source or destination.

**Master Data Suite**

The iWay Master Data Suite is a comprehensive, unified toolset that ensures consistency, uniformity, and accuracy across all critical data assets. This powerful, scalable multi-domain master data management environment allows organizations to enable effective data governance by rapidly creating and efficiently maintaining a single view of their core entities.

Master Data Suite enables data analysis and manipulation, 360-degree views, and hierarchy management to ensure information consistency, uniformity, and accuracy across an enterprise.

The Master Data Suite ensures commonality and interoperability between its various components, which include:

**Data Quality Suite**

The Master Data Suite includes all the features of the Data Quality Suite, in addition to specialized master data management capabilities. This foundation for data governance allows users to manage
configurable process-based remediation workflows to correct and resolve data quality and master data management issues in real time. Notifications can be sent to mobile devices when human intervention is required. Data quality issues can also be monitored, managed, resolved, and tracked in real time, with an end-to-end audit trail of all changes.

**Hierarchy Manager**

This specialized interface for jobs dealing with hierarchical master data, such as product hierarchies and categories or organizational structures, allows users to productively manage complex hierarchies spread over one or more domains. Features include:

- Efficient creation of hierarchies, including the definition of hierarchy nodes and the relationships between nodes
- The ability to attach domain golden records, or associated data records, to nodes within hierarchies
- Management of previously-configured hierarchies, such as cloning, merging, absorbing, and updating
- Tools to define and manage hierarchies
- An intuitive interface for viewing and managing hierarchies
- Configuration and storage of node-specific presentation information

**360 Viewer**

The 360 Viewer is an end-user interface for visualization, job management, and master data management. It offers a complete, web-based view of golden records mastered across all functional domains. Additionally, data stewards, data supervisors, and system administrators can unlink and delete instance records that are associated with a particular golden record group. To preserve data security, all users must be authenticated and authorized with roles and privileges before accessing the 360 Viewer.

**Pre-Built Business Models and Processes**

Pre-built business models, optimized for customer, contact, and product, are provided. Models may be imported, modified, or created from scratch to suit business needs. Business processes can directly leverage a variety of pre-defined and pre-configured master data management services, from simple customer management to customer profile searches and unification with additional business entities. Vertical solutions for healthcare, financial services, insurance, and government are also available.

**Integration Suite**

The iWay Integration Suite delivers powerful integration capabilities that enhance information unification and accessibility.

**A Robust Integration Infrastructure**

The most flexible and agile integration foundation on the market today, iWay provides unmatched interoperability between disparate systems and data. iWay Integration Infrastructure Middleware is the fastest way to create powerful, reusable business services from existing applications. All
aspects of an existing infrastructure – every integration, application, and development platform – work in concert with modernized service-oriented architectures (SOA) and event-driven architectures (EDA), to rapidly develop new business applications.

**Application Integration**
A unique and powerful enterprise service bus (ESB) enables the creation, composition, and management of services – whether invoked as web services or through other interfaces. It also provides event-driven integration and B2B interaction management and, unlike other ESBs, interoperates with proprietary technologies as well as industry standards.

The iWay Integration Suite lays the foundation for real-time integration, and web-oriented or event-driven architectures. It delivers rapid throughput and latency, making it a prime component of high-performance networks and business processing solutions.

**Data Integration**
Direct access to any data, in any source, allows organizations to address the individual information needs of all users and systems. iWay accelerates deployment and reduces risk for all types of data integration projects, including enterprise information management (EIM), data migration, web services, on-premise or cloud-based implementations, and co-existing or federated integration platform-as-a-service (iPaaS) environments.

The iWay Integration Suite is available on Amazon Elastic Cloud Computing (EC2), offering comprehensive integration in a highly secure environment. It allows for fully protected connections between Amazon EC2 instances of iWay solutions and all internal corporate back-end applications and systems, such as SAP or DB2.
Conclusion

No data governance plan is complete without data stewardship. In spite of the challenges that go hand-in-hand with data stewardship, it is an achievable goal – no matter how complex or dynamic a business environment is.

Technology plays an important role in data stewardship, and with the right solutions in place, companies can ensure success in their data stewardship initiatives through greater management and control over their information sources. The iWay Information Asset Management Platform from Information Builders delivers three comprehensive suites – Data Quality, Master Data, and Integration – to facilitate efficient and effective data stewardship in any traditional, complex, and big data scenario.

About Information Builders

Information Builders helps organizations transform data into business value. Our software solutions for business intelligence and analytics, integration, and data integrity empower people to make smarter decisions, strengthen customer relationships, and drive growth. Our dedication to customer success is unmatched in the industry. That’s why tens of thousands of leading organizations rely on Information Builders to be their trusted partner. Founded in 1975, Information Builders is headquartered in New York, NY, with offices around the world, and remains one of the largest independent, privately held companies in the industry. Visit us at informationbuilders.com, follow us on Twitter at @infobldrs, like us on Facebook, and visit our LinkedIn page.