THE POWER TO KNOW®

Data Governance for Master Data Management and Beyond

A White Paper by David Loshin

WHITE PAPER

Table of Contents

As a result of both external pressures, such as compliance, and internal pressures triggered by aggressive enterprise information management programs, there is growing interest on behalf of both data management professionals and senior business managers to understand the motivations, mechanics, virtues and ongoing operations of instituting data governance within an organization. The objective of data governance is predicated on the desire to assess and manage the many different kinds of risks that lurk hidden within the enterprise information portfolio. And while many data governance activities are triggered by a concern about regulatory compliance, the definition, oversight and adherence to information policies and procedures can create additional value across the enterprise.

One of the major values of a master data management (MDM) program is that, because it is an enterprise initiative, a successful initiative will be accompanied by the integration of a data governance program. As more lines of business integrate with core master data object repositories, there must be some assurance of adherence to the rules that govern participation. Yet while MDM success relies on data governance, a governance program can be applied across different operational domains, providing economies of scale for enterprisewide deployment.

There are many different perceptions of what is meant by the term “data governance.” Data governance is expected to address issues of data stewardship, ownership, compliance, privacy, data risks, data sensitivity, metadata management, MDM and even data security. What is the common denominator? Each of these issues revolves around ways that technical data management is integrated with management oversight and organizational observance of different kinds of information policies.

Whether we are discussing data sensitivity or financial reporting, the goal is to integrate the business policy requirements as part of the metadata employed in automating the collection and reporting of conformance to those policies. Especially in an age where noncompliance with external reporting requirements (e.g., Sarbanes-Oxley) can result in fines and prison sentences, the level of sensitivity to governance of data management will only continue to grow.

Aligning Information Objectives with the Business Strategy

Every organization has a business strategy that reflects both the business management and the risk and compliance management objectives. And the success of the organization depends on the ability to manage how all operations conform to the business strategy. This is true for information technology, but because of its centrality in the application infrastructure, it is particularly true in the area of data management.

It is important to communicate the business strategy and engineer the oversight of information in a way that aligns the business strategy with the information architecture. This alignment is twofold – it must demonstrate how the organization understands and uses its information assets as well as how the asset is managed over time.

Clarifying the Information Architecture

Many applications are created in a virtual vacuum, engineered to support functional requirements for a specific line of business without considering whether there is any overlap with other line of business applications. While this tactical approach may be sufficient to support ongoing operations, it limits enterprise analytic capability and hampers any attempt at organizational oversight.

Before a data governance framework can be instituted, management must understand and document the de facto information architecture, creating a preliminary phase to governance: assessment. In other words, a prelude to governance involves taking inventory to understand what data assets exist, how they are managed and used, how they support the existing application architecture, and evaluating where existing inefficiencies or redundancies create roadblocks to proper oversight.

The inventory will grow organically – initially the process involves identifying the data sets used by each application and enumerating the data attributes within each data set. Each data element must have a name, a structural format and a definition, all of which must be documented within a core metadata repository. Each data set models a relevant business concept, and each data element provides insight into that business concept within the context of the "owning" application. In turn, each definition must be subjected to review to ensure that it is correct, defensible, and grounded by an authoritative source.

This collection of data elements does not constitute a final architecture, though. A team of subject matter experts and data analysts must look at the physical components and recast it into a logical view that is consistent across the enterprise. This requires harmonization of the data elements to look for similarity (or distinction) in meaning. While this activity overlaps with the technical aspects of master data object analysis, its importance to governance lies in the identification and registration of the organization's critical data elements, their composed information structures, along with the application functions associated with the data element life cycle.

Mapping Information Functions to Business Objectives

Every activity that creates, modifies or retires a data element must somehow support a business activity that contributes to the organization's overall business objectives. In turn, the success or failure of the business activity is related to the appropriate and correct execution of all the information functions that support that activity. For example, many website privacy policies specify that data about children below a specified age will not be shared without permission of the child's parent. A data element may be used to document each party's birth date and parent's permission, and there will also be functions to verify the party's age and parental permission before the information is shared with another organization.

When assessing the information architecture, one must document each information function and how it maps to business objectives. A standardized approach for functional description will help assess functional overlap, which may be subject for review as the core master data objects are identified and consolidated. However, in all situations, the application functionality represents the ways that information policies are implemented across the enterprise.

Instituting a Process Framework for Information Policy

The goal of the bottom-up assessment is to understand how the information architecture and its associated functionality support the implementation of information policy. But in reality, the process should be reversed – information policy should be defined first, and then the data objects and associated services should be architected to implement and document compliance with that policy.

This process framework is supported by and supports a master data management environment. As critical data objects and their associated attributes are absorbed under centralized management, the ability to map the functional service layer to the deployment of information policy enables the embedding of monitoring probes that collectively (and reliably) report on compliance.

Data Quality and Data Governance

A data quality and data governance assessment clarifies how the information architecture is used to support compliance with defined information policies. It suggests that data quality and data standards management are part of a much larger picture with respect to oversight of enterprise information.

In the siloed environment, the responsibilities – and ultimately the accountability for ensuring that the data meets the quality expectations of the client applications – lie within the management of the corresponding line of business. But looking at the organization's need for managing information oversight provides a conduit for reviewing the dimensions of data quality associated with the critical data elements, expressing the data rules that affect compliance, defining quantitative measurements for conformance to information policies, and ways to integrate these all into a data governance framework.

Areas of Risk

What truly drives the need for governance? While there are many drivers, a large component boils down to risk. Both business and compliance risks drive governance, and it is worthwhile to look at a few of the areas of risk associated with master data that require data management and governance scrutiny.

Business and Financial

If the objective of the MDM program is to enhance productivity to improve the organization's bottom line, then the first area of risk involves understanding how nonconformance with information policies puts the business's financial objectives at risk.

Reporting

Certain types of regulations (e.g., Sarbanes-Oxley, 21 CFR Part 11, Basel II) require that the organization prepare documents and reports that demonstrate compliance – establishing accurate and auditable reporting as an area of risk. Accuracy demands the existence of established practices for data validation, but the ability to conduct thorough audits requires comprehensive oversight of the processes that implement the information policies. Consequently, ensuring report consistency and accuracy requires stewardship and governance of the data sets that are used to populate (or materialize data elements for) those reports.

Entity Knowledge

Maintaining knowledge of the parties with whom the organization does business is critical for understanding and mitigating both business risks (e.g., credit rating to ensure that customers can pay their bills) and regulatory risks. Many different industries are governed by regulations that insist on customer awareness, such as the USA PATRIOT Act, the Bank Secrecy Act and Gramm-Leach-Bliley, all of which require the ability to distinguish between unique individual identities. Ensuring that the tools used to resolve identities are matching within expected levels of trust and that processes exist for remediating identity errors falls under the realm of governance and stewardship.

Protection

The flip side of entity knowledge is protection of individual -- potentially private -- information. Compliance directives that originate in regulations such as HIPAA and Graham-Leach-Bliley require that organizations protect each individual's data to limit data breaches and protect personal information. Similarly to entity knowledge, confidence in the management of protected information depends on conformance to defined privacy and data protection constraints.

Limitation of Use

Regulations and business arrangements (as codified within contractual agreements) both establish governance policies for limiting how data is used, how it is shared, what components may be shared, the number of times it can be copied, as well as overseeing the determination of access rights for the data. Data lineage, provenance, and access management are all aspects of the types of information policies whose oversight is incorporated within the governance program.

The Risks of Master Data Management

If these types of risks were not enough, the deployment of a master data management program introduces organizational risks of its own. As a platform for integrating and consolidating information from across vertical lines of business into a single source of truth, MDM implies that independent corporate divisions (with their own divisional performance objectives) yield to the needs of the enterprise.

As an enterprise initiative, MDM requires agreement from the participants to ensure program success. This leads to a unique set of challenges for companies undertaking an MDM program.

Establishing Consensus for Coordination and Collaboration

The value of the master index or repository is the agreement (across divisions and business units) that it represents the highest quality identifying information for enterprise master data objects. The notion of agreement implies that all application participants share their data, provide positive input into its improvement and trust the resulting consolidated versions. The transition to the use of the master repository suggests that all application groups will work in a coordinated manner to share information and resources to make sure that the result meets the quality requirements of each participant.

Data Ownership

Each line of business may have its own perception of data ownership, ranging from an information architecture owned by the business line management to data that is effectively captured and embedded within an application (leading to a high level of data distribution). As diffused application development likely occurs when there is an absence of an official data ownership policy, each business application manager may view the data owned by the group or attached to the application.

When centralizing shared master data, though, the consolidation of information into a single point of truth implies dissolution of the traditional implied data ownership model. This requires the definition and transference of accountability and responsibility from the line of business to the enterprise, along with its accompanying policies for governance.

Form, Function and Meaning

Distributed application systems designed and implemented in isolation are likely to have similar, yet perhaps slightly variant definitions, semantics, formats and representations. An MDM migration requires establishing processes for resolving those subtle (and presumed meaningless) distinctions in meaning that can become magnified during consolidation, especially in relation to data element names, definitions, formats and uses.

This can be addressed by providing guidelines for the capture of data element metadata and its subsequent syntactic and semantic harmonization. Organizations can also establish guidelines for collaboration among the numerous stakeholders so that each is confident that using the master version of data will not affect the appropriate application.

Managing Risk Through Measured Conformance to Information Policies

Preventing exposure to risk requires creating policies that establish the boundaries of the risk, along with an oversight process to ensure compliance with those policies. Yet while each set of policies may differ depending on the different related risks, all of these issues share some commonalities:

- Federation – In each situation, the level of risk varies according to the way that information is captured, stored, managed, and shared among applications and individuals across multiple management or administrative boundaries. In essence, because of the diffused application architecture, to effectively address risks, all of the administrative groups must form a federation, slightly blurring their line-of-business boundaries for the sake of enterprise compliance management.
- Defined Policy – To overcome the challenges associated with risk exposure, policies must be defined (either externally or internally) that delineate the guidelines for risk mitigation. While these policies reflect operational activities for compliance, they can be translated into rules that map against data elements managed within the enterprise.
- Transparent Oversight – One aspect of each of the areas of risk described here is the expectation that there is some person or governing body to whom policy conformance must be reported. Whether that body is a government agency overseeing regulatory compliance, an industry body that watches over industry guidelines or public corporate shareholders, there is a need for transparency in the reporting framework.
- Auditability – The need for transparency creates a need for all policy compliance to be auditable. Not only does the organization need to demonstrate compliance to the defined policies, it must be able to both show an audit trail that can be reviewed independently and show that the processes for managing compliance are transparent as well.

The upshot is that whether the objective is regulatory compliance, managing financial exposure or overseeing organizational collaboration, centralized management will spread data governance throughout the organization. This generally happens through two techniques. First, data governance will be defined through a collection of information policies, each of which is mapped to a set of rules imposed over the life cycle of critical data elements. Second, data stewards will be assigned responsibility and accountability for both the quality of the critical data elements and the assurance of conformance to the information policies. Together these two aspects provide the technical means for monitoring data governance and provide the context to effectively manage data governance.

Critical Data Elements

Critical data elements are those data elements that are determined to be vital to the successful operation of the organization. For example, data elements that are used in reports (both internal and external), can capture identifying information of master data objects (e.g., customer, vendor, or employee data), or are critical for decision-making processes or for measuring organizational performance.

Part of the governance process involves a collaborative effort to identify critical data elements, research their authoritative sources, and then agree on their definitions. In turn, the master repository will become the source of truth for critical data elements.

Defining Information Policies

Information policies embody the specification of management objectives associated with data governance, whether they are related to management of risk or general data oversight. Information policies relate specified business assertions to their related data sets and articulate how the business policy is integrated with the information asset.

For example, consider the many regulations requiring customer knowledge, such as the anti-money laundering (AML) aspects required by the USA PATRIOT Act. The protocols of AML imply a few operational perspectives:

- Establishing policies and procedures to detect and report suspicious transactions.

- Ensuring compliance with the Bank Secrecy Act.

- Providing for independent testing for compliance to be conducted by outside parties.

But in essence, AML compliance revolves around a relatively straightforward concept: know your customer. Because all monitoring centers on how individuals are conducting business, any organization that wants to comply with these objectives must have processes in place for customer identification and verification.

Addressing the regulatory policy of compliance with AML necessarily involves defining information policies guiding the management of customer data, such as the suggestions in Figure 1. These assertions are ultimately boiled down into specific data directives, each of which is measurable and reportable, which is the cornerstone of the stewardship process.

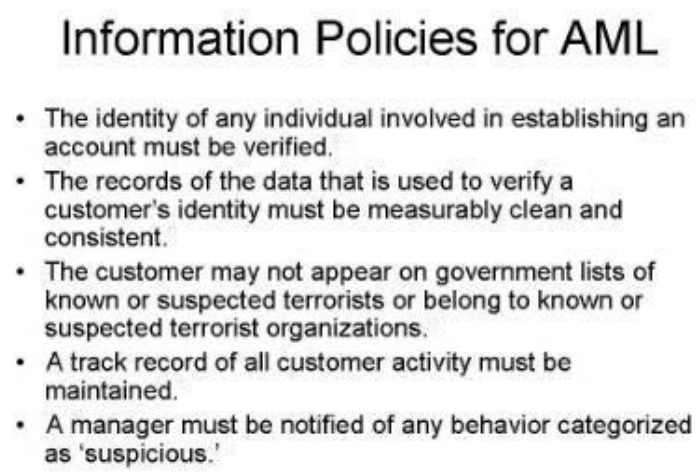

Figure 1: Example of information policies supporting a business policy

Metrics and Measurement

Any information policy might be further clarified into specific rules that would apply to both the master data set as well as the participating applications. For example, our example that required the tracking of customer activity might translate into a rule prescribing that each application that manages transactions must log critical data elements associated with the customer identity and transaction in a master transaction repository. Conformance to the rule can be assessed by verifying that all records of the master transaction repository are consistent with the application systems, where consistency is defined as a function of comparing the critical data values with the original transaction.

Metrics reflecting conformance with an information policy can be viewed as a roll-up of the various data rules into which the policy was decomposed. As long as each rule is measurable, we can create a hierarchy of metrics that ultimately can be combined into key performance indicators for the purposes of data governance.

Monitoring and Evaluation

The collection of key performance indicators provides a high-level view of the organizational performance with respect to the conformance to defined information policies. In fact, we can have each indicator reflect the rolled-up measurements associated with the set of data rules for each information policy. Thresholds may be set that characterize levels of acceptability, and the metrics can be drilled through to isolate specific issues that are preventing conformance to the defined policy, enabling both transparency and auditability.

But in order for the monitoring to be effective, those measurements must be presented directly to the individual that is assigned responsibility for oversight of that information policy. It is then up to that individual to continuously monitor conformance to the policy, and if there are issues, to use the drill through process to determine the points of failure and to initiate the processes for remediation.

A Framework for Responsibility and Accountability

One of the biggest historical problems with data governance is the absence of follow-through; while some organizations may have well-defined governance policies, they may not have established the underlying organizational structure to make it useful. This requires two things: the definition of the management structure to oversee the execution of the governance framework and a compensation model that rewards that execution.

A data governance framework must support the needs of all the participants across the enterprise, both from the top down and from the bottom up. With executive sponsorship secured, a reasonable framework can benefit from enterprisewide participation within a data governance oversight board, while all interested parties can participate in the role of data stewards. A technical coordination council can be convened to establish best practices and to coordinate technical approaches to ensure economies of scale. The specific roles include:

- Data Governance Director.
- Data Governance Oversight Board.
- Data Coordination Council.
- Data stewards.

Data Governance Director

The data governance director is responsible for the day-to-day management of enterprise data governance. The director provides guidance to all the participants and oversees adherence to the information policies as they reflect the business policies and necessary regulatory constraints. The data governance director plans and chairs the Data Governance Oversight Board. The director identifies the need for governance initiatives and provides periodic reports on data governance performance.

Data Governance Oversight Board

The Data Governance Oversight Board (DGOB) guides and oversees data governance activities. The DGOB is composed of representatives chosen from across the community. The main responsibilities of the DGOB include:

- Review corporate information policies and designate workgroups to transform business policies into information policies, and then into data rules.
- Approve data governance policies and procedures.
- Manage the reward framework for compliance with governance policies.
- Review proposals for data governance practices and processes.
- Endorse data certification and audit processes.
- Obtain support at the C-level.
- Warrant the enterprise adoption of measurably high-quality data.
- Negotiate quality SLAs with external data suppliers.

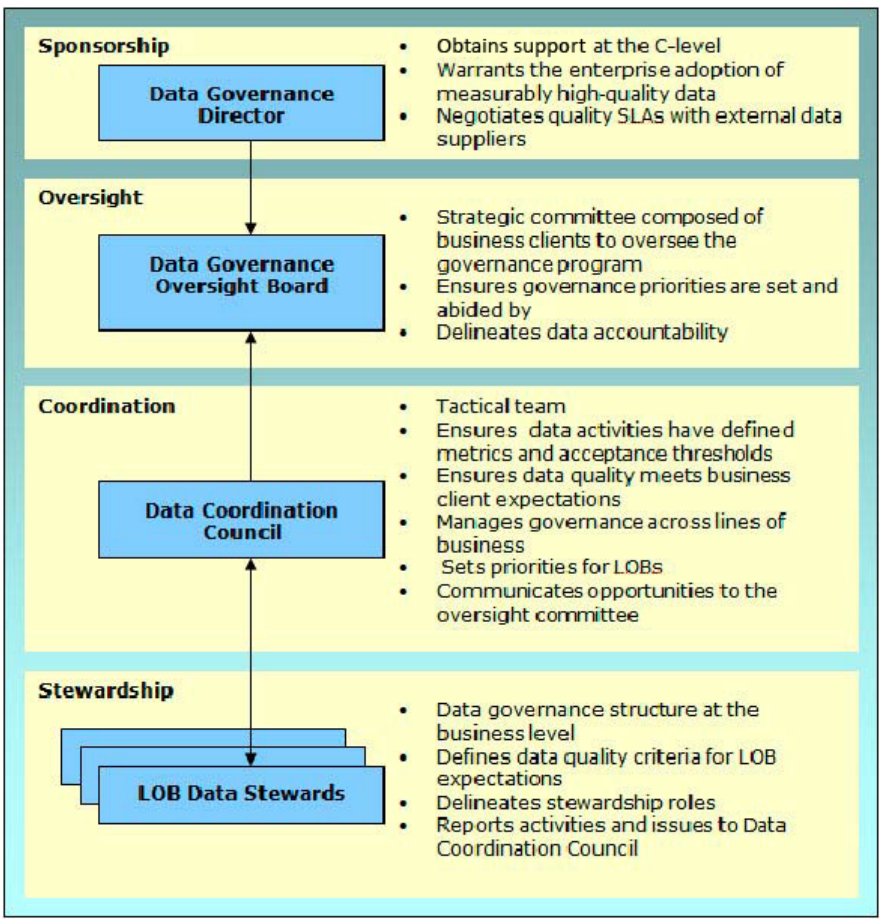

Figure 2: A framework for data governance management

Data Coordination Council

The actual governance activities are directed and managed by the Data Coordination Council, which operates under the direction of the Data Governance Oversight Board. The Data Coordination Council is a group composed of interested individual stakeholders from across the enterprise, and is responsible for adjusting the processes of the enterprise as appropriate to ensure that the data quality and governance expectations are continually met. As part of this responsibility, the Data Coordination Council recommends the names for and appoints representatives to committees and advisory groups.

The Data Coordination Council is responsible for overseeing the work of data stewards. The coordination council will also:

- Provide direction and guidance to all committees tasked with developing data governance practices.
- Oversee the tasks of the committee's advisory groups related to data governance.
- Recommend to the Data Governance Oversight Board the endorsement of output of the various governance activities for publication and distribution.
- Recommend data governance processes to the Data Governance Oversight Board for final endorsement.
- Nominate stewards and oversee the practices managed by the data stewards for data certification and managing audit information.
- Advocate for the enterprise data governance by leading, promoting, and facilitating the governance practices and processes developed.
- Provide progress reports, review statuses, and to discuss and review the general direction of the enterprise data governance program.

Data Stewardship

The data steward's role essentially is to support the user community, with responsibility for collecting, collating, and evaluating issues and problems with data. Prioritized issues must be communicated to those individuals that may be affected. The steward must also communicate issues and other relevant information (e.g., root causes) to those staff members that are in a position to influence remediation.

As the person accountable for the quality of the data, the data steward must also manage standard business definitions and metadata for critical data elements, and oversee the enterprise data quality standards, including the data rules associated with the data sets. This may require using technology to assess and maintain a high level of conformance to defined information policies within each line of business, especially its accuracy, completeness, and consistency. Essentially, the data steward is the conduit for communicating issues associated with the data life cycle – the creation, modification, sharing, reuse, retention, and back up of data. If any issues regarding the conformance of data to the defined policies over the data lifetime emerge, it is the responsibility of the steward to resolve them.

Data stewardship is not necessarily an information technology function, nor should it necessarily be considered to be a full-time position, although its proper execution deserves a proper reward. Data stewardship is a role that has a set of responsibilities along with accountability to the line of business management. In other words, even though the data steward's activities are overseen within the scope of the MDM program, the steward is accountable to his or her own line management to ensure that the quality of the data meets the needs of both the line of business and of the organization as a whole.

Pulling It All Together

For companies undertaking MDM, a hallmark of successful implementations will be the reliance and integration of data governance throughout the initiative. There are three important aspects of data governance for MDM and beyond:

- Managing critical data elements – Ensuring consensus in identifying data elements associated with common business terminology, researching their authoritative sources, agreeing on their definitions, and managing them within the master repository as the enterprise source of truth.
- Setting information policies and data rules – Determining the critical business policies that relate to data, and devising the information policies that embody the specification of management objectives associated with data governance, whether they are related to management of risk or general data oversight.
- Enforcing accountability – Empowering the right individuals in the organization to enforce well-defined governance policies and to establish the underlying organizational structure to make it possible by defining a management structure to oversee the execution of the governance framework along with the compensation model that rewards that execution.

Keeping these ideas in mind during the development of the MDM program will ensure that the master data repository doesn't become relegated to the scrapheap of misfired enterprise data management initiatives. Rather, developing a strong enterprise data governance program will benefit the MDM program as well as strengthen the ability to manage all enterprise information activities.

About the Author

David Loshin, president of Knowledge Integrity Inc., is a recognized thought leader and expert consultant in the areas of data quality, master data management and business intelligence. Loshin is a prolific author regarding data management best practices and has written numerous books, white papers and Web seminars on a variety of data management best practices.

His book, *Business Intelligence: The Savvy Manager's Guide* has been hailed as a resource allowing readers to "gain an understanding of business intelligence, business management disciplines, data warehousing and how all of the pieces work together." His book, *Master Data Management*, has been endorsed by data management industry leaders, and his valuable MDM insights can be reviewed at mdmbook.com. Loshin is also the author of the recent book, *The Practitioner's Guide to Data Quality Improvement.* He can be reached at loshin@knowledge-integrity.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 60,000 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**.

SAS Institute Inc. World Headquarters +1 919 677 8000
To contact your local SAS office, please visit: **sas.com/offices**