

Operationalizing Data Governance through Data Policy Management

Prepared for Kalido by:

David Loshin
Knowledge Integrity, Inc.
June, 2010

Introduction

The increasing dependence on (and frequent failure of) enterprise-wide business applications such as Enterprise Resource Planning (ERP), Business Intelligence and analytics, Master Data Management (MDM), and Supply Chain Management (SCM) only demonstrates that the traditional attempts for ensuring high quality data are not living up to enterprise expectations. The concern that source data sets will not live up to the requirements of enterprise business applications has prompted senior managers with many organizations to introduce the concept of data governance. Although the typical approach to data governance involves the creation of a data governance council or board, in the absence of well-defined policies and processes, this council often degenerates in terms of efficacy, participation, as well as management support. This “data governance gap,” occurs when the prerequisites for data governance policies, practices, and procedures are not established prior to the creation of the council.

Avoiding this pitfall means understanding what operationalizing data governance means and instituting the right policies and practices prior to launching the data governance council. In this paper we consider how data- or application-centric views of data governance reflect earlier stages of data governance maturity, and as the organization’s data management and governance practices mature, we see a transition to effective data governance that focuses the business processes and how they meet corporate success objectives.

With repeatable processes ensuring conformance with defined data policies that are correlated to business information requirements, an organization can make compliance with data policies pervasive through the definition, documentation, implementation, and then measurement of performance metrics relating to achieving business objectives. In turn, we explore how data policy management encompasses these aspects to provide a fundamental tool set for the data governance practitioner.

The Data Governance Gap

The growing interdependencies among collections of enterprise business processes and their corresponding applications expose the amplified risks associated with the absence of best practices for data management. Alleviating these risks drives the desire for formal data management oversight. Having read the latest articles and papers, and in their enthusiasm to institute formal data governance, senior managers at many organizations will often take the first step of establishing a data governance board populated with key business and technical stakeholders. Although the formal organizational structure is a critical component to exercising control over the enterprise information asset, there is the risk of creating a “data governance gap.”

This data governance gap occurs when corporate management creates a formal data governance organizational structure before there is a clear definition of the roles and responsibilities associated with ensuring that the governance policies and processes can be operationalized. The conventional wisdom recommends the creation of staff roles such as “data governance director,” “data owner,” “data governance board member,” or “data steward.” However, when there is a data governance gap, these roles are assigned to individuals even before there are policies in place for proposing, reviewing, and approving data policies. In addition, the assignment of these roles to staff members without the proper

structure in place often leads to *increased* fear and confusion as opposed to the desired effect of increased compliance with enterprise data expectations. Without a means for defining and documenting data policies as well as establishing methods for monitoring accountability of each role's performance with respect to the compliance with those data policies, these roles have accountability, but no tools for enforcement.

Why the Data Governance Gap is a Problem

At first, many data governance council members are enthusiastic about the opportunity to share their data management concerns and actively participate in the periodic meetings. But the risk of the data governance gap is that during the time delay between assigning roles and determining clear responsibilities with corresponding performance measures, issues are introduced and discussed, but no clear steps can be taken to address and remediate them. Frustration at the lack of progress grows, while interest in attending the meetings wanes. Essentially, these councils devolve into a technical forum for discussing, but not resolving, data issues.

The prime directive of establishing a data governance board and corresponding data stewardship teams is to ensure compliance with data policies that are directly linked to the achievement of the corporate business objectives. The data governance council is born out of the need for establishing best practices for data management. However, data governance cannot be operationalized through the creation of an organizational structure alone because a number of prerequisites for meeting the prime directive are missing, namely:

- Processes for correlating data errors with measurable business impacts;
- A process for collecting and documenting enterprise data consumer expectations;
- A process for defining and agreeing to data policies directly corresponding to business policies;
- Procedures for inspecting data values to assess compliance with defined data policies;
- Procedures for alerting key individuals when data does not comply with data consumer requirements;
- Processes for prioritizing and remediating data issues; and
- A framework and processes for monitoring performance with respect to data policy compliance.

The nascent data governance board provides a forum for developing those required processes and procedures. And because data policies are the instruments of governance, it is incumbent upon the data governance leaders to not just clearly articulate roles, responsibilities, and data management policies but also to employ the right tools and techniques for capturing data policies, their corresponding methods for measurement, acceptability thresholds, and remediation. Integrating the organizational framework with a means for operationalizing the definition and implementation of data policies is a way to minimize the potential risks of the data governance gap.

Levels of Maturity

In fact, an informal consideration of the maturity of these data management practices reflects the degree to which the practitioners are prepared to manage compliance with the data policies that reflect business needs. As is described in Table 1, at the initial or “awakening” level, there are few, if any formal governance processes available to support any data management best practices at the enterprise level. But as the data governance leaders refine the organization’s business requirements into defined and documented data policies, data compliance is incrementally integrated into the system development lifecycle, providing data controls that alert data stewards to emerging issues in accordance with data quality service level agreements.

	Awakening	Developing	Practicing	Fully Governed
<i>Business impact analysis</i>	Initial considerations of correlation between noncompliance with data expectations and business impacts	High-level categorization of business impacts related to data policy compliance	Well-defined impact categories, processes for mapping data expectations to business impacts, definitions of data rules	Hierarchies of impact categorization, data policies directly linked to business requirements, data rules for measuring compliance with policies
<i>Data requirements analysis</i>	Siloed application development, no perception of enterprise requirements	Limited understanding of cross-application data completeness and syntactic consistency expectations	Potential downstream consumers are identified and interviewed to solicit data quality expectations; data requirements are captured as policies and shared with implementation teams	Data requirements processes integrated into system development life cycle
<i>Data policy management</i>	No concept of a data policy	Data policies exist for data used in applications for some lines of business	Data policies defined and documented for common data sets and data domains that are used across different business applications	Data policies defined, documented, linked to business policies as well as appropriate level of data precision, and managed within a central repository
<i>Compliance inspection and monitoring</i>	Data consumers discover data issues and errors	Information Technology department runs data cleansing tools in ad hoc manner over static data sets	Validation of data policies inserted as inspection probes within some line of business applications	Data policy validation methods embedded within business processes across lines of business; data governance integrated within system development life cycle

	Awakening	Developing	Practicing	Fully Governed
<i>Alerting and remediation</i>	Information Technology team members reactively address most recent issues, with limited capability to eliminate root causes of problems	Information Technology has defined processes for remediation, with little or no business involvement	Formal processes for data stewards for analysis and remediation for most defects	Clearly established processes for triage, prioritization in relation to business impacts and remediation costs, shared responsibilities between business data owners, data stewards, and
<i>SLAs and Performance monitoring</i>	None	Information Technology department is presumed to be accountable for all data issues	The reliability of data exchanges is governed with defined data quality service level agreements, complete with data policies, corresponding validation measures, and escalation directives	Performance of data stewards with respect to agreed-to service levels is monitored in relation to compliance with service level agreement directives

Table 1: Selected considerations of data management maturity

Each of these levels corresponds to an evolving perspective of data governance, incrementally migrating responsibilities from the Information Technology department to the various lines of business. At the earlier stages of maturity, the focus is largely data-centric, with an emphasis on reacting to data issues by seeking to correct the data. Informal governance may exist in silos centered on systems or data repositories, but when similar data sets are inadvertently replicated in different locations, they may be governed by potentially different policies and processes. This is inefficient and ineffective, so mature companies want to manage policies in a more centralized way, and standardize on a uniform set of processes.

As the organization matures, the focus cycles through a series of transitions; as initial data governance policies and processes are defined, the focus gradually changes to concentrate on how application processes handle the data. Gradually, with the definition of enterprise data policies directly related to business success criteria coupled with methods for inspecting and monitoring compliance with those policies, data governance can emerge as a business priority.

Business Process Success Relies on Data Predictability

The challenge is that most organizations look at data errors from a data- or application-centric perspective, and not a business process perspective. With the advent of statistical analysis and data profiling tools, it is easy to assess quantifiable objective characteristics of the data, such as the percentage of missing values or a histogram of value frequencies. Although these objective measures often do not provide significant insight into the business reliance on high quality information, such insight is exactly what is required to understand and prioritize violations of data policies. For example,

counting missing values is less relevant than measuring the degree to which missing values impact revenue growth.

Data expectations and requirements are inextricably linked to meeting enterprise information usage needs. Data policies must reflect business relevance, and that means defining those policies within the context of the people, business processes, data, and systems. Therefore, in addition to the organizational infrastructure needed to facilitate the definition of data governance policies, there needs to be a way to operationalize data governance that supplements the data governance board’s ability to function effectively. Ensuring the predictability of the quality of enterprise data contributes to business process success.

Example

Data rules are often hidden in many forms, requiring some review and analysis for clarification and extraction. As a discrete example, consider the task of reviewing the data rules associated with a small subset of the payment terms described in a sample professional services agreement, as shown in Table 2.

Each term and clause of the contract details a specific component of the agreed-to business policies directing payment terms. The corresponding information policies for each clause detail the information expectations in order to ensure compliance with the business policy clause. In turn, each information policy is refined into a set of data rules describing validity constraints for submitted invoices at various stages within the Payment Process business workflow.

Business Policy Term	Business Policy Clause	Information Policy	Data Rules
<i>Billing Rate</i>	<i>Contractor’s hourly rate is \$50.00/hour</i>	Contractor invoice must specify hourly rate of \$50.00; any other rates are not allowed	<ul style="list-style-type: none"> • Company must have a persistent, unique record of each agreement • Company must have a persistent record of specified labor categories and corresponding rates for each contractor and each agreement • Each submitted invoice must have a non-null agreement identifier • Each submitted invoice must have a non-null value for labor category • Each submitted invoice must have a non-null value for the associated labor rate • The (labor category, labor rate) value pair must be consistent with one of the labor categories and its corresponding labor rate associated with the agreement
<i>Billing Rate</i>	<i>Total payment will not exceed \$10,000.00</i>	The total accumulated invoiced amount must be less than or equal to \$10,000.00	<ul style="list-style-type: none"> • The sum of the invoice totals for all invoices submitted for services rendered under the specified agreement <= \$10,000.00
<i>Payment</i>	<i>Contractor will</i>	The contractor may not	<ul style="list-style-type: none"> • For each month, the count of the invoices

Business Policy Term	Business Policy Clause	Information Policy	Data Rules
	<i>invoice company on a twice-monthly basis for actual services rendered including labor and expenses</i>	submit more than 2 invoices per month. Invoiced amounts must either be for services or for expenses	submitted for services rendered under the specified agreement <= 2 <ul style="list-style-type: none"> • Every invoice line item must have an invoice type of either services or expense
Payment	<i>Company will have 30 days from the date of receipt to pay a submitted invoice</i>	Invoices are documented and time-stamped when received; Workflow process for payment must be completed within 30 days	<ul style="list-style-type: none"> • Each invoice record must have a non-null date of receipt value • Each invoice record must log whether the invoice has been paid or not (i.e., the paid data attribute must be either 'N' or 'Y') • In the invoice has been paid, the invoice record must have a non-null date of payment • There may not be any invoice records where the duration between date of receipt and today's date is > 30 days and where the paid data attribute is not set to 'Y' • There may not be any invoice records where the paid data attribute is set to 'Y' and the duration between date of receipt and date of payment is > 30 days

Table 2: An example of refining business policies into data rules.

Having refined the business policies into data rules, compliance can be automatically monitored.

Operational Data Governance

Each of the defined data rules can be directly related to a specific business constraint as well as well-defined business value drivers. In our Payment Process workflow example, each data is intended to do more than validate the structural consistency and completeness of submitted invoices; the rules are meant to flag any data inconsistencies that could lead to a negative business impact, namely improper payments. Ultimately, this rolls up into a generalized corporate value driver of improving and streamlining accounts receivables efficiency.

Instituting the data controls provides a trigger to alert the right chain of command when business policies are not observed. However, instituting the data controls is not sufficient – the organization needs to operationalize the data governance processes so that when the right person is alerted to a potential noncompliance, there are well-defined steps taken to understand the impact and address the issue in an efficient and complete manner.

Operational data governance means instituting the policies and processes for the entire lifecycle of defining, implementing, and observing data controls related to potential information issues within the context of business success criteria. This requires a framework of:

- Definable and measurable controls associated with underlying data concepts based on recognized business constraints;
- Inserting controls within the variety of business processes that employ one or more instantiations of the underlying data concepts, positioned at the earliest place in the business process flow;
- Monitoring, reporting, and alerting when there are issues;
- Prioritization methods for issues in relation to enterprise business needs; and
- Methods for investigation and remediation of root causes.

Data Policy Life Cycle and Management

This suggests that operational data governance imposes a life cycle for the definition and implementation of data policies, and Figure 1 provides a broad overview of this life cycle.

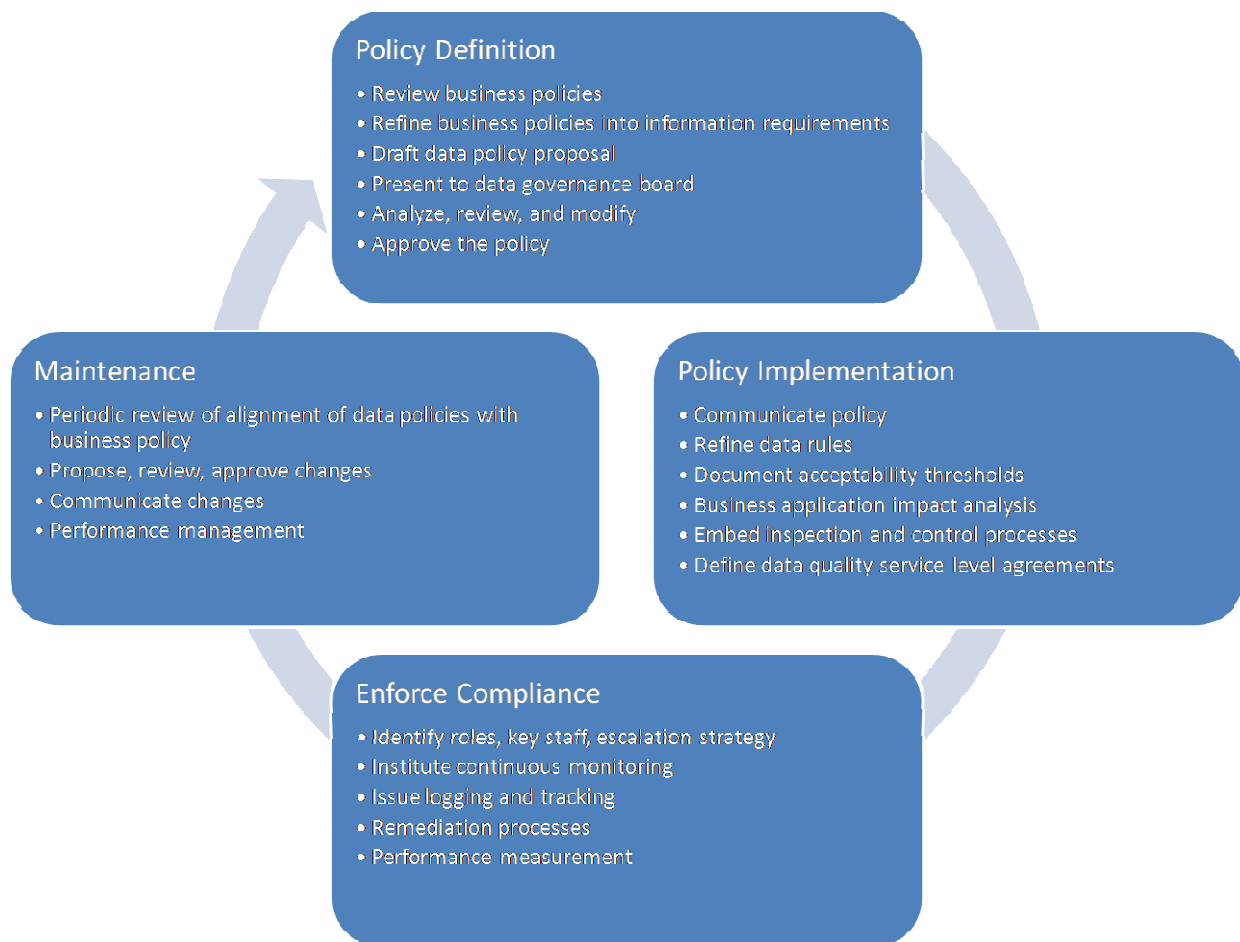


Figure 1: Defining and deploying data policy management

By necessity, operationalizing compliance for data policies involves four stages:

1. **Policy definition and approval** – Business analysts must always be alerted to the introduction of business policies that imply a need for data governance. Business policies can be imposed from outside the organization, such as regulatory compliance, industry standards, or extra-enterprise systemic interoperability requirements. Business policies may also be imposed internally, based on generally accepted operating principles, compensation/benefit programs, supplier management practices, contractor management, among other examples.

It is the job of the business analyst to review business policies to identify the reliance of policy compliance on information, and then draft a data policy that reflects the necessity of observing information requirements to ensure that the business policies are observed. The draft proposal is presented to the data governance board, whose members will review it and iteratively refine it until there is a level of confidence that it accurately represents the business information needs. At that point, the draft proposal becomes a data policy.

2. **Policy implementation** – Once a data policy has been approved, the members of the data governance board must marshal it through the system development life cycle so that it will ultimately be embedded within the application infrastructure. The first stage is communicating the data policy to the collection of data stewards, data and system analysts, and developers. In turn, the data stewards review the data policy along with the original business policies and refine assertions as data rules, such as those shown in Table 1. The business consumers help to define the required level of acceptability for compliance.

The data and system analysts will then review the business processes to assess where the corresponding data rules must be asserted. These locations must be documented, as the developers develop services for inspection that can be directly embedded within the business processes. At the same time, the data stewards facilitate the documentation of a data quality service level agreement that captures the data rules, acceptability thresholds, and the business data producers and consumers who are parties to the agreement in preparation for enforcement.

3. **Enforce compliance** – The data quality service level agreement will identify the key data stewards associated with the operational aspects of data policy compliance. The results of the embedded controls are forwarded to the data steward, who is held accountable for addressing any emergent data issues. Notification and escalation strategies are put in place.

When data issues are identified, they are logged with an incident reporting and tracking system, and the data steward is instructed to analyze the root causes and develop a remediation plan. The data compliance metrics are collected and can be communicated through a data compliance portal or dashboard. If issues are not resolved in accordance with the data policy, they are escalated as described in the service level agreement. The data stewards' ability to ensure compliance is continuously measured and can be reported to the members of the data governance board.

4. **Maintenance** – Business policies are not immune to change, and accordingly, if there are changes then they must be reviewed to determine if there is a need to modify the associated data policies. If so, the draft changes must be submitted for review and approval to the data governance board, and any agreed-to changes have to be communicated and deployed accordingly.

In addition, the performance of the data analyst and data stewardship team is continually reviewed to ensure the effectiveness of the compliance program. Any opportunities for improvement are communicated through the data governance board, and appropriate actions can be taken.

Conclusion Technology - Support for Data Policy Management

When reviewing the processes described in this paper, it becomes clear that the success of implementing a data governance program is critically dependent on the organizational structure of the data governance board, the policies and processes for operationalizing the decisions of that data governance board, and the techniques and methods supporting both.

This is facilitated through tools that support the full life cycle associated with defining, approving, communicating, and fully integrating data policy compliance throughout the application infrastructure. These tools should support data policy management through:

- Data policy definition;
- Metadata management;
- System impact analysis;
- A centralized repository for sharing information about data policies;
- Interoperability with existing tools;
- Documentation of role definitions and associated procedures;
- Documentation of the terms of Data Quality SLAs;
- Guidance for operational roles based on defined policies;
- Services for measurement and monitoring of compliance to data quality rules;
- Preset reports reflecting compliance with data policies; and
- Performance management in the context of expectations defined in data quality SLAs.

About the Author

David Loshin, president of Knowledge Integrity, Inc, (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data quality, master data management, and business intelligence. David is a prolific author regarding BI best practices, via the expert channel at www.b-eye-network.com and numerous books and papers on BI and data quality. His book, "Master Data Management," (August 2009) has been endorsed by data management industry leaders, and his valuable MDM insights can be reviewed at www.mdmbook.com. Learn about data quality improvement and data governance in "The Practitioner's Guide to Data Quality Improvement," (October 2010).

David can be reached at loshin@knowledge-integrity.com.

About the Sponsor

Kalido delivers industry leading solutions for turning bad data into meaningful business information. Developed through years of successful best practice implementations, Kalido's robust, business-model driven information management engine governs data across the enterprise and automatically feeds information to industry standard BI platforms. Kalido® software is implemented at over 250 locations in more than 100 countries with market leading companies. More information about Kalido can be found at: <http://www.kalido.com>.