

Enterprise Data Architecture and Data Governance: Use Metadata to Get to the Starting Gate

Prepared by:

David Loshin
President,
Knowledge Integrity, Inc.
(301) 754-6350
loshin@knowledge-integrity.com

Sponsored by:



Introduction – The Brave New World of Information

Disruptive forces are radically changing the face of enterprise information management. While the prior generation of information management professionals might have been satisfied with augmenting the organization's transactional systems with data warehouses supporting reporting and analytics, today's data practitioner is faced with three factors that are influencing the evolution of the organizational enterprise information management paradigm:

- **Analytics-driven processes:** To remain competitive, companies have developed a thirst for comprehensive predictive and prescriptive analytics requiring orders of magnitude greater visibility and accessibility to data managed within or acquired by the organization.
- **Expanding external user community:** Organizations are increasingly providing their customers with direct access to internal reports and analyses, thereby expanding the number of data consumers beyond traditional administrative boundaries of the enterprise.
- **Broadened data inclusion:** While conventional approaches to operational reporting and analysis may be satisfied using internally-generated data, customer profiling, vendor evaluation, and product analytics are increasingly informed by open data sets or third-party data sets acquired from outside the enterprise.

Business analysts and data scientists are transforming their organizations to be information-driven to create corporate value. Yet as the data horizons widen and both intra- and extra-enterprise data use increases, organizations run the risk of allowing misinterpretations and inconsistencies influence decision-making, impacting the creation of value. Even disagreements about reference data sets will affect the ways that analysts slice and dice data to look for actionable intelligence.

Instituting data governance within the organization will help ensure compliance with business policies as well as reduce the aforementioned risks. But how do we get started with data governance? This paper explores some challenges in bootstrapping a data governance program, and then considers key methods for using metadata to establish the starting point for data governance. The paper will then focus on how metadata management facilitates progress along three facets of the data governance program:

- **Assessment:** Assessment and documentation of the current data architecture.
- **Collaboration:** Creating a collaborative platform for sharing data standards.
- **Operationalization:** Developing the data stewardship policies and procedures to operationalize data governance.

Challenges in Operationalizing Data Governance

Data governance is intended to establish practices and procedures for defining and operationalizing policies to ensure the proper acquisition, management, handling, and

utilization of data by the community of data consumers. Moving these policies into a production mode means deploying the processes and tools for communicating data governance policies, monitoring compliance with those policies, and alerting the right data stewards when the policies are not observed. Yet many organizations are finding that defining data governance policies is often much easier than implementing them through operationalization of processes and procedures. Some of the challenges in operationalizing data governance include:

- **Organic data architecture:** It is unusual for there to be an organization whose application environment did not evolve organically, with systems designed to address operational needs of specific business functions. Correspondingly, few organizations have an enterprise data architecture that guides application design and development in a way that is amenable to governance.
- **Vertical metadata:** A byproduct of organic application development is that the reference data sets and core data models required by each business function are defined within the context of their applications. However, siloed development means that there is limited awareness of overlapping aspects of application metadata that cuts horizontally across the enterprise.
- **Business glossary disharmony:** To ensure consistency and agreement about business term semantics, there is a need for collaboration among data architects and data consumers to ensure harmonization of business terms and data representations.
- **Poorly-defined roles and responsibilities:** One cannot expect to define data policies and expect compliance without articulating what roles are key to enforcement and what the individuals assigned those roles are supposed to do.
- **Need for monitoring and reporting:** Supporting the data stewards requires the right tools for capturing business data rules and monitoring data compliance with corresponding business policies.

While these appear to be stand-alone issues, in practice all of these challenges touch upon the need for visibility into the existing use of information across the enterprise as a prelude to documenting the data architecture. Yet seldom is there a single individual who completely understands enterprise data utilization, creating the need for accumulating intelligence about existing data models, semantic metadata, and lineage maps that show how information is acquired, managed, and importantly shared across the organization.

Information Intelligence

Instituting data governance requires two different aspects of awareness of the enterprise information landscape. The first is the holistic aspect that maps business concepts to information assets that persist within the organization – what are the subject areas (like “customer” or “product”), how they are employed within transaction, operational, and analytical processing, and how those conceptual models are sourced and managed. The second includes the details of the data architecture.

Knowledge Integrity Incorporated

Business Intelligence Solutions

It is the intersection of these two aspects that demand governance, especially when application development has been organic, with little imposed coordination from the beginning. One cannot govern what one is not aware of, and this implies that the starting point for operationalizing data governance is accumulating intelligence about information: data models, data elements, business terms, business rules, etc., as a prelude to developing and managing a governed enterprise data architecture.

The essence of data governance is ensuring that the information assets comply with business policies. In turn, operationalizing data governance relies on developing a uniform data architecture plan that provides the foundation for layering data policies for ensuring usability and quality and consistency. That data architecture plan must embrace the vision for a unified set of conceptual and logical models while integrating the details of the existing data artifacts in use across the organization.

That requires accumulating *intelligence* about the corporate data assets. The use of the term “intelligence” is deliberate in that it encompasses both the collection of data for its utilitarian value and the need for being smart in how that data is collected. Often, the data governance program managers trigger rampant collection of all data set metadata. Yet one must be clever in the way this is done so that the right resources spend the right amount of time in collecting the metadata that will best facilitate the definition, deployment, and enforcement of business data policies.

Using Metadata to Get to the Starting Gate

Operationalizing data governance leverages the practical aspects of the data architecture by enabling business analysts and their IT partners to layer business rules on top of the data models through the adaptation of metadata. Data implications are embedded within the natural language used to define most business policies, but an iterative review and analysis will help identify key terms, determine their association with defined data systems, and enable analysts to define business rules that model the intent of the policy. These business policies link rules about compliance with policies directing how reference data sets are used and shared across the enterprise, how common data element concept definitions and semantics are aligned, as well as how data quality expectations can be configured as data quality rules for validation and quality assurance.

Yet defining the data architecture is a long-term program, requiring commitment over time. Older systems will be assessed to extract their relevant metadata, and some will be selected for renovation and modernization. At the same time, data models and standards will be established to guide new application development to ensure that governance is engineered into the application environment.

Metadata is key to this process, and instituting good practices is critical to get to the starting gate when devising an enterprise data architecture. Recognize that the right tools must be employed to facilitate the development of the data architecture, especially when it comes to

using a metadata tool to help operationalize data governance. Some of those features include:

- An integrated **business glossary** used as an enterprise resource for managing business term collection and documentation of their authoritative definitions and semantics.
- Integration with **data modeling** tools to integrate data elements with conceptual, logical, and physical models across the application landscape.
- Specific metadata **extensions around governance** (such as “user defined properties”) enabling data consumers and managers to define reusable properties that can be standardized and used across the enterprise.
- **Role based access** to the metadata repository enabling individuals taking on the different data stewardship and governance roles to facilitate enterprise information management best practices.
- **Forward engineering** metadata to guide assessment of aging systems and applications, the interpolation of inherent structure and architecture, and the ability to automate the generation of evolved data models (and corresponding data definition language) targeted at specific databases.
- **Collaboration and visualization tools** that enable publication of definitions, standards, models, and data policies for review, interpretation, discussion, and agreement.

Summary

Tools alone will not address the need to operationalize data governance. Metadata tools help get you to the starting gate, but a holistic approach will blend technology with people and process to define the operational roles and responsibilities, that will engage the business users to partner in the process. Some of the key roles and responsibilities include:

- **Data governance manager** – This individual’s responsibility goes beyond running data governance council meetings. The operational role of the data governance manager is to ensure that there is a thoughtful plan for accumulating metadata artifacts, vetting business term definitions, and documenting how data element concepts are applied across the organization’s business processes.
- **Data steward** – As the practitioner with the greatest degree of interface between the data management team and the business users, the data steward must be able to differentiate relevance among the collected metadata, engage the business users for clarifications when necessary, and manage the harmonization of business information concepts on an ongoing basis.
- **Business analyst** – The business must be engaged and involved in the translation of business policy into data rules. This implies collaboration between the business analysts and the data stewards in isolating data dependencies and documenting business rules.

Knowledge Integrity Incorporated

Business Intelligence Solutions

- **Data quality analyst** – Business policies imply data policies, which can then be articulated as rules that impose quality and usability constraints for data. The role of the data quality analyst is to ensure that those rules are defined, reviewed, and put into production.
- **Metadata analyst** – The metadata analyst facilitates the deliberations over definition, semantics, and the practical aspects of both documentation of the metadata and its institutionalization within the application design processes.
- **Data modeler** – Last, but certainly not least, the data modeler puts the pieces into place by ensuring that conceptual and logical models are adequately adapted into both renovated and newly-designed application designs.

The organization's data architecture cannot be developed in a vacuum away from the data governance professionals. The two go hand-in-hand: data architects need to be aware of how governance influences the architecture, and governance practitioners need to be aware of how the data architecture enables compliance. Collaborative metadata platforms make this happen, and ensuring that your team is equipped with the right tools will simplify the data governance operationalization process.

Knowledge Integrity Incorporated

Business Intelligence Solutions

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, with numerous articles published at searchBusinessAnalytics.techtarget.com, and numerous books and papers on Big Data, Analytics, Business Intelligence, Data Warehousing, data governance, and data quality. Visit <http://dataqualitybook.com> for more of his insights. David can be reached at loshin@knowledge-integrity.com.

About the Sponsor

Sponsored by CA ERwin® Modeling — providing a collaborative data modeling environment to manage enterprise data through an intuitive, graphical interface. With a centralized view of key data definitions, you can have a better understanding of corporate data, managed in a more efficient and cost-effective way — placing ERwin at the Center of Data Management. For more information visit www.ERwin.com