Operationalizing Data Governance with Living Metadata

Prepared by:

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

Sponsored by:

erwin

As combinations of both internal and externally-imposed business policies imply dependencies on managed data artifacts, organizations are increasingly instituting data governance programs to implement processes for ensuring compliance with business expectations. One fundamental aspect of data governance involves practical application of business rules to data assets based on data elements and their assigned values. Yet despite the intent of harmonizing data element definitions and resolution of data semantics and valid reference values, most organizations rarely have complete visibility into the metadata associated with enterprise data assets.

That means that most data governance activities are often tightly coupled with assessing, understanding, and solidifying the corporate view of metadata. In turn, many data governance managers feel that potential metadata awareness gaps are easily addressed with broad-scale inventories of data assets and interpolation of their associated metadata. It is surprising, then, to recognize that most metadata projects are not successful!

The failure of many metadata management activities can often be boiled down to a single root cause: the presumption that the collection and documentation of system metadata constitutes the completion of the process. While many metadata projects begin with a bang, with tasks initiated to scan through all enterprise data assets to profile and amass their structural metadata, these tasks are often brute force activities that largely constitute busy-work instead of thoughtful assessment.

Is it critical to recognize that as the business continues to change as time goes on, the data and metadata must continually be updated and reviewed. The inability to continually use and maintain the enterprise metadata inventory will consign the repository to an early demise, as its usefulness will continue to degrade. This paper explores best practices in maintaining a "living" metadata management platform. In turn, we discuss the key factors in ensuring timeliness and utility of an enterprise metadata resource and how collaboration and utilization will enhance data governance and stewardship.

Metadata: Not Once-and-Done

Enterprise business metadata encapsulates more than just data element names and their structural details. Semantics, context, and content are all relevant within the business data utilization frameworks. For example, for analytical reporting using data warehouse models, the values in a table's column represent the facets of its dimensionality, and aggregation and presentation both rely on meanings and the fidelity of the data element's values.

And knowing that data elements are constantly subject to change over time, ensuring that governance can be properly deployed means having an active and living metadata asset. That suggests three key guidelines that need to be followed:

- *Metadata should be used* – A metadata management tool should not be a "roach motel," in which metadata checks in, but does not check out. If the metadata repository is "input-only," it means that what is being collected and managed is

effectively irrelevant. Ensure that team members are aware of, and have access to use the metadata for data governance purposes.

- *Metadata must support stewardship* – The metadata program should provide a control panel through which enterprise semantics are continuously shared, communicated, debated, and ultimately agreed to. This supports a proactive approach to implementing and ensuring compliance with data policies.
- *Metadata must be kept up to date* – As perceptions and interpretations of information assets change, the underlying semantics must be also continuously monitored for changes so that the metadata must be kept up to date. Because business contexts evolve and policies change, governance relies on currency and accuracy in the underlying metadata.

Metadata management is not a "once and done" prospect, but rather that it involves programmatic activities that must be continuously monitored, managed, updated, and governed.

Building Upon the Foundation

Data governance requires the definition of data policies to meet business expectations. Once business glossary terms have been captured and definitions documented, the organization is able to more effectively specify the data policies whose observance will indicate compliance with business expectations.

For example, consider the need to comply with a regulatory policy for protection of personally identifiable data. The criteria for compliance imply that in any corporate data asset, it must be clear which attributes contain information considered to be personally identifiable data. This information, of course, is critical metadata, and the definition and implementation of the business policy requires data policies that refer to the metadata and the actual data values.

If you reflect on that concept, you will see that a data policy is fundamentally composed of business data rules that would also be characterized as metadata. In turn, each data policy represents assertions about the values and dependencies among the attributes and records of managed data assets. This implies that the definition of a data policy is effectively a metadata exercise. The metadata provides the means by which the direct connections between the specification of a data policy and its implementation, suggesting that the stewardship and governance of implemented approved data policies hinges on successful metadata management.

Benefits of Metadata-Driven Data Governance

Establishing a data governance program that is motivated and driven via a metadata program provides a number of valuable benefits to the organization. One of the greatest challenges of instituting data governance is its frequent perception of a bloated bureaucracy

imposing extraneous rules, and as a necessary evil to maintain control over the data. In the best scenarios, however, organizations embracing data governance see the positive side effects including:

- Increased trust in enterprise data as data stewards are able to more rapidly identify and resolve data issues;
- More precise clarity surrounding regulatory compliance efforts in relation to oversight of data element definitions;
- Lowered operational costs attributable to analysis of data issues and the time spent in their necessary remediation;
- Reduced effort in assuring levels of data quality as a result of proactive validation and monitoring; and
- Increased data utilization across the enterprise stemming from all of these aforementioned benefits.

These effects are perpetuated by well-organized data modeling and holistic approaches to metadata management that expand beyond an ethereal scan of database catalogs or cursory glances at the results of a data profile. Clearly-defined data element concepts and semantics that are shared and actively kept up-to-date ease the perceived onerousness of data governance. Instead of restricting the organization, this approach actually frees people from the constraints of ambiguous interpretation and variance in understanding.

Establish processes for "living metadata" as best practices that can be implemented and operationalized as a core component of data policy management. Active, ongoing metadata management can ensure consistency in analysis and process execution that eliminates a large part of the overhead of reactive tasks performed to identify, reverse engineer, and address data problems.

Planning for Living Metadata

What are the core components of a "living metadata" program? If you are thinking about the benefits of transitioning from a brute-force metadata scan to a more thoughtful and insightful approach to harmonizing context and semantics across the enterprise, here are some facets of a metadata solution to consider:

- **Establish an enterprise platform**: Seek out vendor products for metadata management that enable a repository for business glossary and data element definitions that can be shared across the organization and provides easy accessibility for both technical and business users.
- **Enable collaboration**: A solution that provides a community-based platform for knowledge sharing and transfer should also support collaboration among communities of active participants to share their thoughts and inputs.

- **Proactive notification**: Provide methods to alert interested parties about adjustments and changes to business terms, definitions, data elements, or data models.
- **Data protection**: To prevent exposure of sensitive information to the wrong individuals, provide a role based access mechanism for data protection. An added benefit is provided when data security tools can be configured using metadata.
- **Standards**: Data policy compliance can be enhanced when the metadata product allows for user-defined properties for establishing standards.
- **Forward engineering**: Evaluating current state analysis, enabling what-if analyses for future visions, and automatically generating models supports impact analysis, renovation assessments, as well as modernization.

Products that have these features will motivate the maturation of processes supporting "living metadata" management, enable data policy definition and compliance, and help establish an evolutionary approach to instituting enterprise data governance.

Putting It All Together

There is no doubt that the data governance lifecycle is intricately entwined with a successful metadata strategy. The organic nature of enterprise application development has left a trail of vertical systems with little or no horizontal integration, leaving business users to grasp at straws when it comes to harmonizing their use of organizational data. As we discussed in our earlier paper, we can use metadata to make the data meaningful by organizing the practical aspects of enterprise data assets and operationalizing metadata utilization by leveraging an integrated business glossary, logical data models, and collaboration tools for knowledge exchange.

This sets the stage for a more comprehensive metadata strategy to drive the analysis of requirements and definition of enterprise data policies. As we discussed in our second paper, using metadata management as a process tool helps to formalize the semantics associated with business policy compliance, including aspects of data protection, data quality, data model renovation, as well as newer approaches to information utilization via data lakes and other big data paradigms.

Finally, embrace metadata management as a core competency for the sustainability of your data governance program. As more business users understand how their participation not only enhances the governance activity, it also simplifies its movement into production, the stakeholders will recognize how operational data governance will benefit the business and create significant value.

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 article 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