

DATA CATALOGS, BUSINESS GLOSSARIES, AND DATA GOVERNANCE FOR CUSTOMER BI ENABLEMENT

EMPOWERING SOPHISTICATED DATA CONSUMERS

For more than three decades, organizations have relied on data warehouses to support business information consumers' needs for descriptive analytics to help inform about the current state and to help influence ongoing business decisions. And although organizational analytics programs are increasingly augmented with machine learning and advanced algorithms for predictive and prescriptive analytics, the ongoing need for business intelligence (BI) supporting descriptive and operational analytics applications will remain. What has changed over time, though, is the increasing sophistication of the data consumers and their growing awareness of the breadth and depth of corporate data assets.

Business BI consumers are no longer the “customers” of the data warehouse team – they are their partners. And this suggests that the best way to empower business information consumers is to provide accessibility to organizational data configured in ways that both simplify the production of analytics and speed time to knowledge. Empowering the data consumers requires some key aspects of operational data governance, including:

- Enterprise data intelligence,
- Collaboration among the data architects, data stewards, and business data consumers,
- A searchable catalog of enterprise data assets, all of which support the simplified engineering of a business intelligence solution.

In this paper, we consider the historical approaches to developing data warehouses and how growing end-user sophistication has increased the criticality of ensuring data clarity and consistency of the semantics across a variety of data sources. The paper then discusses the concept of enterprise data intelligence, and how that is facilitated through the use of a data catalog. Automated data catalogs are excellent for data discovery and documentation. However, they are enhanced with corporate knowledge that data architects can use to supplement and expand enterprise data awareness, and this paper examines how a modern enterprise data catalog will capture and document a broad array of metadata that help the data architects and practitioners enable data consumers. When considering how data catalogs enhance the context for reporting and analytics, we identify two insights that can influence the modern business intelligence stack. Finally, we provide some recommendations associated with the characteristics of data intelligence tools and look at ways that data models, data governance tools, and data catalogs should interoperate to help re-envision the ways that data governance can drive business intelligence solutions.

PRACTICAL DRIVERS FOR BUSINESS INTELLIGENCE

In recent years, the evolution of underlying data warehousing, data integration, and data preparation technologies has facilitated the evolution of the enterprise data warehouse from a conceptual “data dump” into a thoughtfully structured set of capabilities for data analyst enablement. Modern reporting and analytics ecosystems enable technologists to more effectively solicit end-user requirements, understand data consumers’ business problems, and assemble customized application environments that the data consumers can use to explore solutions to those problems.

But increased end-user sophistication implies corresponding sophistication for the data practitioners in terms of soliciting data consumer needs and expectations and synthesizing the collected requirements into a design for a business intelligence solution. As the number of potential data sources increases, some particular challenges include:

- **Ensuring metric clarity:** The team that is supporting the business analysts needs a clear understanding of what information the business user needs, with clarity, context and without ambiguity, and must have a well-defined process and corresponding tools for soliciting and documenting requirements and critical business terms.
- **Identifying and managing information sources:** Once the operational measurements are defined, the technical team must identify which organizational information sources can be used to architect the BI solution and develop an inventory of those data sources for the benefit of the downstream consumers. This inventory would help the end-users identify the best source of information to help develop their analytics applications.
- **Ensuring consistency:** Different data sources might represent the same or similar concepts in slightly different ways. The data architects must be able to provide harmonized definitions for the range of business terms and ensure consistency in both definition (i.e., how the business terms are defined) and semantics (i.e., how those terms are used) so that the data consumers’ needs are met.
- **Maintaining relationships among terms:** Business users are increasingly aware of the relationships among and between different pieces of information. For example, a “person” or an “organization” are different types of “parties,” and the term “client” and “customer” may share the same meaning. Data architects must become trained in discovering, documenting, and managing these relationships.

ENTERPRISE DATA INTELLIGENCE AND THE DATA CATALOG

The key to addressing the aforementioned challenges is knowledge – knowledge about the enterprise data landscape, knowledge about the processes and procedures for accessing and consuming information, and knowledge about the “data mechanics” employed to produce ultimately a usable reporting and analytics environment. There are three aspects to managing this enterprise data knowledge: the collection of data intelligence, its organization and enhancement, and facilitation through use. Not surprisingly, all three of these aspects are enabled by data governance procedures and techniques for enterprise data intelligence, such as:

- **Surveying** the array of data storage environments across the enterprise to identify and inventory data assets.
- **Profiling** identified data assets to collect structural metadata, assess data quality, identify any semantic tags, and surface any potential issues hidden within the data.
- **Documenting the object metadata** including data owner/producer as well as aspects of the production process and lineage of the data.
- **Classification** of data elements by assessing the meaning of each structural element in the data asset and attach to the appropriate business term that reflects that meaning.
- **Populating an index** and providing a searchable semantic registry of business glossary terms and data elements that provides insight into the inventory of available data sources.
- **Auditing** the quality and usability of each data element in the data asset using defined and captured data quality rules, which then enables operational reporting of the actual performance measures for data quality compliance.

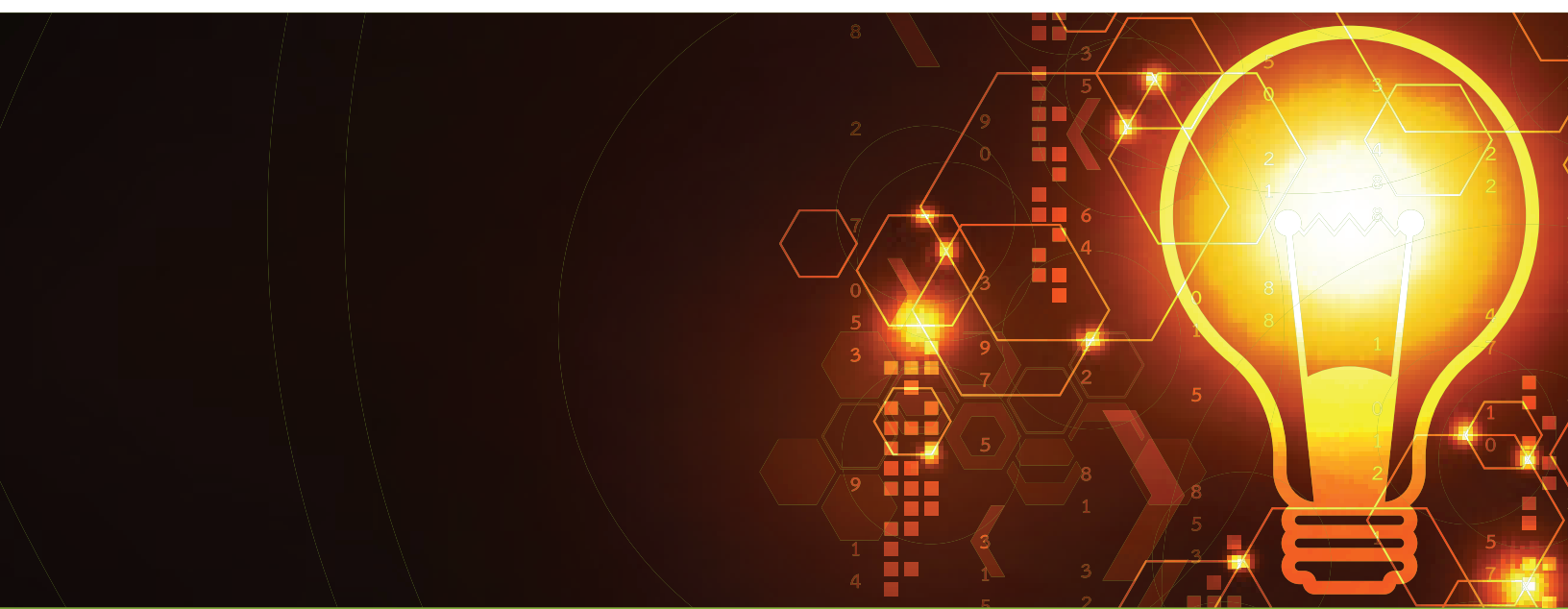
The result of these techniques is the construction of an enterprise data catalog that provides several services and benefits to the data consumer communities. This data catalog is a repository for a data asset inventory that serves as an enterprise resource through which all data architects and data consumers can share knowledge. It documents the high-level object metadata associated with each data asset, enumerates data elements, and maps those data elements to a semantic business glossary. In turn, this information can be absorbed into an intelligent search capability that allows data consumers to browse the array of data concepts, help determine which enterprise data assets are best suited for any specific business scenario’s reporting and analytics needs.

ENHANCEMENT AND ORGANIZATION: LEVERAGING THE CORPORATE KNOWLEDGE BASE

The introduction of data discovery and surveillance tools is certainly a boon to the organization, especially those with limited (or no) existing data architecture or data stewardship. These tools are key to establishing an enterprise data intelligence baseline by listing data assets and inferring core data asset metadata. And while automated tools can contribute to the collection of knowledge about data that presumably represents the structure of a variety of corporate information assets, recognize that automation is never going to be a replacement for “ground truth” knowledge.

That knowledge is supported through existing data governance and data stewardship best practices associated with data modeling and data architecture. A mature data model contains the logical model which is a well-considered understanding of the information of the organization expressed in business language and arranged into a structured model. For data assets under the control of data architects, physical models will be produced as a faithful snapshot of the elements and structure of the data asset expressed in technical terms. Data Architects will map the physical layer to the logical. This mapping is valuable knowledge concerning the content and meaning of data assets. The existing data architecture provides a valuable knowledge base that combines the collected conceptual, logical, and physical data models with the corporate lore that data architects, modelers, and developers have to develop and maintain the enterprise data environments.

The data architects that are familiar with this corporate lore about existing conceptual, logical, and physical data models (and their corresponding artifacts and tools) can provide a qualitative dimension that organizations can leverage as part of the enterprise data intelligence collection process. Their knowledge not only enhances what automated tools can discover, but it also helps in fleshing out undiscovered details associated with data assets such as business terms and corresponding definitions, mappings between those business terms and data asset structure, classifications associated with business domain, operational use, and (with increasing criticality) characteristics of data sensitivity of each data asset’s contents.



CHARACTERISTICS OF A MODERN ENTERPRISE DATA CATALOG ENVIRONMENT

These aspects of collection and enhancement of data intelligence are meant to support the aspect of data use, and a modern application stack for enterprise data intelligence that will adequately facilitate data usability will most likely incorporate these capabilities that support all three aspects:

- Data asset **structure metadata management**, including data profiling to assess data asset structure and associated structure, existing conceptual, logical, and physical data models, as well as details of the table layouts, data elements, sizes and datatypes, and other aspects of the structure of data assets.
- Data asset **object metadata management** including data source, owner, responsible parties, size, refresh rate, and other aspects of the object itself.
- Data asset **business metadata**, including a business glossary, definitions, and increasingly facets of semantic models such as defined logical namespaces, semantic graphs capturing taxonomies and relationships, and classifications
- Data **intelligence utilities** that can help find data assets, perform automated discovery help identify and inventory distributed and heterogeneous data assets across a (potentially hybridized) data environment, assign classifications, document data lineage, and assemble an intelligent search index.
- A **data portal** that provides visibility into the data catalog's inventory and includes an intelligent index that supports searching and data browsing.
- Support for **data governance** that supports data quality policy management, data classification and tagging, data protection policy management, and access controls associated with users, roles, and data asset classifications, among other aspects of stewardship and management.

BRINGING IT ALL TOGETHER

Examining the modern context for business intelligence, reporting, and analytics yields two valuable insights about the value of data governance. Together, these insights suggest an interesting conclusion about expanding the scope of the modern BI software stack.

The first insight is that the need to balance the increased data awareness on behalf of the information consumer communities, the voracious appetite for more sophisticated reporting and analytics, and the amplified demand for leveraging enterprise data assets reinforces the value of good data governance practices that rely on the use of a data catalog. Instituting repeatable processes in managing the enterprise information assets and capturing the details of data location, lineage, source mappings, classification of data sensitivity, and requirements for data protection establishes a level of trust in data accessibility, quality, and usability. The data catalog better positions your organization for BI/analytics and enables analysts and other data consumers to make good business decisions.

The second insight is that a data catalog's automated capabilities are significantly enhanced through human interaction. Data architects can collaborate to verify the data in the data catalog and validate consistency with documented conceptual, logical, and physical data models. Furthermore, data architects can employ the information held in the business glossary and the corresponding mappings within the data catalog to drive the development of the appropriate information models that support the design, development, and delivery of end-user BI applications. And importantly, data consumers can search and browse the business glossary and the data catalog to identify corporate data sources that can support their BI, reporting, and analytics needs.

But what is interesting is that more precise, accurate, and comprehensive information and metadata in a shared data catalog ultimately simplifies the data architect's job while empowering the users in the data consumer communities. The data catalog becomes the conduit for self-service, as the data consumers can find the right sources for their needs. Additional utilities that support integrated data warehouse automation can leverage the accurate structural and object metadata data catalog's inventory of enterprise data assets and simplify the speedy implementation of data marts and data warehouses. In other words, a comprehensive data warehouse production solution can leverage enterprise data intelligence (embodied within a business glossary and data catalog) as input to a data warehouse automation (DWA) system that simplifies customer enablement by helping sophisticated data consumers understand the content of that available data assets and determine the information they are looking for instead of just relying on a massive data dump pulled from a variety of operational systems.



NEXT STEPS

A quiet revolution is underway in the world of reporting and analytics, inspired by the increasing compatibilities between data, models, data catalogs and business glossaries. When organizations institute effective data governance policies and practices, the roadblocks and barriers to the success of the past quickly diminish while customer enablement and self-service come to the forefront. Empowering communities of increasingly sophisticated data consumers demands the kind of enterprise data intelligence that leverages collaboration, interoperability, and data awareness. This revolution is powered by core technologies that need to work together to allow the data architects to support the data consumers best. If the objective is to bridge the gap between data models, business glossaries, and data catalogs, look for products that can:

- Align with industry standards for representation and management of metadata (such as the ISO 11179 standard) to simplify interoperability with other data management tools.
- Allow Data Architect access to Business Glossaries to use in the classification of data assets.
- Harvest business terms and phrases from conceptual and logical data models and surface them for definition and documentation.
- Simplify and automate the generation of logical data models using semantic models and business terms from a business glossary.
- Integrate with data governance tools to synchronize business terms and link data policies to data design and implementation within a harmonious ecosystem.
- Manage a variety of relationships among business glossary terms manifested in taxonomies and ontologies that reflect the underlying business realities.
- Provide a user-friendly interface for searching for business terms, providing fundamental attribution of data elements, and effectively visualizing relationships, taxonomies, and ontologies.
- Integrate with data warehouse automation tools to rapidly forward engineer, design, and implement new systems for business intelligence, reporting, and analytics.

These types of products will help in re-envisioning the ways that operational data governance best leverages enterprise data intelligence to simplify the creation, implementation, and management of business intelligence solutions.

IDERA'S SOLUTION

ER/Studio Enterprise Team Edition

Design and share data models and metadata across the enterprise

ER/Studio Enterprise Team Edition is the leading business-driven data architecture solution that combines multi-platform data modeling, business process modeling, and enterprise metadata for organizations of all sizes. With an extensive feature set, the ER/Studio suite provides robust logical and physical modeling with ER/Studio Data Architect, business process and conceptual modeling with ER/Studio Business Architect, business glossaries with ER/Studio Team Server, and more, to build the foundation for data governance programs.

Start for FREE

Request a DEMO

