



## Using metadata to enable data governance

*Beate Porst  
IBM Software Group  
Information Platform and Solutions*

---

<b>Contents</b>
<b>2 Introduction</b>
<b>3 Why is data governance important?</b>
<b>5 The role of metadata management for data governance</b>
<b>7 Metadata helps bridge the semantic gap</b>
8 Common controlled business vocabulary
11 Data stewardship
12 Instant business understanding
<b>13 How metadata supports compliance and risk mitigation</b>
15 Data lineage
16 Impact analysis
17 Customized reporting and search
<b>19 Data governance helps preserve data quality</b>

## **Introduction**

Organizations today understand that data is a fundamental asset to their business. Yet they are faced with the challenge to derive business value from data spread across many different systems—data that is often redundant and exists in numerous formats.

On the technical side, IT stakeholders are challenged to extract meaning from an increasing number of systems to provide comprehensive viewpoints. Constrained by small update windows and tools that often do not work together, they may leave existing data flows unexamined, which can result in data duplication and inconsistencies. On the business side, business stakeholders are challenged to make accurate business decisions when they do not fully understand or trust the data they are provided.

Because IT complexity and data quantity will continue to grow, organizations need data assurance methods to help improve their ability to adapt to system change, mitigate risk and deliver strategic information to business users. A unified metadata management strategy and a fully integrated metadata toolset are proven methods for supporting business and technical users in their efforts for effective collaboration.

Information is a powerful tool—but its true business value relies on the delivery of the right information to the right people at the right time. This white paper explains how the unified metadata management architecture and the metadata-driven tools of IBM® Information Server aid organizations in their data governance programs and collaboration efforts. This paper also illustrates how the metadata tools of IBM Information Server deliver data lineage and traceability to support change activities and to help organizations meet regulatory compliance requirements.

### **Why is data governance important?**

Today, an organization's success largely depends on an intangible asset: data. Data is the foundation upon which critical business decisions are made. Yet the value of data is often at great risk because of insufficient data management processes and standards. Poor data management ultimately causes an increase in data redundancy and degradation of data quality, which can eventually result in poor business decisions and greater exposure to compliance violations. On the other hand, the ability to leverage reliable, complete and trusted information when needed can greatly assist organizations in providing quality analysis results to business users. Those quality results can help an organization become more agile and responsive to the various demands of the business and ultimately, more competitive in the marketplace.

In many organizations, data is stored in numerous formats and spread across many different and often siloed systems. To maximize information value, organizations need to understand the relationship between the different data sources and how they are being used across the enterprise. Customers often have difficulties identifying the right data, its trustworthiness and validity due to missing or incomplete documentation of these systems and their relationship to one another. These indicators often showcase a deficiency in effective alignment with data governance programs.

To enhance data quality, availability, integrity and auditability, organizations need to establish standards, policies and processes for data usage and management. Data governance has emerged as an important discipline for doing just that: providing a framework for customers to manage information assets to drive organizational efficiency. Additionally, data governance fosters the understanding and management of data from both a top-down business and bottom-up technical point of view.

Through an effective data governance program, IT managers and business users alike will become more effective in business measures including risk mitigation and value creation. From a data perspective, these initiatives include data provenance or lineage, stewardship and enhanced collaboration through effective use of technology. This white paper will show how these initiatives are fueled by metadata throughout the enterprise data integration lifecycle.

### **The role of metadata management for data governance**

A data governance program can be implemented to organize important measurements in a data-intensive landscape. Controls for this type of information are often rooted in the creation, use and lifecycle of the data and are fueled by both the decisions made and the data itself.

The term metadata describes the data used within an organization. It describes where the data originated, the person or persons defining it, the ownership or stewardship bodies responsible for it and the state or usage specifications, as well as potentially other annotations in more optimized organizations. While metadata concerning a single data asset is important, it will not allow assumptions to be made regarding data quality, currency or integrity with respect to the entire organization. Understanding the bigger picture of how data traverses through different systems and its usage requires a holistic approach, typically referred to as *metadata management*.

The role of metadata is a fundamental underpinning of data governance. This is reflected in the Data Governance Blueprint created jointly by customers and the IBM Data Governance Council.<sup>1</sup> This blueprint defines a maturity model, measuring data governance competency and providing assurance based on 11 disciplines. One of the paramount disciplines is Metadata/Business Glossary.

Effective utilization of metadata and business glossaries—or enterprise data dictionaries—enables capabilities that are significant for day-to-day business activities including:

- **Consistency of definitions.** *One department refers to “revenue;” another refers to “sales.” Are both referring to the same activity? One subsidiary unit talks about “customers;” another about “users” or “clients.” Are these different classifications or different terms for the same classification? Effective management of business metadata can help ensure that the same descriptive language applies throughout the organization. The glossary for IBM Information Server enables business users to author and own business metadata.*
- **Clarity of relationships.** *Analysis of metadata illuminates the associations and interactions among all components of the warehouse environment: business rules, tables, columns, transformations and user views of the data, to name a few. By clarifying relationships throughout the integration project, managed metadata enables IT managers and business users to understand the larger perspective while remaining focused on departmental issues—leveraging the enterprise meaning of the data assets to help enable the broadest understanding of the assets within the enterprise.*

IBM Information Server is a revolutionary new software platform from IBM that helps organizations derive more value from the complex, heterogeneous information spread across their systems. It enables organizations to integrate disparate data and deliver trusted information wherever and whenever needed, in line and in context, to specific people, applications and processes. IBM Information Server helps business and IT personnel collaborate to understand the meaning, structure and content of any type of information across any type of source. It provides high-performance capabilities for cleansing, transforming and delivering this information consistently and securely throughout the enterprise so it can be accessed and used in new ways to drive innovation, increase operational efficiency and lower risk.

Specifically for metadata management, IBM Information Server includes tools that can help deliver not only consistent data language and definitions and clear views of data relationships, but also data lineage reports and data tracking. IBM Information Server includes four major metadata components:

- **IBM WebSphere® Metadata Server:** *A metadata services framework and a dynamic metadata repository*
- **IBM WebSphere Business Glossary, IBM WebSphere Business Glossary Browser and IBM WebSphere Business Glossary Anywhere:** *The business user's interfaces for authoring, browsing and initiating searches from external applications*
- **IBM Metadata Workbench:** *A visual tool to track, analyze and assess data throughout the enterprise*
- **IBM Information Server bridges:** *Metadata interchange interfaces*

### **Metadata helps bridge the semantic gap**

The metadata semantic gap refers to a lack of common understanding between business and technical users. Data governance depends on a common business vocabulary. As a first step to facilitating this common understanding of information across these two roles, the organization must establish a baseline across two types of metadata:

- **Business metadata** *is critical for end users—or consumers—of information. It allows them to be confident that the data they rely on for making business decisions is exactly what they expected. It provides business users a roadmap for navigating enterprise data and provides context for interpreting the data.*
- **Technical metadata** *helps companies understand what information they have today and assess the reliability of that information. It streamlines development efforts by providing technical users with information about the data elements and how they are implemented currently across various systems. Additionally, this metadata can help improve auditability and provide visibility through impact analysis and direct lineage reporting.*

WebSphere Business Glossary helps organizations create, manage and share an enterprise-wide controlled vocabulary that acts as the common language between business and IT. This is a critical step in better aligning the efforts of the technologists with the goals of the business.

### ***Common controlled business vocabulary***

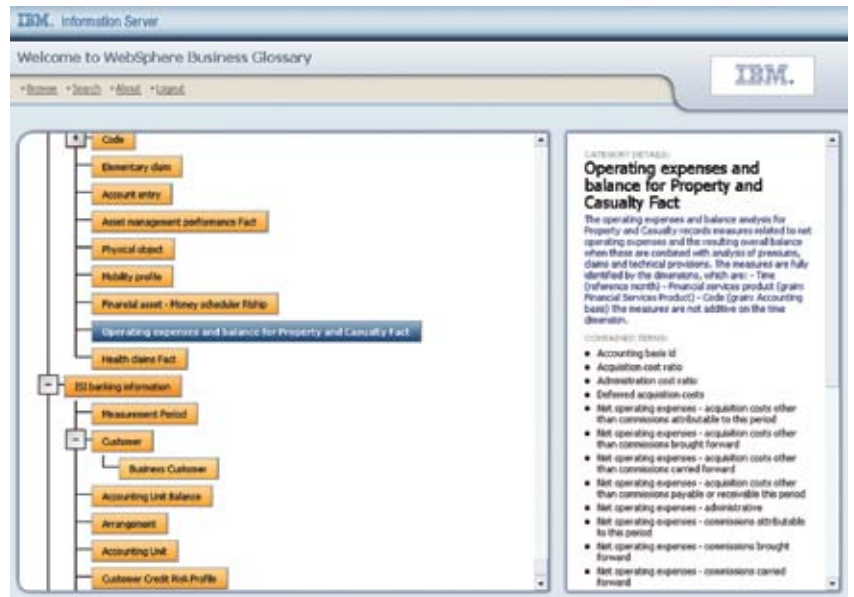
Many organizations wrestle with inconsistency of business definitions across the enterprise. This may be caused by events such as mergers and acquisitions or by systemic siloed approaches to the definition of a common business vocabulary. The root cause is often attributed to the absence of an enterprise-wide data governance and stewardship program.

The standardized definitions and business concepts in an organization comprise a common controlled business vocabulary. A business glossary is created by organizing the controlled vocabulary (business definitions) in a hierarchical structure and establishing relationships between definitions and classifications to its technical metadata.

WebSphere Business Glossary provides a Web-based portal for the definition, management, search and exploration of business vocabulary and its rules and relationships. To ensure high quality, only authorized data stewards can use the administrative functions within WebSphere Business Glossary to create and manage the glossary. Any business user, however, can use the intuitive WebSphere Business Glossary Browser to search and explore the meanings defined in the glossary (see Figure 1).



Figure 1: The graphical explorer view of IBM WebSphere Business Glossary Browser



All too often, tools intended to aid users in day-to-day activities fail to be adopted because of usability barriers caused by over-engineering. With this in mind, the WebSphere Business Glossary Browser design was based on two principles: “simplicity lasts” and “cut right to the chase.” Figure 2 shows an example of this streamlined design: the browser’s entry page from which business users can search for categories and terms in free form.

WebSphere Business Glossary bridges business requirements and technical implementations. Companies can accelerate the process of developing a glossary by leveraging IBM Industry Models as the foundation for their common vocabulary rather than building a glossary from scratch.

Figure 2: From the IBM WebSphere Business Glossary Browser entry page, a business user can search for categories and terms in free form



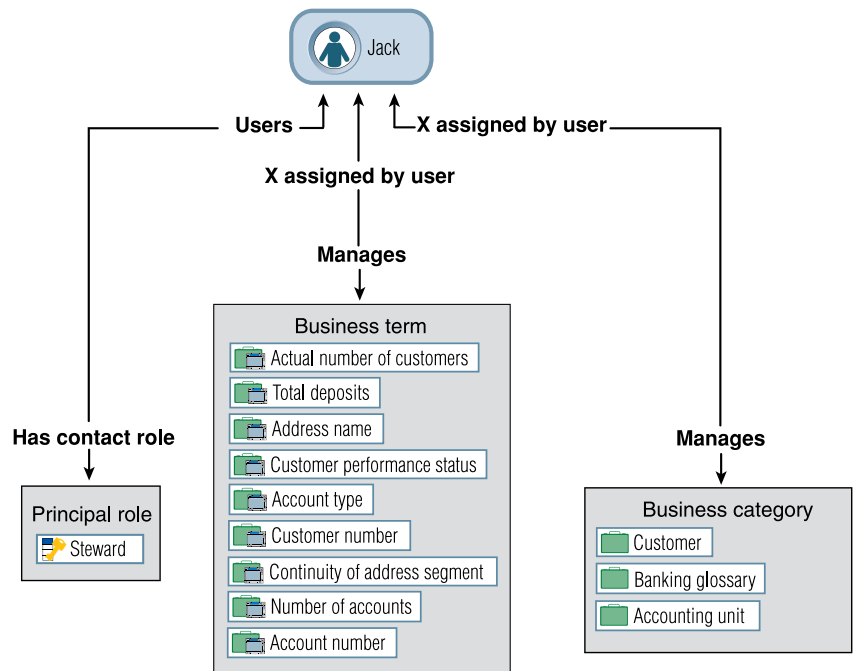
To address the Information Lifecycle Management (ILM) entry point, IBM provides a methodology to help organizations define a storage infrastructure and service levels appropriate for information at each phase of its user access lifecycle. Using this methodology, companies can apply higher standards to active information that is more critical to their business and lower standards to information that is less active and less critical to access instantaneously. It is important to consider risk, compliance, operational efficiency and cost effectiveness in order to execute this process properly. Through this entry point, businesses of all types can begin to classify their data and take a vital step forward in helping to align the value of their information with the most appropriate IT infrastructure.

**Data stewardship**

Data stewardship is the management of data resources throughout the data lifecycle. A formal data stewardship committee with executive sponsorship is a crucial prerequisite to an effective data governance program. Typical activities performed by the committee include the creation of business definitions and rules, the definition of domain values and the establishment and validation of data quality rules. Data stewards bear responsibility for the accuracy, accessibility and currency of metadata, but they do not own the data. Rather, data stewards act on behalf of the committee, which is driven by the business.

IBM Information Server supports the concept of data stewardship and allows users to set and retrieve stewardship information for all data assets. Figure 3 shows a graphical representation of a data steward and all data assets for which he has responsibility.

Figure 3: A graphical representation of a data steward and associated data assets



WebSphere Business Glossary administrative functions supporting stewardship include:

- **Managing business categories**—Users can manage business categories and the objects in them
- **Managing business terms**—Users can manage business terms and the repository objects they identify
- **Managing stewards (people)**—Users can manage single-person data stewards (users with stewardship responsibilities over metadata objects in the repository)

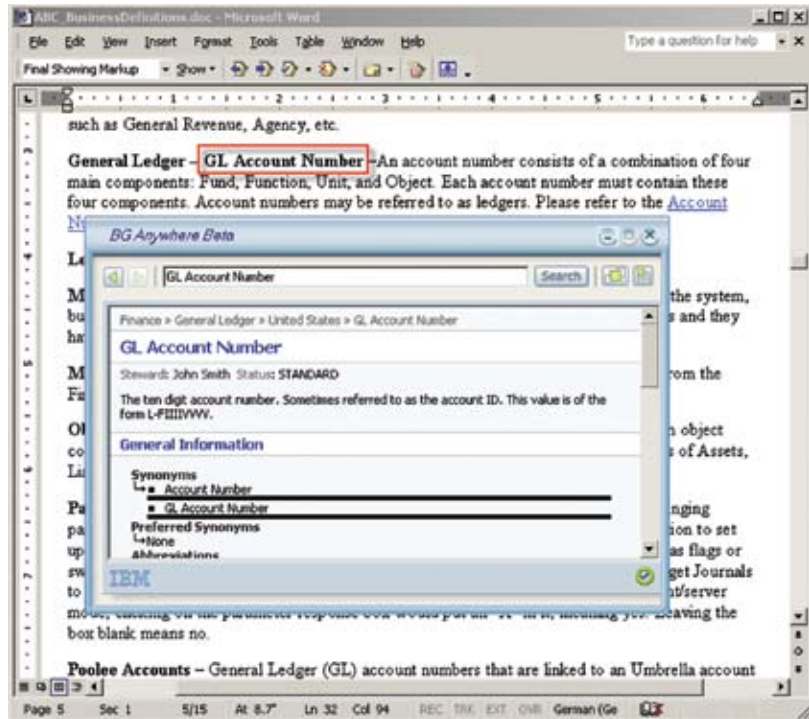
### ***Instant business understanding***

Business understanding is crucial, yet time pressure often keeps business users from leveraging available resources. For example, you read an e-mail or a white paper and come across a term or phrase that makes you wonder how this is being defined or used in your organization. You know you can find it in your company's online glossary by simply opening a Web browser, but instead of diverting from your current tasks you postpone it for later. By then, you have often forgotten about it and may have missed some important information.

What if you could get the information immediately from where you are?

WebSphere Business Glossary Anywhere provides instant business understanding to make an organization's business language the lingua franca of its business users. To find out the business meaning of a term, simply highlight the term where it is found, click a button and the definition pops up immediately (see Figure 4).

Figure 4: WebSphere Business Glossary Anywhere pops up business term definitions with a single click



### How metadata supports compliance and risk mitigation

Exposure to financial fraud—whether intentional or otherwise—is a key business driver of risk mitigation for organizations. Many still remember the high-profile accounting scandals which eventually resulted in the U.S. federal government’s establishment of a new enterprise regulation, the Sarbanes-Oxley Act, with the anticipated goals of establishing and implementing measurements that enable trust in companies’ accounting practices.

Demands for risk mitigation and compliance affect the way organizations manage their information. One measure for fraud protection or even fraud detection on data leveraged in financial reports is to demonstrate the source of the data, where it flows and how it is transformed as it travels through the enterprise. However, the proliferation of products that result from traditional project-based data integration practices or merging business units creates a seemingly impossible exercise in navigation and integration. Making the process more transparent and efficient requires a consolidation or integration of toolsets and a metadata-driven approach to building a common foundation of answers.

Because data governance programs include “appropriate data usage,” measures for risk mitigation and compliance are primarily derived from the policies defined in an organization’s data governance program.

IBM Metadata Workbench—a Web-based tool within IBM Information Server—utilizes the unified metadata management of IBM Information Server to provide end-to-end visibility into how data flows through an organization. The tool’s sophisticated metadata management and reconciliation techniques assist with the automation and tracking of relationships across disparate systems. This confidence in the perspective of metadata enables an organization to quickly assess where the data was introduced and the various derivations as it flowed through integration technologies.

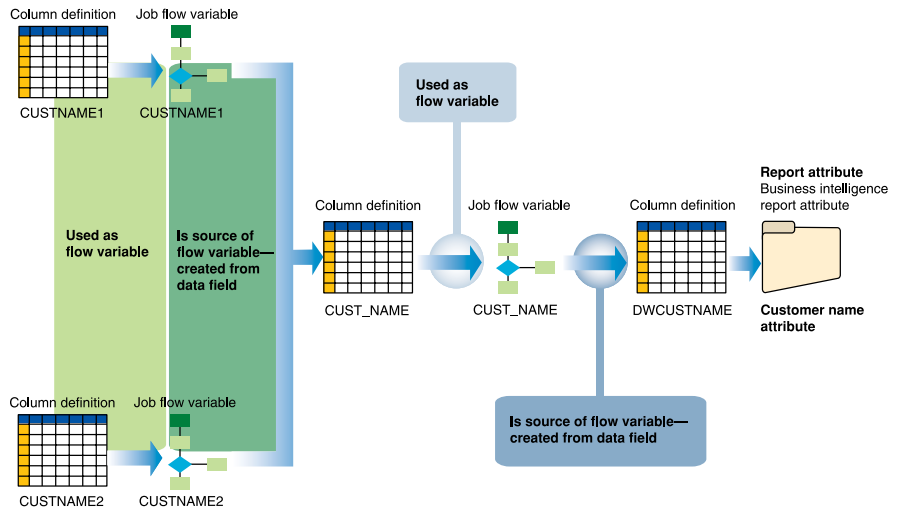
The subsequent sections highlight three functionalities where IBM Metadata Workbench can aid organizations with efforts such as fraud prevention and compliance.

**Data lineage**

Data lineage or provenance provides an audit trail for data movement through integration processes. The result of a data lineage process is the answer to basic questions such as “Where did this data come from?”, “Where does this data go?” and “What happened to it along the way?”

Data lineage reports cover events that occurred to create and update data (see Figure 5). When event or operational metadata is connected to design metadata, it provides a complete view of the data, giving users confidence in the results of the analysis.

Figure 5: Connecting event metadata to design metadata provides users with a complete view of the data lineage



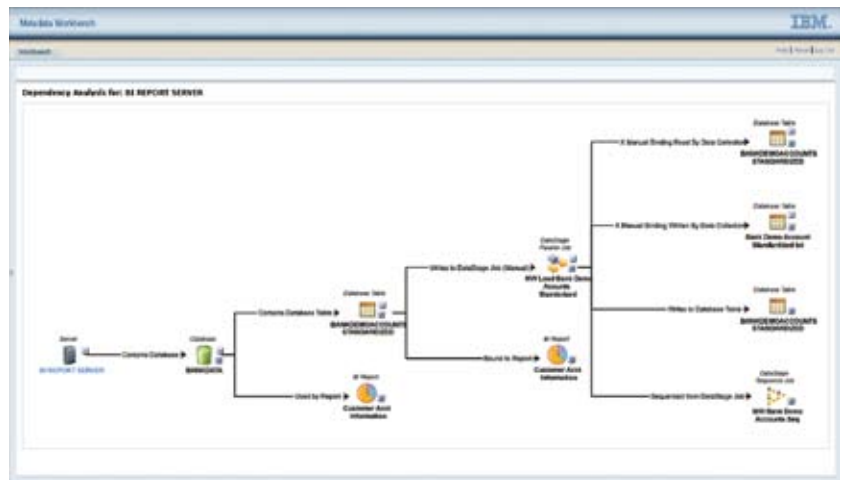
Data lineage reports create an end-to-end view of the environment that can help optimize data integration efforts. IBM Metadata Workbench includes out-of-the-box reports on multiple objects and varying granularity levels, depending on the level of detail required.

**Impact analysis**

Impact of change and dependency analysis helps accelerate the understanding of dependencies that exist across multiple, diverse assets. This capability provides an enterprise-wide view that extends across integration tools and showcases the impact of objects, enabling a full understanding of the impact or cost to change integration workflows in the data integration lifecycle. Impact analysis enables users to examine all of the relationships associated with an object, thereby providing the ability to assess and mitigate risk before creating the change. Taking into account that changes are inevitably introduced during the development lifecycle, understanding the impact of changes will allow companies to manage projects more effectively.

Figure 6 shows a graphical dependency analysis report for a server, listing which databases, jobs or business intelligence (BI) reports would be affected if this server had to be taken offline for maintenance. Traditionally, gathering this type of information would require navigation via multiple users and multiple toolsets to assess the potential risk.

Figure 6: An example impact analysis report for a server, showing the potential impact of changes across the entire environment



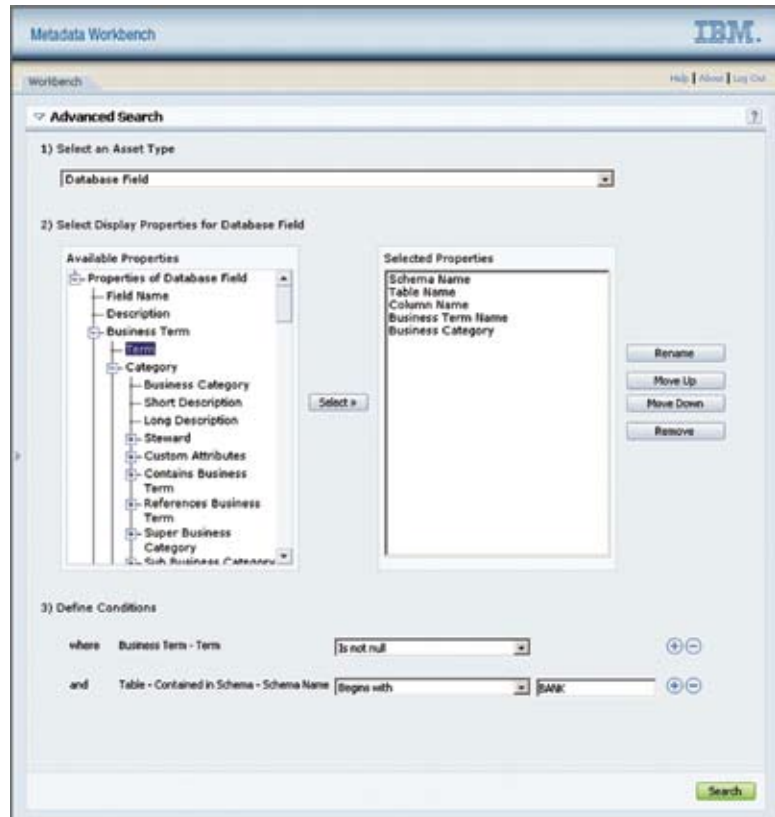


***Customized reporting and search***

IBM Metadata Workbench provides a wide range of out-of-the-box reports for data flow and impact analysis, along with the ability to customize the view of metadata from a variety of perspectives. Consider this example: An IT specialist experienced a failure on one of the company's data warehouse databases. The specialist has certain facts about the failure, such as the time window in which the incident occurred. The next step would be to understand the impact of data integration processes that rely on the affected server being available.

The advanced search interface of IBM Metadata Workbench provides users with a graphical interface—like the one in Figure 7—to build complex requests against business, technical and operational metadata. Individual requests can be run once or they can be persisted to make them available to a wider audience and enable them to re-execute at any point in time.

Figure 7: The advanced search interface of IBM Metadata Workbench provides users with a graphical interface to build complex requests against metadata



Returning to the example above, the system report showed that jobs failed during the database outage. The IT specialist can create a report to analyze if any of the affected tables are being used in BI reports. Because each object contains information about the responsible data steward, the IT specialist can contact the data steward of the report and share information about the outage, as well as contact the data stewards of the affected integration jobs so that they can take appropriate steps. All these actions can be done directly from IBM Metadata Workbench.

**Data governance helps preserve data quality**

As the quantity of information and technologies increases, the complexity of answering these questions also continues to rise. Enterprises are required to take the appropriate actions to ensure effective use of data regardless of data volumes that are increasing through acquired or organic growth. Leveraging assets across the integration landscape requires effective metadata management and, therefore, highly optimized and metadata-driven technologies to support these goals.

A unified metadata management strategy forms the foundation of IBM Information Server. Its metadata components aid organizations in their data governance programs, helping to ensure that analysts, developers and architects have information they can trust.



### For more information

To learn more about IBM Information Server, please contact your IBM marketing representative or visit [ibm.com/software/data/ips](http://ibm.com/software/data/ips)

© Copyright IBM Corporation 2007

IBM Software Group  
Route 100  
Somers, NY 10589

Produced in the United States of America  
September 2007  
All Rights Reserved

<sup>†</sup>The IBM Data Governance Blueprint: Leveraging best practices and proven technologies. [ibm.com/software/tivoli/governance/servicemanagement/data-governance.html](http://ibm.com/software/tivoli/governance/servicemanagement/data-governance.html)

IBM, the IBM logo and WebSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

**TAKE BACK CONTROL WITH** **Information Management**