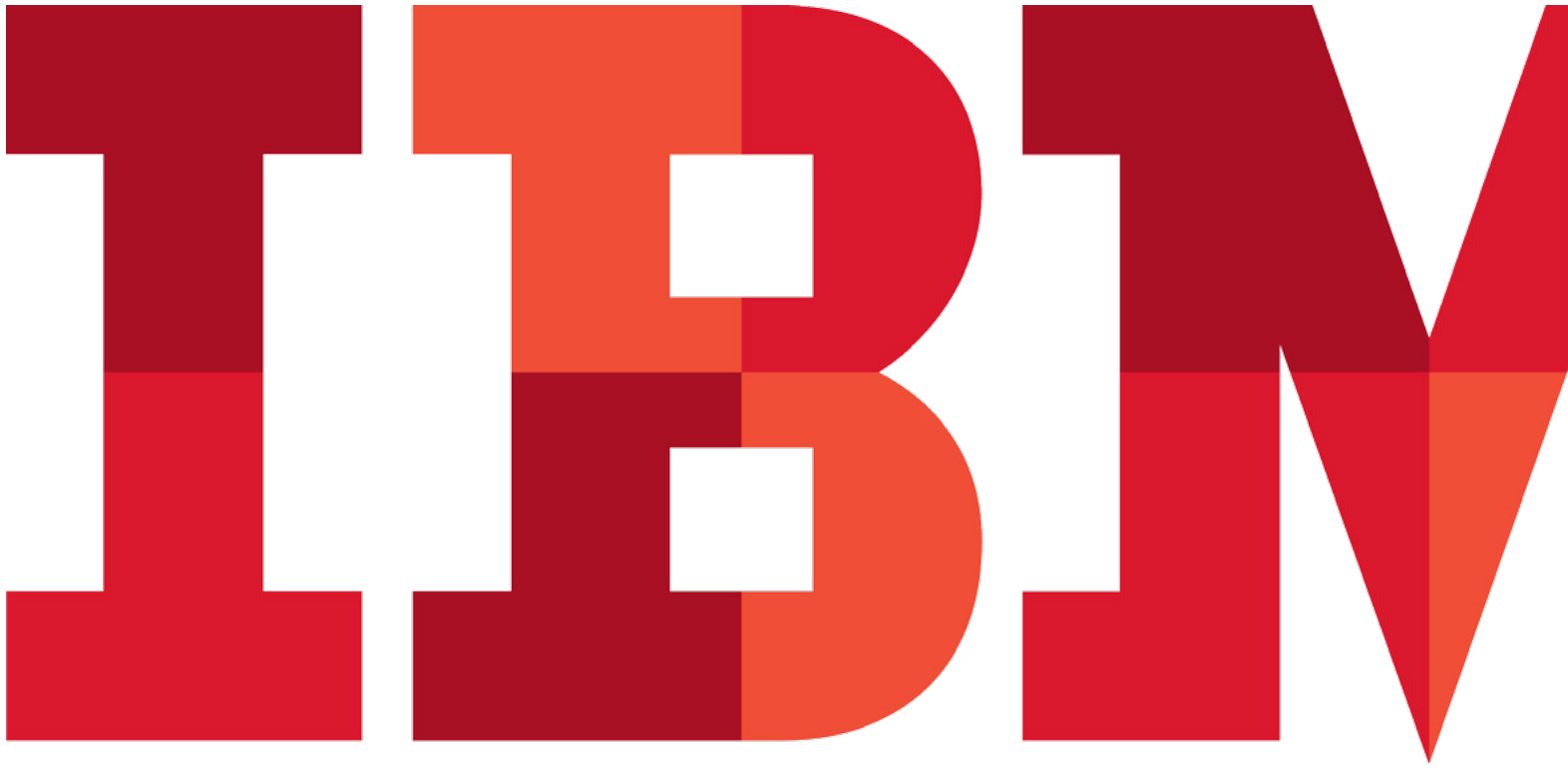
IBM Analytics

Make sense of your data

Using metadata to understand data in a hybrid environment



IBM

Table of contents

Introduction: Imposing order on chaos

Cloud-based data presents a wealth of potential information for organizations seeking to build and maintain a competitive advantage in their industry. However, as discussed in "The truth about information governance and the cloud," most organizations will be confronted with the challenging task of reconciling their legacy on-premises data with new, third-party cloud-based data. It is within these "hybrid" environments that people will look for insights to make critical decisions.

A hybrid environment blends data and computing from both public cloud sources and on-premises systems. The fact that hybrid environments generally occur without much advance planning makes the task of managing ever-growing data stores even more difficult.

Yet, there is a way to make sense of the chaos. As always, the first step is to understand the nature of the problem. The focus needs to be

on the data itself, and much less on the source of the data and on systems used to manage the data. If you make data and ownership of the information derived from the data the highest priority, everything else falls into place quickly.

The four pillars

How can your organization realize the financial benefits of the cloud while ensuring information culled from cloud sources is secure and trustworthy? The answer is governance.

Good hybrid information governance rests on four key priorities for IT and the business:

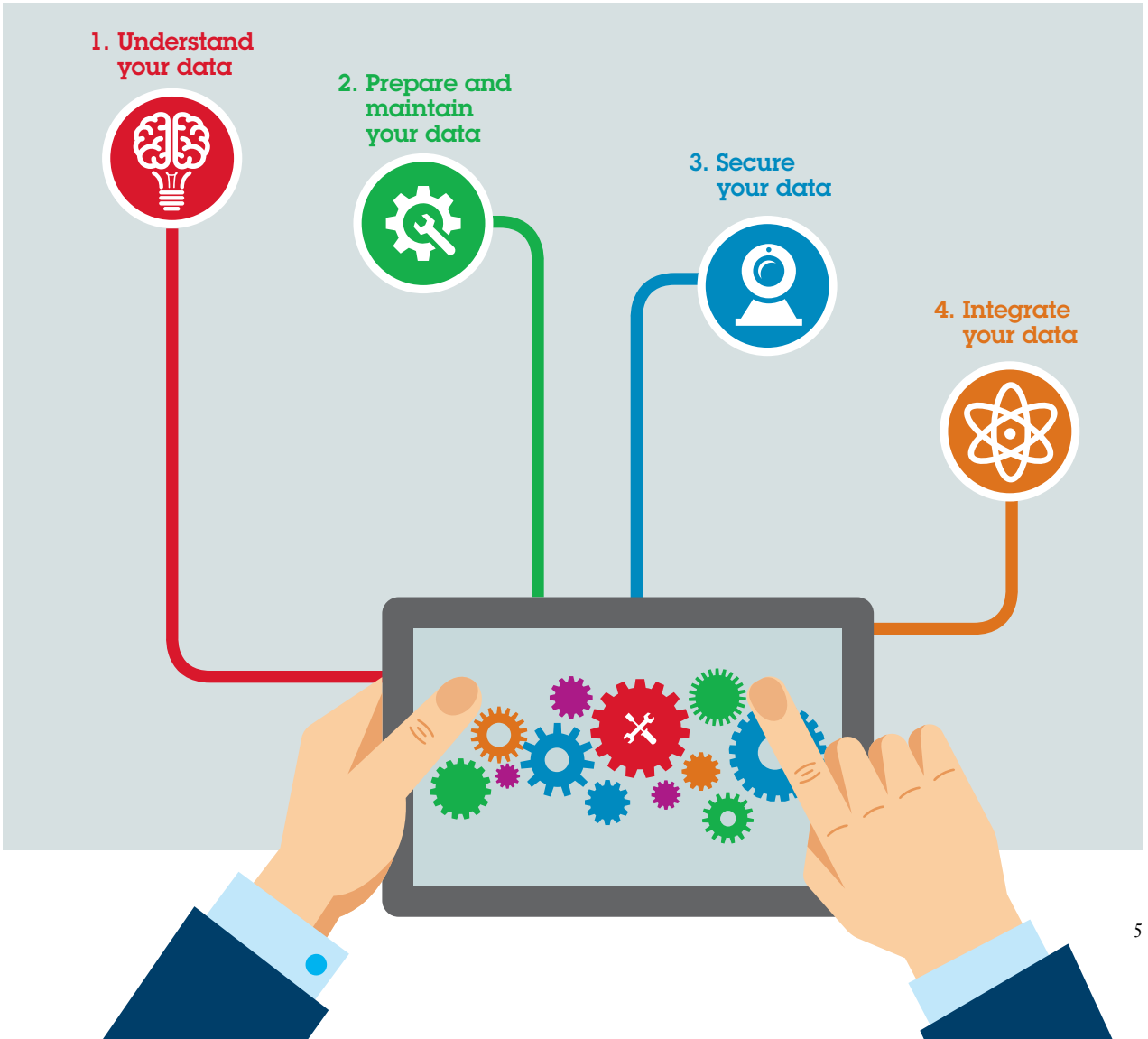
1. **Broad agreement on what information means,** including metadata on common policies and plain-language rules for the information the business needs and how it will be handled.

2. **Clear agreement on how owned-information assets will be maintained and monitored**—for example, operational data quality rules to master data management in on-premises systems.

3. **Enterprise- and departmental-standard practices for securing and protecting strategic information assets,** such as articulating role-based access to information, creating rules governing how information is shared and protecting sensitive information from third parties.

4. **An enterprise data integration strategy** that includes lifecycle management, architecting how data will flow and be assembled into strategic information, and also understanding how that information will be maintained over time.

Roll over the icons below for more on the top priorities for good hybrid information governance.



These components form the foundations of information governance in a hybrid environment. In each case, you need a blend of process, organizational and technical enablers for success. With these pillars in place, your organization will have the flexibility to move with speed and confidence.

This e-book will focus on the first pillar: Understanding your data.

What is information governance?

Information governance practices provide a holistic approach to managing, improving and using information to help you gain insight and build confidence in business decisions. The goal is to:

- **Bring together data from diverse sources for diverse targets, manage its quality and maintain it throughout its lifecycle**
- **Protect data and maintain privacy requirements**
- **Facilitate information-based collaboration across business and technical teams**

These broad capabilities help organizations increase the value of data for information-intensive projects including big data and analytics, application consolidation and retirement, security and compliance, master data management and many more.

Taking ownership of strategic information

Adopting a hybrid environment does not imply you must have your IT strategy completely worked out. In fact, cloud-based aspects of the environment will evolve rapidly in response to business priorities. However, even if only a small percentage of data is flowing in from cloud-based sources, IT needs a plan for data integration and security. IT must help the organization ensure it "owns" the information created from all data and processing, no matter where it is located. The hybrid infrastructure and decentralized computing are means to the ultimate end of creating strategic information assets.

Trusting your information: A business requirement

The most important business requirement for organizations today is the ability to completely understand and trust their information.

Consider this example: A financial analyst is looking at a loan application for a prospective borrower. She needs to know exactly what "mortgage risk" entails: the approved definition of risk, acceptable levels of risk, the business rules that are associated with high-risk and low-risk clients, and where the specific prospect falls on that scale.

The analyst not only must understand this information, she needs to trust that the data in the application she is using aligns with the same business rules and definitions. With this understanding and confidence, she can make better approval decisions, more quickly and confidently.

Unfortunately, in most organizations, business analysts and data analysts often have multiple definitions for a given business term. Even worse, all of those definitions may be correct, depending on context and usage. For example, "mortgage risk" for existing customers and "mortgage risk" for new prospects are two completely different concepts, definitions and related business processes. This problem is amplified when third-party cloud-based data sources are added to the mix. Not knowing the business context of third-party data could negatively impact analysis.

The dilemma of business definition ambiguity and inconsistency is often due to the absence of an enterprise-wide business glossary and stewardship program, which should be part of a larger metadata strategy plan. A metadata strategy comprises two elements: *technical metadata* and *business metadata* (see Figure 1).

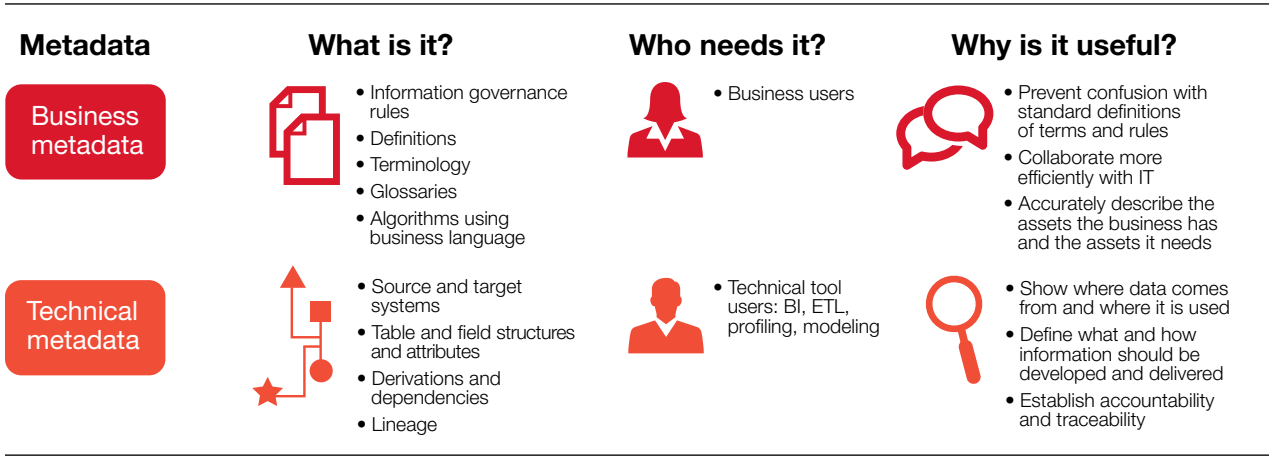


Figure 1. Metadata describes your company's information from a business and technical perspective.

The role of business metadata

Business metadata includes business terms and their definitions, examples of usage, business rules policies and constraints. Together, they define the semantics of a business concept and its realization in physical data assets. Business metadata satisfies the needs of businesspeople and the user community at large by answering questions like these:

- A report shows profit margin, but what does it mean?
- Where does the data for calculating the profit margin come from?
- What calculations went into the determination of the profit margin?
- Who (which data steward) owns this term?
- What are the business rules that are applied to profit margin?
- What other reports show profit margin?

Although business users take advantage of business metadata, anyone can use it to understand what things mean. Examples include learning how, when and by whom a data asset is used, and which policies, rules and restrictions might apply to its use.

Making sense of data has an impact across the entire data ecosystem, not just behind the firewall of the organization. The first step in understanding data is building a business glossary.

Helping business and IT talk the same language

It's essential to establish a common vocabulary for your organization. It improves communication and removes ambiguities within the organization, leading to higher productivity and better utilization of resources. A business glossary provides a framework where such a vocabulary can be created, nurtured and promoted for the benefit of the organization (see Figure 2).

A common, shared business vocabulary is at the heart of information governance and the metadata management practices deployed by an organization. It is a vehicle of communication that brings business and IT into complete agreement.

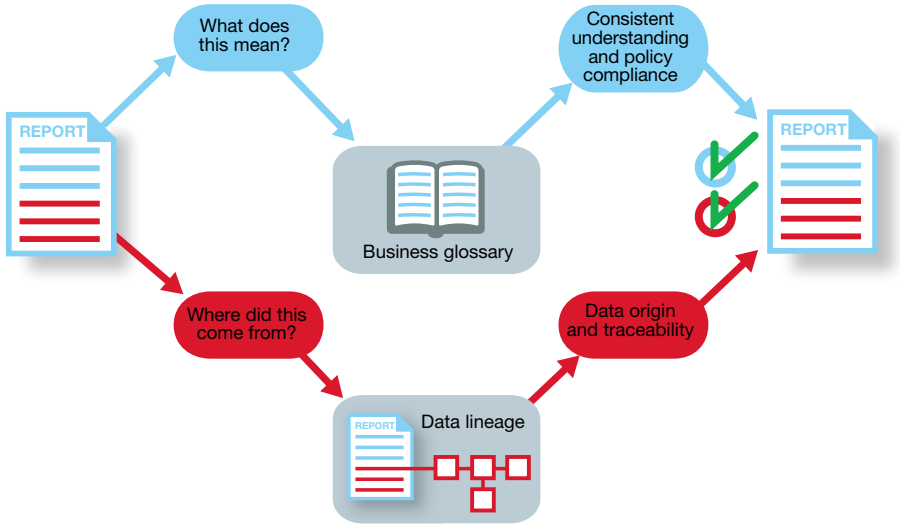


Figure 2. Business glossaries and data lineage tracking enables users to quickly answer questions about the meaning and source of information.

Consider the following questions:

- What is the other party talking about?
- What do the terms in a requirements document mean?
- Do the specifications accurately reflect the requirements?

By maintaining a common vocabulary in a business glossary, business and IT communities have access to a comprehensive body of information that enables them to answer questions like these. They also have knowledge about the data the company generates, processes, stores and uses to support its operations. Through a well-designed hierarchy (taxonomy) and carefully selected and properly formatted business terms (business vocabulary), users can navigate, browse and search for information about business terms, their meaning and usage, and the IT assets used to realize them.

The ability to retrieve information about data; its source, meaning and usage; and how it was processed promotes trust in the data. It also enhances the efficiency of the processes involved in generating and using the data.

In addition to providing an authoritative source for terms and their meaning, the combination of a business glossary and other types of metadata provides access to a rich store of knowledge and analytical capabilities. Businesspeople have immediate answers to data questions such as:

- What information is out there?
- What does it mean?
- Where is it coming from?
- How is it being processed and used?
- When was this information last updated?
- Who owns this information?

A business glossary has many of the attributes required to support information governance. In essence, all of these attributes depend on the ability to create, preserve and publish knowledge about information across your organization.

Linking business terms to data sources

A business glossary goes beyond just a list of terms. Linking terms to data sources—whether on-premises or cloud-based—establishes a connection between business and data sources that enhances collaboration between parties. This capability helps a variety of people within the organization:

- **General business users** have fast access to commonly used vocabulary terms and their meaning, often with additional information such as any constraints and flags indicating special treatment.
- **A business analyst** has a better understanding of the terms used in business requirements, which results in better and faster translation into technical requirements and specifications.
- By knowing the sources of data associated with a term, **data analysts and developers** can be more precise in their job development or report design.

A business glossary provides a collaborative environment to help your organization achieve information governance goals. Here are some ways a business glossary helps achieve these goals:

Data governance

- Supports compliance with regulations such as Basel II through a common language
- Represents and exposes business relationships and lineage
- Tracks change history

Accountability and responsibility

- Assigns stewards as a single point of contact

Improved productivity

- Allows administrators to tailor the glossary to the needs of business users
- Provides access to enterprise information on demand
- Enables the use and reuse of information assets based on a common semantic hub

Increased collaboration

- Captures and shares annotations between team members
- Enables a greater understanding of the context of information
- Promotes use and reuse of trusted information

The role of technical metadata

Technical metadata consists of the technical description of data assets, including the following:

- Schemas, tables and file layouts
- Source and target data store identification and physical attributes (for both on-premises and cloud)
- Data mappings
- Formal specifications of transformation jobs, business rules and other processes

People and systems both use technical metadata. Technical staff—including analysts, developers and administrators—uses technical metadata daily to perform IT tasks. For example, they might analyze requirements and write specifications for a new job, develop a method and write the code, or diagnose a problem and develop a fix.

Answering operational questions

Operational metadata consists of information about the execution of an application or a job. Such information includes times and dates, counts of records processed and rejected, and statistics about processes, processors and servers that are generated in the course of executing a job. Operational metadata answers the following questions:

- At what date and time was a job last executed?
- How many records were processed?
- How many records were rejected?
- How long did it take to process the job?

Users of operational metadata are primarily operations workers who monitor the execution performance of various processes. IT and business management are often interested in the overall system performance and throughput as they consider adding more applications and computing resources. Businesspeople might also be interested in the timeliness of the data, looking for the last time a job ran or a data store was updated.

Building confidence with metadata management

Metadata management involves cataloging information about data objects. A good management strategy provides the tools, processes and environment to enable an organization to answer the question, “How do we know what we know about our data?”

The ability to easily and conveniently locate and retrieve information about data objects, their meaning, source, characteristics and usage is powerful and beneficial to the enterprise. An organization that can do this is better equipped to deal with risk, meet regulatory requirements and improve productivity.

Good metadata management gives organizations confidence in their information. This helps organizations make better business decisions because they know they can trust what they see.

IBM InfoSphere Information Governance Catalog: The IBM metadata management solution

IBM® InfoSphere® Information Governance Catalog helps you understand your information and foster collaboration between business and IT groups by establishing a common business vocabulary on the front end, and managing data lineage on the back end. Using the comprehensive capabilities in InfoSphere Information Governance Catalog, you can better align IT with your business goals.

With InfoSphere Information Governance Catalog, your organization can build and maintain a strong data governance and stewardship program that turns data into trusted information. You can leverage this trusted information in various information integration and governance projects, including big data integration, master data management (MDM), lifecycle management, and security and privacy initiatives.

Significantly, InfoSphere Information Governance Catalog allows business users to play an active role in information-centric projects. They can collaborate with their IT teams without the need for technical training.

This level of governance and collaboration creates an environment where decisions are more accurate and business opportunities are more readily captured. The end result is an organization with a consistent understanding of information, what it means, how it is used and why it can be trusted.

InfoSphere Information Governance Catalog creates trust in an organization's information assets by:

- Establishing a common business language and managing business perspectives about information, aligning those views with the IT perspective
- Managing and exploring data lineage for data governance and compliance efforts
- Creating a well-documented, end-to-end information blueprint to ensure business intent and requirements are aligned with data design and enterprise and reference architectures before starting a strategic project

Next steps: Continuing the cloud governance discussion

Cloud-based data and processing services present too much opportunity for business users to ignore, and IT is charged with maintaining the integrity of internal, on-premises transactional and reporting systems. Charting a governance strategy for a hybrid environment is not something to consider at a future date. It needs to happen now.

This e-book discusses the role of metadata in helping you understand your data, regardless of its source. **For a look at other pillars of information governance in hybrid environments, download one or all of the e-books in this series:**

- The truth about information governance and the cloud
- Prepare and maintain your data
- Securing data in the cloud and on the ground
- Developing a data integration and lifecycle management strategy for a hybrid environment

For more information on IBM governance thought leadership and supporting technologies, visit: **ibm.com**/software/data/information-integration-governance

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing

Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
April 2015

Please Recycle

IMM14174-USEN-00