Gartner

Research

Publication Date: 19 August 2010 **ID Number: G00206115**

Gartner Clarifies the Definition of Metadata, 2H10-1H11

Mike Blechar, Mark A. Beyer, Jess Thompson, Anne Lapkin, Nicholas Gall

Gartner's definition of metadata will help the constituencies within an organization reach a consensus on what metadata is, and will allow them to cooperate in managing metadata across the enterprise. IT managers can use Gartner's definition to begin rewarding discussions about metadata. This research is a critical part of delivering an enterprise metadata management project.

Key Findings

- The term "metadata" is widely used, but with different meanings. The common interpretation of metadata as "data about data" has misled IT professionals and information workers alike.
- Metadata unlocks the value of data and, therefore, requires management attention.
- The classification of metadata varies based on the context and manner in which it describes data.
- Failure to manage metadata properly will hamper critical activities such as information management, business process management and service-oriented architecture (SOA) initiatives.

Recommendations

- Define metadata in a way that demonstrates the value of information to the organization.
- Identify the uses of the term "metadata" throughout the organization to help organizations determine the level of effort necessary to implement metadata management.
- Form a cross-organizational team to reach agreement on the fundamental purpose of metadata and on what information asset is being described.

WHAT YOU NEED TO KNOW

The common, shorthand definition of metadata is data about data, which often confuses efforts to actually manage metadata. Well-meaning but often misguided, efforts to define metadata have blurred the line between data and metadata. *Generally speaking, the more valuable the information asset, the more metadata there will be about it, and the more valuable the metadata will be.* Metadata unlocks the value of data and, therefore, requires management attention. Metadata management needs are not the same for all companies, or for all roles within one company. IT managers should focus on achieving consensus on what metadata is, based on what it describes (information assets) and how those assets are used to create value for the organization. The term "information assets" may help define any aspect of the enterprise architecture (including assets related to the business, technology, application and data architectures). Gartner's definition of metadata provides a good starting point for these cross-architecture discussions involving IT and the other business units.

ANALYSIS

Context

Metadata makes information valuable by allowing people to put that information to use. Therefore, critical advice for any company that wants to maximize the value of its information assets is to "manage your metadata." However, efforts to do so often sink into intractable disputes over what is and is not metadata. These disputes occur because different constituencies are each talking about their own perspective of an information asset without realizing it.

Consider the ramifications of an X-ray taken at a hospital:

- The radiology department needs to describe the X-ray as a *content object;* for that department, therefore, the metadata includes the doctor's name, patient's name and the body part subjected to the X-ray.
- To build an *X-ray management system,* designers need structural metadata, such as "the doctor's-name field is 45 characters long."
- A medical insurance company building a *payment process* needs data that will also serve as metadata, such as the doctor's name, the patient's name and preauthorization status for the X-ray to be taken.

Each of these perspectives is valid but incomplete when it comes to managing metadata enterprisewide.

Analysis

IT managers in charge of managing metadata and unlocking the value of the organization's information should start by bringing together the various constituencies that create or use metadata, including database managers, content managers, business process owners, software engineers and users. The discussion will bear fruit only if all parties first reach a consensus — it doesn't have to be perfect agreement — about what metadata is. Many clients have asked Gartner for a definition of metadata. We have formulated a definition that organizations can adapt to reach a consensus within days, rather than weeks or months. Our explanation of this definition identifies the source of the dispute that keeps people from working together to manage metadata.

Gartner's Definition of Metadata

Instead of fighting over what metadata is, the various parties should try to agree on the fundamental purpose of the metadata and the information asset that is being described. Gartner's definition of metadata will provide a useful starting point for these discussions.

Metadata is information that describes various facets of an information asset to improve its usability throughout its life cycle.

Let's look at what each of the key terms means. It is important to note that the Gartner definition complements and clarifies other industry definitions (see Note 1).

"Information Asset"

This term encompasses a vast range of possibilities, including but not limited to:

- Database tables
- Books (both printed and digital)
- XML documents
- Spreadsheets
- Videotapes
- Card catalogs
- Paper forms
- Business Process Execution Language (BPEL) scripts
- System administration logs
- Digital images
- Software engineering artifacts (such as Unified Modeling Language [UML] diagrams and SOA assets)

For the purposes of this research, the term information asset does not include traditional hardware assets (for example, servers, storage devices and networking equipment). However, it may include any software or data structures associated with such hardware, such as OSs, firmware and routing tables.

"Facets"

Organizations such as the U.S. National Information Standards Organization and the Digital Library Federation define three basic types or categories of metadata:

- *Descriptive metadata* describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author and keywords.
- *Structural metadata* indicates how compound objects are put together — for example, how pages are ordered to form chapters.
- *Administrative metadata* provides information to help manage a resource, such as when and how the data was created, file type and other technical information, and who can access it. This category includes two important subgroups: *rights management*

metadata, which deals with intellectual property rights, and *preservation metadata*, which contains the information needed to archive and preserve a resource.

Distinguishing between these facets will enable the different constituencies to speak with greater precision about the kinds of metadata they work with. For example, the database community traditionally focuses on structural metadata, while content managers focus on descriptive metadata.

"Life Cycle"

Discussions about managing metadata enterprisewide also need to reflect the full range of uses to which information assets are put and the roles that use them. A number of activities define the life cycle of any information asset from creation to destruction. Some major activities and associated metadata (with illustrations based on the X-ray example above) will include:

- Capturing (for example, camera-setting metadata, such as f-stop and lighting)
- Finding (patient ID metadata)
- Accessing (file format metadata)
- Viewing (image thumbnail metadata)
- Understanding (highlighting metadata — circles/arrows point to areas of interest on the X-ray)
- Changing (editing metadata)
- Transmitting (Web services metadata)
- Storing ("area of interest" metadata)
- Securing (Security Assertion Markup Language [SAML] certificate metadata)
- Archiving ("last used" metadata for a retention period)
- Sharing (process model or workflow metadata)
- Evolving (versioning metadata)

Different roles within the organization handle different life cycle activities. For example, "understanding" might be the responsibility of a radiologist who adds explanatory notes to an X-ray image, while a records manager would handle "archiving." In addition, each role may add "understanding" metadata, rather than one role exercising responsibility for an entire class of life cycle metadata. Key roles and examples of the metadata they focus on include:

- Data management professionals (data dictionary entries)
- Content management professionals (content attributes)
- System administrators (Common Information Model attributes)
- Web administrators (HTML tags)
- Software engineers (configuration files)
- Business analysts (business processes and flows)
- SOA architects (Web Services Description Language [WSDL] documents)

- Information and records managers (Machine Readable Cataloging records)

"Improve Its Usability"

The core purpose of metadata is that it enables people to use an information asset more effectively. It is important to consider that data used to describe other data is serving the role of metadata. Metadata makes the completion of activities in the life cycle better or more effective:

- Easier (explanation metadata makes the X-ray image easier to understand)
- Simpler (tagging metadata makes it simpler to find the image)
- Faster (indexing metadata makes it faster to sort images)
- More efficient (compression metadata makes storage more efficient)
- Lower cost (caching metadata lowers the cost of accessing the image)
- More reliable (error correction code metadata makes transmission of the data more reliable)
- Lower risk (embedding patient metadata into the image reduces the risk of using the wrong image for diagnosis)
- Higher quality (color correction metadata improves the color quality across displays)
- Less resource-intensive (hyperlink metadata enables many e-mails to refer to one stored image)

Typically, an organization places a priority on improving the usability of its most valuable information assets — for example, customer, product and supplier data. *Generally speaking, the more valuable the information asset, the more valuable the metadata about it.* Each organization should define metadata to reflect the value of the information assets. The primary reason to amass metadata is to provide value statements that describe an information asset and the life cycle activities associated with it. The value statements benefit the organization and its partners.

Applying Gartner's Definition of Metadata

Companies can use the parts of Gartner's definition presented in this research to test whether something is metadata — or just data — in a given situation. However, it is critically important to recognize that metadata is data. It is data that is used to reference a specific facet of understanding an information asset to improve its usefulness.

To continue the X-ray example, suppose the hospital's imaging center receives word from the X-ray machine manufacturer that a particular make and model has a malfunction that bends images 2% from top to bottom. There are several instances of metadata built in here (for example, linking the imaging problem to a specific make and model of machine is a metadata point; the degree of error is a data point). After fixing the machines themselves, the imaging center uses the make and model data elements stored in X-ray images taken previously as metadata to find all affected images and use digital manipulation software to correct them. In this example, the:

- Make and model data elements are used in respect of certain *information assets* (X-ray images).
- Applied facet of the make and model is *administrative* (identifies the machines that created the images).

- Make and model are used in *finding* those images.
- Make and model are used to guide a quality rule to improve the value of the images by making them *more reliable* (correcting an error).

Thus, the make and model data element meets the requirements of Gartner's definition and should be considered metadata in this context.

Further Cautions and Recommendations

- Before convening the various constituencies in your organization to discuss how to manage metadata, conduct a survey to identify the different uses of the term metadata across the organization. This will enable you to determine how much effort is necessary to unify metadata and then structure it.
- While surveying the organization for examples of metadata, be aware that metadata about an information asset is frequently collected independently of the manager or owner of the asset (for example, users or customers may tag your content for their own purposes).
- There is no final arbiter of the value of metadata to the organization. Liability claims, compliance and regulatory issues can supersede internal guidelines, and different business units can dispute the value of metadata.

Key Facts

- Metadata is information that describes various facets of an information asset to improve its usability throughout its life cycle.
- An information asset is a digital rendering or description of a real or logical party (such as a person or corporation), location or place, financial or physical asset (thing), or concept.

RECOMMENDED READING

"Metadata Business Threats and Opportunities, 2H10-1H11"

"Metadata Management Technology Integration Cautions and Considerations"

"Best Practices for Metadata Management"

"Managing the Multiple Dimensions of Metadata"

"Metadata Repositories Address Disparate Sets of Needs"

"The Evolving Metadata Repository Market"

"Metadata Management Trends in Master Data Management and Data Warehousing"

"Applying Data Mart and Data Warehousing Concepts to Metadata Management"

"Metadata Management Is Important to MDM's Long-Term Success"

Note 1
Other Definitions of Metadata

The Gartner definition of metadata is compatible with the definitions of a variety of industry organizations. For example, the U.S. National Information Standards Organization says: "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (see "Understanding Metadata" p. 1, http://www.niso.org/publications/press/UnderstandingMetadata.pdf).

The International Organization for Standardization (ISO) says: "Metadata is the information and documentation which makes data understandable and shareable for users over time. Data remain usable, shareable, and understandable as long as the metadata remain accessible" (see http://www.iso.org/iso/home.html).

REGIONAL HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brazil
Av. das Nações Unidas, 12551
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509