Metadata

Idea In Short

Metadata is defined as data that provides information about other data. In other words, it is data about data. Metadata is used to help organize, find, and understand data.

Metadata is data about data. It is information that describes an item and its content. In a library, Metadata is used to catalog and organize books and other materials. The catalog entry for a book includes the author, title, subject, and other information. This information is the Metadata.

Metadata is important because it makes information easier to find. When you are looking for a book on a particular subject, you can use the library's catalog to find all the books that have that subject in their Metadata. Without Metadata, data would be difficult to locate and interpret. Most data today is stored electronically, and Metadata is often stored along with the data it describes. This is especially true for digital data, such as images, video, and audio.

The problem

Today, many enterprises use several applications that process data. For example, enterprises use Digital Asset Management [DAM] software to safely store, organize, manage and share their Digital assets, such as documents, photos, videos, etc. Such applications require Metadata to, not only properly process data, but also ensure consistency and trustworthiness of data as it flows among enterprise systems. Poor Metadata impairs system performance, promotes data loss and data breaches, increases cost of information search, decreases user productivity and experience, to name a few issues.

Business value

As data itself has become currency, the Metadata describing it — and what happens to it —

has also emerged as a core asset of modern business. Metadata interweaves itself throughout all information; like DNA, it serves as the genetic makeup of data. So, even though Metadata may not be the most obvious data created, it holds tremendous value in unlocking and exploiting the value of enterprise information. Metadata management is a cross-organizational agreement on how to define informational assets for converting data into an enterprise asset. As data volumes and diversity grow, metadata management is even more critical to derive business value from the gigantic amounts of data. Metadata management drives business value, improves innovation and collaboration among IT and business stakeholders, reduces information search costs, increases trust and reliability in data among data consumers, and helps mitigate risk. Particularly, it enables Data Citizens to access high-quality and trusted data, thus ensuring that they work with the right data to deliver accurate insights.

IT value

Metadata management and its use in enterprise information management is one of the critical Information Technology [IT] focus areas for most public and private-sector organizations. These organizations seek to reduce their IT portfolio and control escalating IT costs through comprehensive Metadata management programs. IT organizations that have implemented well-architected enterprise-wide Metadata have achieved tremendous successes in consolidating disparate IT systems, reducing total IT spend, improving market responsiveness, reducing information search costs and increase their value to their organization by enabling previously unavailable capabilities. Furthermore, such IT organizations have been able to take advantage of the state-of-the-art developments in Information Technology, such as Cloud computing, Blockchain, Internet of Things [IoT], Digital Twins, etc.

Metadata creation

Metadata can be created manually or automatically. For example, when you fill out the fields in a library catalog, you are creating Metadata manually. When a software program generates Metadata automatically, it is called automatic Metadata. Some examples of automatic Metadata are the data that is generated by a digital camera when a picture is taken, or the data that is generated by a scanner when a document is scanned. Digital cameras, for example, typically store Metadata in the form of EXIF data. This data includes information such as the date and time the photo was taken, the camera settings, and the

GPS coordinates of the location where the photo was taken. Many software applications also generate Metadata. For example, when a document is created in Microsoft Word, the application stores Metadata about the document, such as the author, the date the document was created, and the file size. When data is shared, the Metadata is often shared as well. This allows people who receive the data to understand it and use it appropriately.

Metadata types

There are two types of Metadata based on the function it serves in information management:

- **Structural Metadata** is data about the structure and relationships among data elements. This Metadata reveals how different elements of a compound data object are assembled. For example, database schemas are structural Metadata. Structural Metadata can include the number of pages, the chapter titles, and the order of the chapters
- **Descriptive Metadata** is data that describes the content, quality, condition, origin, and other characteristics of data. Examples of descriptive Metadata include file format, size, date created, and date modified
- **Administrative Metadata** allows administrators to impose rules and restrictions governing data access and user permissions. It also furnishes information on required maintenance and management of data resources. Often used in the context of ongoing research, administrative Metadata includes such details as date created, file size and type, and archiving requirements
- **Technical Metadata** describes the technical aspects of an item. Technical Metadata is a synonym most closely associated with items in digital libraries. This can include the file format, the size of the file, and the software needed to view the file
- **Legal Metadata** provides information on creative licensing, such as copyrights, licensing and royalties
- **Preservation Metadata** guides the placement of a data item within a hierarchical framework or sequence
- **Process Metadata** outlines procedures used to collect and treat statistical data. Statistical Metadata is another term for process Metadata
- **Provenance Metadata**, also known as data lineage, tracks the history of a piece of data as it moves throughout an organization. Original documents are paired with Metadata to ensure that data is valid or to correct errors in data quality. Checking

the provenance is a customary practice in data governance

- **Reference Metadata** relates to information that describes the quality of statistical content
- **Statistical Metadata** describes data that enables users to properly interpret and use statistics found in reports, surveys and compendium
- **User Metadata** is data that is sorted and analyzed each time a user accesses it. Based on analysis of use Metadata, business can pick out trends in customer behavior and more readily adapt their products and services to meet their needs

Metadata standards

A Metadata standard is a set of rules for creating and formatting Metadata. There are many Metadata standards, including those for library catalogs, learning objects, and geospatial data.

Some common Metadata standards are the Dublin Core and the Resource Description Framework [RDF], Text Encoding Initiative, and Metadata Encoding and Transmission Standard [METS]. These standards help ensure that Metadata is consistent and can be understood by people and machines.

Usage

Metadata can be used for many purposes, including:

- Describing the content, structure, and format of digital information
- Organizing and retrieving digital information
- Controlling access to digital information
- Preserving digital information

Storage

Metadata is typically stored with the digital information it describes. For example, the Metadata for a digital photograph might be stored with the photograph file in a file format that includes Metadata tags. Metadata can be stored in a variety of formats, including text files, XML files, and databases.

Components

The components of Metadata are:

- **Metadata element:** a unit of information that describes a piece of digital information. For example, a title is a Metadata element that can be used to describe a document.
- **Metadata schema:** a set of rules for the structure and content of Metadata elements. A Metadata schema defines the element set, or set of Metadata elements, that can be used to describe a particular type of digital information. For example, the Dublin Core Metadata schema provides a set of elements that can be used to describe a Web page
- **Metadata registry:** a database of Metadata elements, element sets, and schemas. A Metadata registry can be used to store and share Metadata. The U.S. Geological Survey's National Biological Information Infrastructure [NBII] Metadata Registry is an example of a Metadata registry

Summary

Metadata is a valuable tool for organizing, finding, and understanding data. By storing Metadata along with the data it describes, we can make sure that the data is properly used and interpreted.