

Metadata Definition

Metadata is *data* about data. Data is basically the same thing as *information*, although it is often in a form that is easier for humans and/or computers to use and manipulate.

Information can be broadly defined as any pattern that can be recognized by some system (e.g., a living organism, an electronic system or a mechanical device) and/or that can influence the formation or transformation of other patterns. The pattern can be in any of a wide variety of forms, for example spoken or printed words, temperatures, visual images, pain, radioactivity, DNA, the structure of a crystal, color, or electron flows. It can range from extremely simple single *binary* values (e.g., *yes* or *no*, or *zero* or *one*) to something so complex that only a few human minds can understand it (e.g., Einstein's theory of relativity).

Metadata is used to organize, locate, manipulate and otherwise work with data when it is not necessary or desired to actually deal with the data itself. This is because the metadata is usually far smaller and easier to work with than the data that it represents.

Metadata can include information about various aspects of the data that it describes, including its structure, content, quality, context, origin, ownership and condition. There is usually some extent of separation between data and its metadata. Metadata can be stored in various forms and in one or numerous locations. Examples of types of locations in which metadata are commonly stored include computer databases, envelopes on letters, special pages on books (e.g., covers and title pages) and even bar codes and RFID (radio frequency identification) tags.

For example, metadata for a book includes its title, author, publisher, copyright date, location of printing, language, price, category, ISBN (international standard book number), number of pages and binding type. Libraries usually add more metadata to a book after they purchase it, such as the date of acquisition, catalog number and copy number. Some of this metadata appears on or in the book, and most or all of it also stored in databases about books, such as electronic card catalogs and reference books about books.

The metadata for an article in a periodical could include its title, subject, author, length, the name of the periodical, publication date, the section or page on which the article begins and whether or not it was *peer reviewed*.

The metadata about a web page that is typically returned by a search engine (e.g., Google or Yahoo) includes brief sampling of the contents and its URL (universal resource locator). Additional metadata about a web page includes its author, the human language in which it is written, the target audience, the dates of creation and last modification, keywords, file size, the *markup language* (e.g., HTML, XHTML or XML) in which it is written and any access restrictions (e.g., registration or passwords required).

Metadata can be about any kind of information or objects, including images, sounds, databases and collections. For example, in the case of an image it would include the name, creator (e.g., photographer or artist), date of creation, subject category, means of creation (e.g., photograph, painting, computer generated), copyright owner, file size (in [bytes](#)) and file format (e.g., jpeg, gif, png or tiff). Examples of collections that have metadata are books in a library, collections in a museum and inventories in a warehouse.

Major functions of computer [filesystems](#) are the storing of metadata about files and facilitating the locating and manipulation of files. The metadata about a file on a [Unix-like operating system](#) includes its file type (e.g., data file, directory, link), name, [timestamps](#) (i.e., dates of creation, last access and modification), location on the [filesystem](#), size (in bytes), its physical location (i.e., the addresses of the blocks of storage containing the file's data on a disk), ownership (usually the same as its creator), access *permissions* (i.e., which users are permitted to read, write and/or execute the file) and file type. There are several levels of file type, the broadest of which is plain text and binary; binary can be further broken down into such categories as word processing, image, sound and executable; each of these can be further broken down (e.g., sound file formats include au, avi, bwf, mp3, ogg and wav).

An [inode](#) is a *data structure* on a Unix-like operating system that stores all the metadata about a file except its name(s)¹; the name(s) and the actual data of the file are stored elsewhere. A data structure is a way of storing information in a computer so that it can be used *efficiently*.

Because metadata is also data, it is possible to have metadata for metadata. This is referred to as *meta-metadata*. An example would be data about a number of databases, each containing the metadata for a library. Another example is a table in a relational database that contains information about all of the tables in the database (e.g., their name, use, date of creation, creator, access restrictions, number of columns, number of rows, etc.).

Metadata has been used effectively for a long time. For example, thousands of years ago, when libraries consisted of piles of rolled-up scrolls, little tags were attached to each scroll so that its contents could be identified without having to open it. The development and rapid improvement in computers has made it much easier to create and use metadata. This has resulted in a rapid growth in the amount of metadata available and the extent to which it is used.

The prefix *meta* is a Greek word that can mean *after*, *among*, *change* or *with*², and *data* is Latin for *information*. The term *metadata* was coined by Jack E. Myers in 1969 and was first used in print in 1973 in a product brochure. The word *Metadata* (beginning with an upper case letter) was registered in 1986 as a trademark belonging to The Metadata Company; however, it is extensively used (beginning with a [lower case m](#)) in a generic sense.

¹This storage of names elsewhere on such operating systems has the advantage of allowing any file to have multiple names.

²This is similar to the term *metaphysics*, which literally means *after physics* and is the branch of philosophy that seeks to explain the nature of being and reality. The *meta* is

Created March 21, 2006.
Copyright © 2006 The Linux Information Project. All Rights Reserved.