

# Demystifying and Debunking Metadata



May 6th 2015 - 09:01 AM  
By Gary Olson, GHO Group



Metadata becomes increasingly important in a world where everything needs to be searchable.

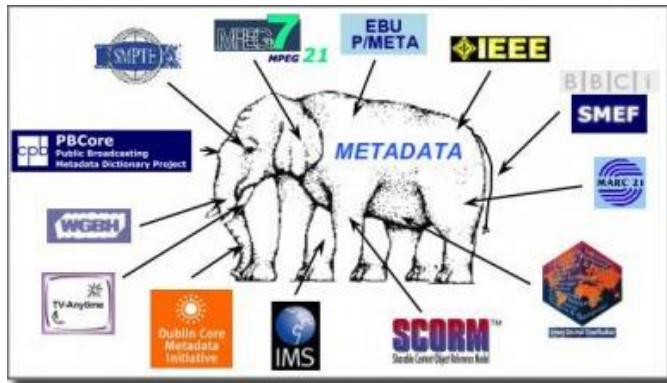
The word “Metadata” was introduced into the broadcast industry and instantly became a technological pariah. The mere mention of the word brings dread and groans. Yet, Metadata is one of the most critical elements in media management, handling, movement, control, rights and monetization.

Maybe it’s the word DATA that turns people off. It’s not glamorous, certainly no awards are given for the Best Metadata Schema. Metadata has become a trendy buzz word that is used without the understanding of the depth and importance it rightly deserves. The way this term is used is different than its actual meaning. In some cases it has also become a generic term to describe all data points very much like Kleenex is the term used to identify all facial tissues.

This will **NOT** be an attempt to make it glamorous. It **WILL** be an attempt to help understand what it is, its value and its importance.

The word metadata comes out of Library Sciences and was used to describe information like card catalogs that tracked books, document, and all forms of printed materials. The Dewey Decimal System of cataloging is a form of metadata. This is where the definition “information about information” came from. As the world became digital and all “information” was now referred to as data, Metadata just became “data about data”.

There are five (5) types of Metadata according to the US National Institute of Standards and Technology (NIST). These are Descriptive, Structural, Administrative, Rights Management, and Library & Preservation metadata.



There is no shortage of groups who claim to have created "metadata" dictionaries. One problem with this solution is that each group has different goals and therefore often use dissimilar definitions for like terms.

- **Descriptive metadata** is the most familiar and what is used by asset management application. This is the metadata used to browse and search for an asset. It can include elements such as title, abstracts, descriptions, author, and keywords. This is the metadata that search engines and program guides use to locate media.
- **Structural metadata** is the control data that is used by automation and orchestration tools. It handles the movement of media, automation, and interaction between applications and databases. An example of structural metadata is identifying the correct file and knowing to send it to a transcoder. Then giving the transcoder a set of instructions to convert the file to a different format using a specific profile. This process controls how compound objects are put together or which instructions are sent to transcoders to format an asset for cable, phone, tablet, and web.
- **Administrative metadata** is used for management. This is defining the media by file type, ownership, it's location in the storage network, internal usage, and access rights. Additionally, this is the metadata that integrates with business systems. Essentially, the administrative metadata captures the technical details of the asset, making it easier to manage.
- **Rights management metadata** is different than administrative usage permission and access rights. This is the protection data for distribution and includes copyright, intellectual property, and distribution rights. It carries the expiration rules, encryption policies, and tracking metadata used in watermarking for piracy protection.
- **Library & preservation metadata** is used for archiving purposes. When metadata is part of any discussion, it typically centers on archiving. There are core elements that need to travel with the asset for preservation. This is to ensure that when the asset is accessed or restored sometime in the future, there is adequate information (both descriptive and administrative) that can identify it. Metadata existed in the

enterprise before migrating into the broadcast industry.

In the enterprise, there is a concept known as the *Managed Meta Data Environment*(MME). The Managed Metadata Environment represents the architectural components, people and processes that are required to gather, retain and disseminate metadata throughout the enterprise in a structured process. Enterprise professionals that have built a metadata repository have realized it is much more than a database that just holds metadata and pointers to metadata.

The technology where the term metadata is most often used is in Asset Management. HOWEVER, this is only one of the use cases for Metadata. Metadata can be very cool. Metadata is what makes content smart from creation to consumer. It assures the accessibility of content both in active storage and archive. Metadata protects content with DRM and it helps the automation and orchestration tools know what content to select and when to move it to the appropriate application for quality control, retention, archive, playout or distribution. Metadata is the EPG or PSIP that allows a consumer to find their program and tell the DVR when to record. Metadata helps phones, tablets, computers and gaming consoles find the program you are interested in watching. It also tell the content originator what you watched.

In the production process, metadata controls how a file moves around, where it can go what processes it needs for finishing. Metadata tells the asset manager what raw elements or B-Roll was used in the finished program and allows the archive system to retain that information. Here is a brief video tutorial about metadata, courtesy EDINADatacentre.

[Try watching this video on www.youtube.com](http://www.youtube.com)

Video tutorial

How much metadata is enough and how much is too much? Is there are standard for Metadata? Well, actually there is, sort of. When the librarians met in Dublin, Ohio, they produced the minimum required amount of information an asset needed for archive. These were **Fifteen (15)** specific identifiers (data fields) for any piece of material that would be archived and called a metadata schema. This became known as the DublinCore. When EBU was looking to establish a minimum requirement for metadata for media, they expanded this to **Sixty (60)** required identifiers for any asset and this became known as EBUCore. Coming back to the US, as PBS was building their digital library, they created PBCore, originally with **Forty Eight (48)** data fields. Now PBCore 2.0 has **Eighty (80)** data points as the required set of information needed to identify and preserve an asset.

And PBS published their schema in attempt to have it adopted as a standard.

However, when the US Library of Congress was creating their digital library, they specified **Eight Hundred (800)** required identifiers in their metadata schema for preservation and archive.

Every organization and institution should determine what information is critical for preservation and also make the asset retrievable through easy to use search criteria (unstructured keywords).

As metadata moved into the lexicon of broadcast so did other terminology that confuses and intimidates such as words like taxonomy and ontology. And this new word vocabulary has new application. These terms used with

metadata now have more relevance when searching and browsing.

OK, both of these terms have their origin in Greek science.

Let's start with **Taxonomy**, the Greek word "*taxis*" means "arrangement" and the word "*nomia*" means "method". Taxonomy is the science of identifying and naming species and arranging them into a classification. Taxonomy is used to create classifications according to a pre-determined system (schema) or a controlled vocabulary. It is used with the resulting catalog to act as a framework for retrieval. In asset management, taxonomies are used to organize assets and manage metadata. By employing a taxonomy to classify content and assets, it makes searching or browsing using a digital asset management tool easier for users who do not know many of the details about what they are looking for.

The next term is **Ontology** is based on the Greek word "*onto*" and translates as "being; that which is" and the word "*logia*" which means "science, study and theory". Very metaphysical! Ontology is the philosophical study of the nature of being, existence, or reality. It is also the basic categories of being and their relationship.

**Ontology** is a classification scheme. It is a method used to define the relationships between objects in the world and by organizing objects by subject categories. Ontology defines how to divide up an object in smaller categories. This might not be by subject; an object may instead be divided by type, format, and location. An entertaining example of this might be: "All metadata is data, however not all data is metadata". Very Deep!

There's one more Metadata term that should be defined before you need a rest. This would be **Controlled Vocabulary**. A controlled vocabulary is a defined list of words and phrases that are used to tag an asset so it can easily be retrieved by a search. These are the keywords that are an integral part of the metadata associated with an asset. Controlled vocabularies provide a way to organize the metadata, making it easier to index.

AND NOW – IN THE CATEGORY FOR MOST INTERESTING METADATA, THE AWARD GOES TO.....

OOPS, got a little carried away. The goal of this article was to provide some understanding of the importance of Metadata, what it is, where and how it's used AND hopefully help demystify some of the terminology.

This is part of Olson's continuing series "[Smoothing the Rocky Road to IP](#)". Other articles include:

[The Anatomy of the IP Network, Part 1](#)

[The Anatomy of the IP Network, Part 2](#)

[The Anatomy of the IP Network, Part 3](#)

[Changing Technology: Panacea or Pandemonium](#)