Original Article

Metadata: Principles, practical application, best practices, optimization and workflow

Frank Chagoya

has made his presence felt in the Print Industry over the past 20 years by seamlessly incorporating innovative technologies into traditional print production processes. He is currently the Executive Digital Asset Manager for Leo Burnett's Print Management Department, where he integrates and deploys Digital Asset Management (DAM) Solutions for effective global reach. He helps to identify and evaluate the processes and technologies needed to streamline digital workflow for many multi-national clients. In doing so, he has achieved greater production efficiencies, faster global delivery and increased profitability for Leo Burnett and its clients.

ABSTRACT Digital Asset Management (DAM) systems are continually impacting production and archiving processes. Post development of taxonomy and metadata schemas as well as sourcing data from various workflow documentation and even data embedded in certain file types, both pre- and post-ingestion have generated a new impact that is now being realized. There is a growing need to standardize how and, more importantly, where we source our metadata. Searchability and ease of use is greatly impacted in a DAM if the taxonomy and metadata schema implemented is too confusing to use or too cumbersome to apply. Only the most commonly used details about an asset should be used to as its metadata. Information about its lifespan or its production 'paper trail' should also be easily accessible. The extended information should be either associated or attached to the asset or group of assets that share this common information. This extended file should also mirror the paper documentation that follows the asset in its pre-existing workflow. This longhand version, the secondary asset, should be easily accessible if anyone requires an in-depth review of the asset for purposes such as validation, compliance or reconciliation.

Journal of Digital Asset Management (2010) 6, 257–261. doi:10.1057/dam.2010.27

Keywords: end user; taxonomy; metadata source; embedded metadata schemas; workflow

Digital Asset Management (DAM) systems seem to be perceived more and more widely as the answer to all our digital repository dreams.

The reality boils down to this: you can put almost anything into a DAM but it's of no use if the end users can't work with it easily, let alone find it. In the not-so-distant past (and for those who have not yet taken the leap into the DAM pool), files were organized in multi-tiered folder structures and stored on a wide variety of magnetic and optical media, leaving us captive to searching through digital or hand-written catalogs, directories or inventory lists to find

our assets. Most often only a small number (two to three) of alphanumeric references were used as search criteria.

Today's DAM systems can house and manage hundreds of thousands of assets unbound to folder hierarchy. Metadata, the structured data that is applied to an asset, is the integral information the DAM system searches through to return the asset we are looking for. Here, organization of these assets is based on association via common metadata. Now a DAM search engine can locate these assets with lightning speed. Or NOT. 'Hey Frank,

Correspondence:

Frank Chagoya
Agency Technology
Solutions, Leo Burnett, 35
W. Wacker Drive, Chicago,
IL 60601, USA
E-mail: frank.chagoya@
leoburnett.com

can you pick up that image of a needle we used three years ago for that tailor-made suit ad we did?’ Without a practical, well-vetted and structured taxonomy and metadata schema the search criteria may not be able to find that proverbial needle in the haystack.

TAXONOMY AND METADATA PRINCIPLES – WHAT YOU REQUIRE AND WHY

The principles on which to base a selection of taxonomy and definitions can be extensive and complex or simple, clean and intuitive. As the person(s) responsible for the decision making in the development and implementation of a DAM system, if you are already tied into the lifespan of the assets to be impacted you will start off with a good perspective to work from. Keeping an open mind to the needs of the other people tied into the asset lifespan, you should include the data they apply and the data they review in the requirements scope. If you are not an end user then there will be quite a bit of learning you will need to do in order to expect success of your DAM.

Every entity implementing a DAM system struggles with the grand scheme of taxonomy, the terms and their definitions that will be used to make up the metadata schema for assets ingested in the system. Part of the struggle is determining how much is too much and who is eventually responsible for applying and using this metadata. Concern over metadata accuracy can lead you down a grueling path to keep you away from the ‘Garbage In, Garbage Out’ dilemma. As we should all know by now, Mr Murphy and his damn law can play a part in interfering with accuracy, which in turn can lead to the lack of searchability of your assets. This is where standardization via validated lists comes in, to help insure accuracy. Standardization can be a beautiful thing, but keep in mind that it is *best* applied where it is *best* suited. Specifically, this should be reserved for lists or metadata sets that are fixed or change infrequently. If validation lists are applied to ever-changing or growing metadata, it will require frequent management, such as weeding out of obsolete entries, to keep such a list in check.

PRACTICAL APPLICATION – HOW DATA IS CURRENTLY MAINTAINED AND HOW CAN IT BE MADE MORE USEFUL

Your best scenario is to first determine the appropriate trail of data and who the end users are, then get them all together in one room and iron out the practical details together at one time. This will not only provide you with a faster return on your efforts, but also serve as a venue for discovery of the various ways taxonomy and terminology are perceived or interpreted. Here is an excellent opportunity to capitalize on by introducing standardization and promoting adoption of the resulting taxonomy and definitions with the end users’ input.

The paper trail should reveal the most common details and descriptors for an asset. The history of its lifespan will also provide the various people that impact, manage and route the asset. These people will most likely be the same that will ultimately annotate or apply information to the asset through the DAM so consideration for them and their input should be valued as well. When it comes down to adding the responsibility of tagging an asset with metadata, I have found that the end user will immediately opt for the shorthand version and rightfully so, knowing how taxed so many people are these days. This is another reason to keep things simple and lean. Your DAM should make this longhand version the secondary asset, easily accessible if anyone should require an in-depth review of the asset for purposes such as validation, compliance or reconciliation.

Determining what is practical to the end users should not be a huge undertaking. The important thing to focus on is how the end user currently acquires the information they need. Taking into consideration their daily routines and workflow is key. Discussing requirements with them will get you fast answers, but reviewing the documentation end users rely on as well as who provides them with these details are also critical to ironing out the appropriate taxonomy. Following this trail both up and downstream in an asset’s lifespan will get you the best practical selection of metadata to be applied. Generally speaking, the most common details about the asset that follow it and are

quickly searched for or recognized should number around 12. Consider the following:

Job Number, Item/Sub Number, Task Number, Version Number, Job Description, Item/Sub Description, Asset Type, Asset Description, Facility/Office Identification, Department Identification, Originated By and Approved By. Other inherent details are obvious and may most likely be automatically applied, such as Asset Name, File Type, Size (megabytes), Dimensions, Date of Origin and Last Modification.

TIERS OF ASSET DETAILS – CASCADING METADATA

I define a metadata tier as a level of details where the schema is based on specific criteria such as an asset type (that is, document, image or artwork). The tier schema will change to reflect further definition of the asset type. For instance, a document may be specifically related to business, creative development or production specifications. The tier would generate a cascade of more relevant metadata fields. This is where the schema may become more and more complicated for each tier of metadata.

One can imagine just how complication sets in when you consider that there may be common fields, terms and validation lists between tiers. Add to this the fact that an end user may have access privileges to only see or edit certain tiers or metadata within tiers. The reason to keep things as simple as possible should become more evident by this point. This is why I feel that a secondary file with all pertinent details for a primary asset would better serve the system and the end user. Providing a cleaner, more simplistic schema will reflect ease of use and less intimidation and frustration.

DAM EXPECTATIONS – THE REALITY OF HOW METADATA SHOULD BE USED IN A DAM SYSTEM

My basic principle for Taxonomy and Metadata is this: keep it as simple and uncomplicated as possible for the end user. The end users, of course, will be the people who will ultimately apply the metadata and the people who search for the assets and also impact their lifespan.

Make sure the taxonomy parallels their pre-DAM workflow. Although the reason for this may seem obvious, I feel that many times it may be overlooked because the people responsible for the implementation are not the end users themselves. The goal should be to provide a system or tool that will be intuitive and easy to be adopted.

How complicated can it get? Well, the number of things that describe an asset can range from a few key words to a multiple-page history or 'paper trail'. I believe that this longhand version of metadata should be retained in a separate or secondary asset that can be associated or attached to its respective primary asset. I will get to that in a bit. For now, I will focus on the primary asset.

One thing that I believe gets overlooked all too often is that a DAM is much like the Internet in the way that it is capable of housing megatons of content that can be searched through to bring you back just as many results. The more assets and metadata that are searched through the longer it takes to present the results. Granted, today's search engines are incredibly fast; nonetheless, the more data there is to search through, the longer it will take to return results. A DAM system will only continue to expand as more and more assets are ingested. Keep your schema simple and lean to keep your system fast. Also keep in mind that the end user doing the searching is looking for the asset itself and will want to use concise terms to find what they need. Large amounts of metadata should not be needed to make an asset findable. Once found, the asset can be reviewed for the content that makes it important to the end user.

HOW TO GET YOUR DAM SYSTEM TO FIT INTO YOUR WORKFLOW – GETTING THE METADATA APPLIED AND BY WHOM

Getting the DAM system integrated into your current workflow is by no means a simple task. During the initial stages of integration it will actually parallel many manual data application tasks. Although this will likely create redundant work, as the integration progresses the older workflow will give way

to the newer digital and more efficient data entry. This process will affect most users that impact the asset during its lifespan, so naturally it is imperative that all are 'onboard' with the workflow changes. This is where involving all end users in the development stages is so helpful in promoting the importance of full adoption to ensure DAM success.

When it comes down to adding the responsibility of tagging an asset with metadata, I have found that the end user will immediately opt for the shorthand version and rightfully so, considering how taxed so many people are these days. As I mentioned earlier, a DAM should make this longhand version, the secondary or associated asset, easily accessible if anyone should need to review it for purposes such as compliance or reconciliation. This file should be either directly attached to the primary asset or referenced in the metadata so that it can be easily retrieved.

Where the DAM system plugs into your workflow depends on how your DAM is set up. Some may use a DAM strictly as an asset repository for distribution or archival while others use it for work-in-progress. No matter how the asset originates on the system, the information that is applied at that point should be the basic essential details for that asset. As the asset progresses through its lifespan the end users that manipulate it should also apply additional detail as they normally would on any documentation that normally follows the asset. Although you may not be able to mirror the document format on the DAM system Graphic User Interface ((GUI)/input screen), the metadata schema should reflect the most common fields for the asset in each stage of its life. It is possible however, to have the secondary asset itself mirror any of the lifespan documentation, and because these two assets should be associated with each other, all details should be easily applied and searched the same way as in the original workflow. I believe this would be the most practical means and the best practice for extended metadata application. This would involve a consolidation and streamlining of the asset's information. An additional benefit to this practice would be that this secondary file could also be associated with or linked to any other asset involved in a project, such as

component pieces to an ad or an entire campaign. This more-optimized and more-efficient application of metadata would ultimately relieve people from applying this information to multiple files or documents as most likely is already being done in the original manual paper process.

Depending on the asset type or other delineating criteria such as project type, product or client, an appropriate template for the secondary asset would be used to apply related details. This presents another excellent opportunity. Here you can compile multiple forms or documents into one file that aggregates information that may become dissociated from the asset in its lifespan. This is especially true when an asset moves from one department or production phase to another. Implementation of this type of process would also facilitate rights management, which is a growing concern in many industries.

Generally, the initial user will be the person that uploads, ingests or generates the asset in the system. This person will also most likely be the one to apply the first pieces of metadata or annotate the asset. So right from the start the assets should be easy to bring into the DAM and add its critical details. Most DAM systems have facilitated this first step by providing the ability to batch upload or ingest multiple files at one time. *Ideally*, the primary and secondary asset would be ingested at the same time and have common details applied at once which may also create their association. Thereafter, the primary asset would be annotated progressively with minimal essential details and the secondary asset with more in-depth lifespan or production details, again that mirror the original workflow documentation.

Thanks to technology and the development of Extensible Metadata Platform (XMP) and Extensible Markup Language (XML) we now have 'smart assets' which are able to carry required applicable data with them. Growing numbers of DAM systems are utilizing and applying this data directly into respective fields in their schemas. This XMP is where someone may have added details to an asset's 'Information' or 'Properties' schema fields. If these fields are commonly utilized by your workflow it is essential that your DAM map or direct this data into appropriate fields of your metadata schema. Standard schema such

as the International Press Telecommunication Council or the Dublin core standards are becoming more widely used even though customization of schema is available.

THE FUTURE OF YOUR DAM – THE EXPANDING UNIVERSE THAT IS DAM

DAM systems are supporting a growing range of asset types as businesses press for expanding their use of these systems. Content management applications have been moving into the DAM arena to help capitalize on the inherent data inside the assets. Being able to utilize search engines to read the content in the assets themselves brings incredible versatility to the fingertips of people needing to review documents for research, reporting and more. Character and image recognition applications, along with the capacities to utilize embedded XMP/XML data, are making searches and use of data easier, but how does this impact affect your use of the DAM? You may find a re-visit to the drawing board necessary. (Mr Murphy enters the scene once again.) Here are a few things to look out for.

Does the available data

- adapt to your DAM;
- fit your schema;

- fit/match your taxonomy; and
- require validation.

THE EVOLVING DAM PARADIGM – LEADING THE INDUSTRY WITH NEW TECHNOLOGY AND APPLICATIONS FOR DAM

Having attended and participated in DAM events over the years, I have had great opportunities to gather information and (one of my favorite expressions) ‘see what there is to see’. DAM has come a long way in a short time and I believe still has a long way to go. If you are like me and seize these opportunities, you may be surrounding yourself with ‘people in the know’. I want to know too! So don’t be one who thinks but regrets not asking, ‘what else does it do?’ or ‘can it do this ...’ or ‘why not?’ or ‘when will it be able to do this?’ In *my* head, the thought that ends this list of questions is the one I ask myself; How can I make this work for us?

Keep the wheels spinning in your head and keep pushing developers to give us more. This paradigm should continually shift to improve the way we work and are able to do our business. Keep this evolution in constant motion.