

Extending Metadata into the Enterprise

Taras Bugir
Sun Microsystems, Inc. USA

Content delivery platform providers (Broadcasters, Cable, Satellite and Telco) are aggressively evolving their infrastructures for delivering rich media in a bid to engage more subscribers with a rich portfolio of diverse content offerings. To date, managing metadata effectively has been essential in supporting workflow efficiencies within the operational context. However, this operational focus on metadata has largely ignored content management across increasingly distributed networks, multiple cultural environments and multiple device requirements. Such issues can directly impact the scalability, and hence, profitability of operations, often necessitating heterogeneous infrastructures with their resultant management headaches.

INTRODUCTION

Organizations that leverage metadata upstream in their value-chain can enable greater automation and scalability of downstream workflows. Therefore, metadata collection, enrichment and management should ideally extend across the enterprise. This approach ensures that departmental activities such as sales, scheduling, CRM, subscription management, access management, rights-management, financial systems all compliment their behaviors to drive the business.

This documents presents the case for, and outlines an approach to, handling metadata that enables providers to profit from a high level of granular control over the flow of content through their infrastructures. It discusses the issues, tools and workflows that will drive the future of content management. Through judicious enterprise management of metadata, organizations can synchronize the descriptive and instructive aspects of content to identify appropriate audiences and subscribers, determine the profitability of associated commercial opportunities, and subsequently move content to multiple consumer devices with greater business and operational flexibility.

WAYNE'S WORLD

This popular 1992 comedy film starred Mike Myers as Wayne Campbell and Dana Carvey as Garth Algar, hosts of a cable access television show called Wayne's World¹ from Aurora, Illinois.

At the time, it seemed quaint that a couple of 'dudes' in a basement would have an audience beyond their community and a minor cult following. However, fast forward to today's user generated content environment on YouTube with the tagline of "Broadcast Yourself" and the concept is now *de-riguer*. What's more, such content on focused channels generate advertising CPMs, that would enable today's Wayne and Garth to reasonably achieve \$1M+ revenues after distribution split with a faithful following of just over 1000²!

What's happening? Is the model failing? Is the audience fracturing? Is it a generational thing?

In 1922, Thomas Edison declared³ *"The radio craze will die out in time."* a few year later, in 1946, Daryl F Zanuck, President of 20th Century Fox at that time stated⁴, *"Television won't be able to hold on to any market it captures after the first six months. People will soon get tired of staring at a plywood box every night."* The most amazing thing about media is it's 'stickiness'. Print hasn't gone yet, nor have magazines, nor radio, nor TV. In fact we have print and radio and TV on the web - a small but important step to a convergent world. What has changed is the mix of how people use different media. And with that mix, so changes the way the money flows...

Hence, the focus of this paper - designing a business infrastructure to accommodate this constantly evolving landscape. This is the true role of metadata.

MORE THAN OPERATIONS

Most discussions on metadata's role in the management of content have centered around broadcast operations or on non-linear content editing.

Missing from these discussions is the larger role of metadata in the content lifecycle. For such discussions to be meaningful, metadata must be considered in a broader context, beyond its physical operational attributes such as titles, SOM, duration, and encoding formats. One must consider the wider financial and business utilization of content, such as rights, cost, and scheduling constraints. All of this must be considered against the backdrop of changing legacy business models.

This integrated approach to metadata usage and management supports increased organizational efficiencies. Such efficiencies lead to increased opportunities for operational scale, a more cost-efficient mechanism to repurpose content and an ability to generate new revenue streams. These are what drive business value.

Consider figure 1. In its journey from the production house or ad-agency, content travels a disjointed path through the media operation. For example, a commercial spot usually makes its way from the distributor directly to the automation facility. If in a digital form (usually MXF compatible), then ingest automatically provides a way to extract the metadata. If it arrives in raw format on tape or via satellite, the ingest function requires that some metadata be generated. In both cases, the metadata is generally limited to those attributes that facilitate or 'automate' operational efficiency, such as formats, timings, descriptors etc. There is often simply not

enough time, energy or resources to invest in capturing more.

Should the material require editing, then it 'inherits' additional metadata from the editing suite in the form of the increasingly prevalent AAF⁵ standard. Often, this content is enriched with metadata, that remains unused, as it is 'flattened' for export and then re-wrapped in an MXF layer for movement over to operations.

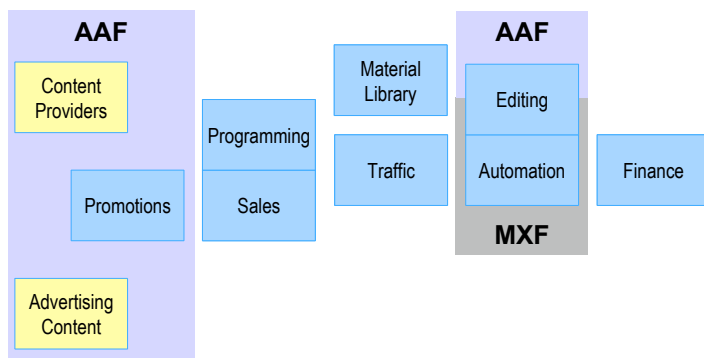


Figure 1 - Typical operational focus on metadata management

Long format items follow a similar path with the addition of segment timings, component content items such as subtitles, different language tracks, censorship items etc. Similarly, in the world of promotions, elemental clips, sound beds, voice-overs, logos, etc. are all assembled with the AFF format, or similar, encapsulating the content for downstream utilization.

Note what is missing from this description.

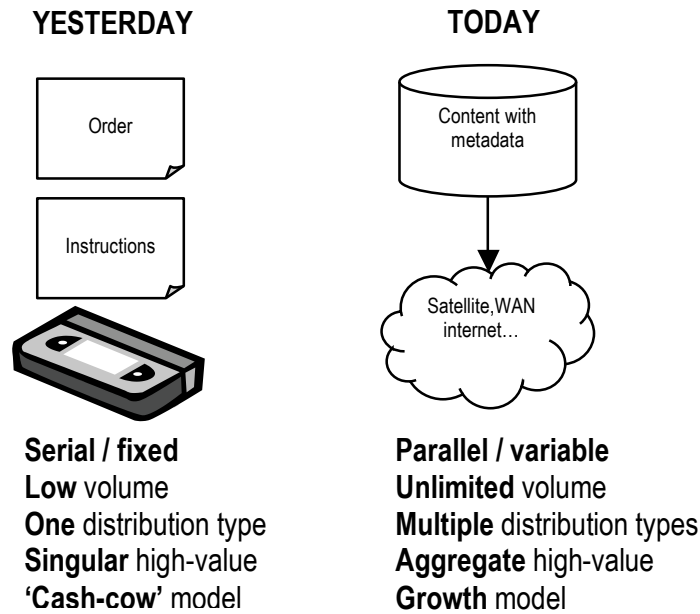


Figure 2 – Metadata – the key to business scalability

Consider the departments of Programming, Sales, Material Library, Traffic and Finance. All ‘touch’ the content from a business viewpoint, yet each department has software applications that abstract and transact the content in order that it is appropriately managed from their perspective – always distant from the content itself. I submit that this distance is the prime cause of discrepancies and inefficiency.

Furthermore, this is a workflow process that was designed for the management of single large valuable content items, not for a multitude of small, lower value items that in aggregate could potentially exceed the value of the aforementioned model. Refer to Figure 2.

SCALABILITY

Scalability is a multi-faceted discussion that is itself is worthy of a paper. However, in this context we highlight only media business scalability issues and their technology implications.

- A proven high-transaction and content volume capability with integrity is a must for any organization committed to growth in media products and services. As the business grows, so must the infrastructure along with the integrated framework components – ideally linearly. Open systems, rather than purpose built components generally, more easily support massive transactional growth.
- Such ‘size’ scalability is one aspect of growth. An infrastructure must be able to support organizational scalability, e.g. locally as well as internationally, one-medium as well as multi-mediums. Additionally, that organic growth or individual business units must be supported by functional scalability, richness of operability that supports diverse marketplace requirement.
- Most importantly, as growth occurs, the cost of that growth must become lower, on a per channel or per service basis. Experience has shown that well designed digital IT

infrastructures enable investments to be amortized across incremental channels and services, actually becoming more affordable with size, because installation, training and configuration are subsequently less.

By utilizing standards Custom Of The Shelf (COTS) solutions and components, organizations can start riding the commoditization curve as they grow. To leverage the commodity curve, requires increased automation and smoother workflows with fewer personnel per channel or service.

This further supports the case for increased propagation of metadata across the enterprise – metadata that can be acted upon, enriched and leveraged automatically at every workflow step.

CREATING BUSINESS VALUE

There are two main ways in which a business can gain a competitive advantage:

- by delivering product or service at a lower cost, or;
- by differentiating itself in order to command a premium price.

A value chain⁶ in our marketplace landscape (as for most other businesses, refer figure 3) can be simplified into the following core value-adding activities: planning & development, operations, and marketing, sales and customer service. All three of these extend beyond the immediate media organizations value chain, linking it with upstream suppliers and down-stream customers in the value system. Examples of this are agencies, media buyers, content distributors, etc. Underlying these activities is the financial management and control of the business, forming in effect the fourth core value adding activity.

For a media organization to remain successful, each of these value chain elements must be addressed, either within a limited service offering, or by providing a platform for integrated solutions.

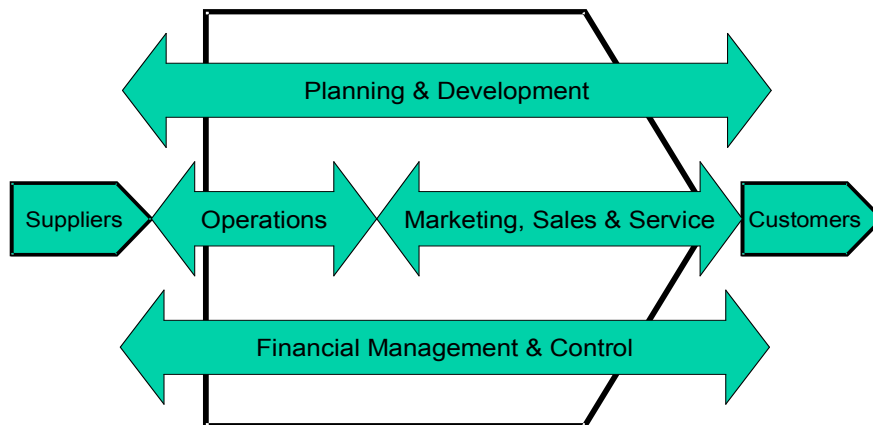


Figure 3 - A company's value system comprises of four core activities

If we accept the fact that the very essence of broadcasting is about providing the right content to the right consumer in order to create business value, then it follows that those businesses which

manage content seamlessly along every link in the broadcast chain will be inherently more efficient.

Consider what those isolated applications in Programming, Sales, Traffic, Library and Finance have in common. Each has a data structure that describes such things as the state of the content i.e. has it been acquired? how and when it can be used, how much it costs, where the material is located, on what media, has it been dubbed? has it been paid for? was it profitable? were the spots invoiced? etc. and many other issues that must be tracked in order that the business maintain financial viability.

All of these individual data structures are function and workflow specific. However, from the enterprise’s perspective, are they not merely subsets of what one knows about the content and its usage at any given point in time? Consequently, we conclude that since content is increasingly expressed in digital form, along with its metadata, then is this not just another specialized data structure, albeit more dynamic in nature?

So if we could define a common metadata structure that provides all of the information or data to run all aspects of the business, then each department, or link in the broadcast value chain could share the same data at all times.

The consequence of this would be an integrated workflow. Because the workflow would be defined by the state of the content, and its usage at any given stage of the business and operational cycle, direct business value is created by operational determinism.

WORKFLOW INTEGRATION

Intuition suggests a technology ‘backbone’ across the enterprise. At Sun, the IT infrastructure is considered to be that backbone in our Media & Entertainment Framework⁷. An open infrastructure that supports a strong Services Oriented Architecture (SOA), enables the integration of component workflow solutions that maximize scalability & re-use. Consider the following framework in figure 4.



Figure 4 – Sun’s Media & Entertainment Workflow Framework

Each of the framework elements are briefly described below:

Planning & Sales Systems - Supports the key “business” aspects of content and its placement,

including long-form acquisition, selling advertising, managing the financial analysis, and operational execution of profitability and revenue optimization.

Media Finance - Enables the financial activities of the business, in the context of transmitted content, be it scheduled, subscribed or paid. Manages the invoicing of advertisers and/or subscribers, as well as paying for content rights, distribution costs etc. Leverages the output of Planning & Sales, and costs from other business systems as sub-ledgers, utilizes the Data Warehouse for Business Intelligence for financial projections, budgets & planning.

Data Warehouse - Aggregates all financial transactions for forward business planning. Consolidates operational and audience data for research, subscription management and reporting purposes. Acts as a repository for all transactional history of the organization to assist with streamlining operational costs, maximizing revenue and mining for new business model opportunities.

Acquisition & A/V Services - Technical shared service for the organization. Manages the acquisition of content, content libraries (both physical and digital), ingest, QA, verification of marketplace legislation for content, encoding and transcoding, routing and quality control of technical standards for the operation.

Catalog & Archive - In addition to housing the content 'essence', supports the repository for, and the generation and verification of content metadata. Preserves and protects the content, and manages the physical archive library. Includes the archive for all content, both video as well as services content, and may include the transactional archive for business systems.

Production - Supports the content production and post-production of content for both static programs and time-critical programming such as news and sports.

Content Distribution - Supports the conditional access/digital rights management, watermarking/fingerprinting of the content. Effectively stages the content for final shipment outside of the protected environment whether to another facility or to a consumer. Also provides the subscription management functionality that integrates back to consumer billing and CRM systems.

Service Delivery Platform - The production of applications and management of support systems which facilitate content to be distributed over an IP delivery channel and manages the return path for value added services. Typically these are VOD, feature search & selection, EPG, games, shopping etc.

Delivery - The infrastructure to deliver the content, be it audio/video, data, interactive content, VOD etc. over the marketplace defined distribution channels.

Infrastructure & Network - Monitoring capabilities for systems and equipment that manage the delivery of Video over IP.

Consumer Edge - Includes the infrastructure of hardware, software and middleware that enables the consumer to correctly receive the content on the specified device and interact with it in accordance with the devices capabilities.

IT Infrastructure - All of the supporting IT systems for internetworking, storage and office systems. Provides an integration and workflow layer to support the management of digital content over the entire business framework, and is the infrastructure for transporting messages and content between non-consumer facing systems.

By adopting this approach, and considering the integration of the various component pieces across such a framework, the enterprise enables maximum integration of workflow, with resultant metadata consolidation. The natural consequence of such an approach is functional scalability.

TYPES OF METADATA

The model described in figure 5 has one more metadata type than is generally discussed or even acknowledged – *transactional metadata*⁸. It is this type of data that bridges the business functions with the operational mandates of the enterprise. Consider the following model:

Transactional Metadata	Instructional metadata that enables organizations to direct how content is to be managed. So, a Telco in Southern China may choose to specify that a particular version of "Wayne's World" be transcoded for local consumption on a handheld device for delivery to its cellular subscribers.
Applied Metadata	The addition of applied metadata to core metadata to support specific contexts. For example, it was dubbed into Cantonese, and so a media organization in Hong Kong may leverage that version, edit it for censorship regulations and change the aspect ratio.
Core Metadata	Fundamental information about content. This information is instantiated by the very production of the content itself, it never changes because it identifies the original content. An example is "Wayne's World", produced in 1992 by Paramount, in color, with sound in English, aspect ratio of 1.85:1 and runs 95 minutes.
Essence	The actual digital 'content'

Figure 5 – An integrated metadata model

Transactional operations are performed with systems today. This is not new. Although, they are external to most content models, they interact directly with the content. Examples of such applications are content schedulers, which check dubbing requirements against channel defaults. What is new, however, is populating the content itself with such metadata.

There has been very little discussion about this type of metadata, because, typically it has been treated as 'data' within transactional systems. However, there is no reason to consider this as being separate from an overall integrated business workflow.

Examples of *transactional metadata* are:

Conditional Access - This metadata can be held as properties of the content but is more often customized for each transmission event of the material. The instructional data defines the logical conditions that must be met for viewing, based upon each entitlement that a viewer could purchase. It is often only used to attach Entitlements to Subscription channels and PPV programming with some geographical blackout control. However the potential extension of this

enables highly targeted television events and commercials. This is an example of metadata information that is transacted with the viewing device by the content.

Program Contracts - This important information is associated with the program titles and instantiates the detailed rights and contractual constraints for the broadcasting of the title. In current business models it is currently used only to schedule and then validate the planned transmission of programming. In a pull world, this information becomes immediately essential to support conditional access and financial mechanisms.

Material Instructions - This metadata is sent by advertisers (or promotions departments) to provide detailed rules for the assignment of content to booked campaign spots. This includes rotation arrangements and frequency of display information. Such information along with commercial orders (refer next) could be intelligently utilized by 'content-aware' placement algorithms. In effect, content could instruct the system where it should be placed – all without human intervention.

Interactive Packages - Another class of material metadata is the linkage of all the related content elements that are needed to compile an interactive TV event. In addition to video, content can be as diverse as HTML, Java as well as static and dynamic image formats. It is important to recognize that script files that instruct the flow of control and display are themselves content and in theory, these scripts can be as complex as a traditional e-commerce web site.

Process Status - The communication of content status to business processes is essential to the management of the preparation of content into a transmission-ready state. This metadata contains all the job steps, their assignment, due dates and completion status. While important during the execution process, it is only of historical value after the material is completed and approved.

Program Format - This metadata is related to the transmission instance of a program title and defines the detailed structure of the breaks, segments and other interstitial events. This information is not only important to the preparation of the transmission schedule but is now being treated as essential metadata to be delivered along with the content to inform receiving parties of the structure. This extremely valuable for affiliate ad break overlays and if sent to the viewing device, could easily support local targeted content insertion.

Clearly this is not a comprehensive list of business driven information that could be embedded into the metadata model, but it does illustrate the point that capturing and enriching metadata upstream of operational processes supports massive automation of downstream operational processes.

METADATA UNIFICATION STRATEGIES

According to Gartner⁹, the integration and consistency of metadata across technologies does not happen automatically. It requires hard work, and there is limited technology automation available to assist in the process. Gartner offers four approaches to accomplish this integration — bridged, physically integrated, extracted and an integrated view — all with advantages and disadvantages. Refer to table 1.

Furthermore, Gartner states, “Alternatives that attempt to physically synchronize or consolidate metadata are usually constrained by semantic problems that can limit what can be shared. Depending on the degree to which the organization wants to resolve inconsistencies among the metadata, political and cultural issues can severely and adversely affect the resolution of

semantic differences. These factors include metadata ownership, ongoing costs and potential bottlenecks, such as the analysis and administration of changes. Regardless of the metadata integration approach implemented, due to differing technical and business semantic issues, it may not be possible to reconcile all metadata across the various technologies”

<p>VIRTUAL-BRIDGED</p> <p>Create a view of the metadata across physical databases</p> <p>Scalability & Change Issues</p>	<p>PHYSICAL-BRIDGED</p> <p>Common metadata model, with messaging infrastructure</p> <p>Hard to Build - Standards</p>
<p>VIRTUAL-EXTRACTED</p> <p>Extract and generate metadata cache for scale</p> <p>Accuracy Windows</p>	<p>PHYSICAL-EXTRACTED</p> <p>Create a new metadata repository by extracting data</p> <p>Synchronization Problems</p>

Table 1 – Metadata consolidation approaches

The actual approaches to consolidating metadata can quickly become very complicated and are organization specific. There is a great body of literature on this topic, and it is beyond the scope of this paper, where the focus is on presenting a case for such consolidation. However, it is noteworthy to summarize that the approaches are fundamentally

- Connect (Bridge) to existing repositories, or
- Extract required metadata from existing repositories

to create unions of abstracted metadata (Virtual) or;
to create new instances of metadata (Physical).

ENTERPRISE METADATA OBSERVATIONS

Sun’s experience with integrated applications suggests that an SOA¹⁰ approach with integrated business operations can reasonably achieve efficiency gains of over 20%. This figure is consistent with organizational studies from supply chain management efficiencies in retail and manufacturing. Furthermore, our experience in deploying systems in Communications, Media and Entertainment has enabled us to make and share the following observations on the impact and value of metadata to organizations:

Operational efficiency decreases in proportion to the degree of metadata abstraction

Basically, the further your business transaction applications are dissociated from the content and its metadata, the more issues you will face in maintaining information in synchronization between the business and operational domain, and the more likelihood of discrepancies and common business practices in your enterprise. This directly impacts efficiency.

System integration effort increases geometrically with the number of metadata databases

The more applications you attempt to integrate with tools that really only facilitate inter-process or inter-application messaging, then the more permutations you have in ensuring that data is consistent across the enterprise. Complexity increases as a function of $n*(n-1)$ database application connections.

Organizational nimbleness is proportional to the number of metadata enabled workflows

If you have to re-key, re-populate, refresh and re-check data and metadata then your ability to scale becomes a limiting factor. The number of parallel workflows can reach economic or operational limits. This is a direct indicator of your ability to quickly change as an organization.

BENEFITS OF ENTERPRISE METADATA MANAGEMENT

Many enterprises have grown from a series of established business practices and subsequent acquisitions and divestments. Clearly, changing this overnight is not realistic. It is essential to define a systemic approach to metadata consolidation and management by creating a framework for integration¹¹ that will sustain scalability, without compromising the speed of deployment.

Some organizations believe that, even with the right approach, capturing enough information to make their metadata really useful or having enough information management personnel is an unquantifiable investment. There are new approaches to automatic metadata management that can overcome such concerns. Nevertheless, a better understanding of organizational metadata can provide several important benefits specifically;

- Increased reuse of established assets
- Reduced change impact
- Identification of redundancy – consolidation
- Facilitate identification business overlaps - streamlining
- Reduced deployment costs
- Savings in training costs
- Accuracy of analysis – consistent data
- Better organizational management decisions

CONCLUSION

As modern media organizations consolidate their holdings, aggregate their media assets and revenue streams, diversify their offer to compete with consumer pressures and new-media offerings, there is only one certainty – the old advertiser-funded, schedule based business model of broadcasting, will not scale and be as profitable in the future.

All around us we see traditional media models:

- becoming increasingly fragmented and ‘multi-media’ in nature
- requiring parallel workflow streams
- for profitability, focusing on less direct staff involvement
- increasingly sensitive to, and stressed by, scalability

At the same time, media and entertainment is becoming more IT-centric. Content is becoming data (e.g. file-based workflows and distribution) and data is becoming content (e.g. RSS feeds). And both are becoming the ‘base currency’ of the digital media enterprise.

Correspondingly, content producers increasingly need to protect their content as consumers challenge that right by demanding access to media across many more platform options. Media operators will need to maintain their distribution costs while at the same time managing much more content and offer many more delivery options.

This is an opportune time to consider the consolidation of metadata across the entire enterprise.

REFERENCES

-
- ¹ Wikipedia - [http://en.wikipedia.org/wiki/Wayne's_World_\(film\)](http://en.wikipedia.org/wiki/Wayne's_World_(film))
 - ² Private client discussions – please contact Author for further information
 - ³ Famous Authoritative Pronouncements - <http://www.av8n.com/physics/ex-cathedra.htm>
 - ⁴ BrainyQuote.com - <http://www.brainyquote.com/quotes/quotes/d/darrylfza401896.html>
 - ⁵ AMW Association - <http://www.aafassociation.org/>
 - ⁶ “Value Chain Framework – Porter” - http://www.valuebasedmanagement.net/methods_porter_value_chain.html
 - ⁷ “Media Workflow Overview”, Taras Bugir - July 30, 2007 – Sun Internal Publication
 - ⁸ “Metadata: The Backbone of Operational Efficiency”, Taras Bugir - November 26, 2005 - Harris Corporation
 - ⁹ “Metadata Management Technology Integration Strategies”, Michael Blechar - July 11, 2005 - Gartner Research
 - ¹⁰ <http://www.sun.com/smi/Press/sunflash/2006-12/sunflash.20061215.1.xml>
 - ¹¹ “Harris’ H-Class Platform for Metadata Usage”, Taras Bugir – February 2006 – Broadcast Engineering