Roll out the Red Carpet: Metadata makes its mark in Hollywood
By Matt Turner, CTO Media & Entertainment, MarkLogic

Not since the heyday of the Great Library of Alexandria has metadata been so popular. Back in 280 BC, grammarian and literary critic Zenodotus ensured his library staff attached a small tag to the end of each scroll, which contained information on each work’s author, title and subject. This way, materials could be easily returned to the area in which they had been classified while giving library users information about each scroll without unrolling it.

Today, metadata is becoming a Hollywood A-lister and playing a vital role ensuring amazing content makes it to the fans who love it. Without metadata, there is no way for the entertainment industry to keep track of the millions of pieces of content and associated details critical to keeping viewers entertained and engaged.

Seismic Data Shifts in Hollywood

The entertainment industry has used data to inform decision making processes - for example using ratings to decide whether to axe a show – but the amount and nature of data available has increased exponentially. Ratings figures are now only part of the picture. Direct to consumer engagement means more data on how, where and when people are viewing and consuming content, while user profiling and behavioural tracking reveals a much more complete view of the audience.

In parallel, there have been huge changes around how audiences consume content with the entertainment industry shifting from analog film and tape to digital file and networked delivery including direct to consumers. With thousands of productions in development at any one time, each an isolated island or silo generating upwards of seven or eight terabytes of data a day, piecing together the metadata generated throughout each step of the process and making sure that data is as useful as the Great Library of Alexandria’s tags is a task that would send Zenodotus and his team of librarians into a spin.

Metadata: It’s complex

Metadata itself is complex and multi-faceted: essentially, it’s all the data around an asset that can be categorised broadly as:

- Descriptive data, which analyses the content in more detail, for example the names of the actors, characters and location and is now being applied at scene level to help find content.
- Technical data is information about the digital files and includes data which comes off the camera as well as the technical attributes of the actual digital content. This is not to be sneezed at bearing mind the typical camera generates roughly a terabyte of content an hour and there could be hundreds of versions of any one item of content across each part of an organisation’s operation.
Transactional data can include details pertaining to rights, funding or insurance. This is often generated in the planning phase when funding is secured - suffice it to say, a tremendous amount of information is created before the clapperboard has made its first snap – but it also has to be updated and tracked as the content is produced and released.

Getting to Grips with Metadata

Just how do the leading entertainment organisations tackle this critical data? Using the real-life example of “The Dark Knight”, Warner Bros. quickly and efficiently delivers metadata to over 150 territories worldwide to all to its digital partners using Smart Content, the white knight of metadata management, which is a combination of NoSQL data and semantics.

At MarkLogic we have pioneered this approach to managing metadata, which liberates media companies and content owners from having to funnel data into the confines of rows and columns associated with relational databases and extract, transform and load (ETL) processes that have many limitations.

Traditional, relational data approaches require you to model upfront the way the data will be used for a given activity, determining in advance the way the data is stored and how it will be retrieved. Using this approach to manage metadata creates metadata suited for a single purpose which creates silos of data throughout the organisation and forces organisations to create complex and costly ETL processes to get access to the data. Then, when the organisation needs to incorporate new types of data, the schemas and ETL processes need to be redesigned and re-executed.

In a fast moving world of entertainment, where the use of metadata is ever evolving, the schema is unlikely to be known ahead of time, Warner Bros. found it increasingly time consuming, costly and inflexible using traditional ETL tools to integrate its metadata.

To meet these challenges, Warner Bros. and other leading lights in the media and entertainment industries have turned to Smart Content – using NoSQL data and semantic relationships to manage metadata and make it available for any purpose. To integrate hundreds or thousands of pieces of data from across the supply chain, these organisations leverage NoSQL’s schema flexibility, enabling metadata from multiple sources to be managed in a single system.

To fully capture the context of metadata, media organisations are using Smart Content to describe and link that data from every part of the supply chain. Semantics allows organisations to use taxonomies and ontologies to model and manage complex relationships in data. This includes classifications, genres and relationships between titles and products as well as how characters, films and series fit together alongside other data from production.
Smart Content captures the complete picture of data in the source systems, accommodates new data and new sources, and allows organisations to use the data for new purposes without the time and expensive of rebuilding the system.

Warner Brothers is now reaping the business benefits of having a scalable and flexible system, which is both operational and transactional, to give them a complete view of their data and metadata.

Another example of how the entertainment world can combine different types of metadata from across the supply chain is NBC’s *Saturday Night Live*.

As part of the promotion to celebrate the 40th anniversary of its iconic show *Saturday Night Live*, NBC created a new *SNL* app. In addition to hiring a ‘metadata army’ to watch and ‘tag’ information from old *SNL* clips, NBC linked and enriched that metadata with semantic ontologies.

By combining the different types of data from the supply chain, the NBC team was able to fully leverage its enormous library of content and create a personalised experience for fans of the show. The app was highly successful, holding the number one position in the app store for several weeks, and delivering over 100 million videos to fans.

Exploring how to capture and re-use metadata in the production cycle was a key goal of another Hollywood project called The Suitcase from The Entertainment Technology Center at the University of Southern California (ETC@USC). This project used new technology approaches in production including cloud storage and C4 IDs. One of the goals was to prove that a complete set of descriptive data could be generated from the assets and data generated during production and didn’t need to be re-created downstream. The film had its Premier in Hollywood in May which included an interactive application powered by Smart Content and MarkLogic using the actual production data. The findings from the project will be delivered as part of a paper presented at the Society of Motion Picture & Television Engineers in October later this year.

**Linking it all together**

Entertainment is a serious business. In today’s multi-platform environment, broadcasters need to maximise their ability to reach audiences by creating content that engages and retains eyeballs and can be effectively monetised. Keeping everyone in the picture by connecting all the data and content across the digital supply chain has never been more important.

Using Smart Content - the mix of NoSQL data and semantics – leading lights in the entertainment industry are already demonstrating the value of linking this data together to customise distribution, connect data in the production cycle and benefit the entire supply chain.
About MarkLogic Corporation
For over a decade, organizations around the world have come to rely on MarkLogic to power their innovative information applications. As the world’s experts at integrating data from silos, MarkLogic’s operational and transactional Enterprise NoSQL database platform empowers our customers to build next generation applications on a unified, 360-degree view of their data. Headquartered in Silicon Valley, MarkLogic has offices throughout the US, Europe, Asia and Australia.

For more information, please visit www.marklogic.com