M AND E JOURNAL
Media & Entertainment
Strategies. Solutions.

A New Age of Data in Media & Entertainment

Data is the single most important resource for players in the M&E space. For M&E companies, operating in a state of hypervigilance about consumer privacy and sensitivity to usage concerns while mining, managing and monetizing that data is the challenge of the day. P. 19

THE AGE OF DATA

TRANSPARENCY

published by

MESA
Media & Entertainment
Services Alliance

The Pattern to Unlock Smart Content

*NoSQL and semantics flexibly bring data together*

By Matt Turner, CTO, Media and Manufacturing, MarkLogic

Abstract: The data around content, sometimes called smart content, impacts everything including how quickly we can get content to our fans and customers, the experiences they have, and the content's impact on the bottom line. Bringing this data together and being able to make use of it can be critical to success.

Everyone in the M&E content ecosystem knows the Holy Grail of content value generation is in serving the right content in the right format through the right channels to the right people. Smart content is the concept of connecting all the data surrounding content. It was introduced as a concept by the Media Entertainment Services Alliance (MESA) in 2014 and anticipated the rise in importance of data for the entertainment industry. Core to the smart content concept is the importance of data for every stage of the content creation process.

Since that time, several projects have successfully demonstrated the impact of smart content. Focusing on delivering the value of the data around the content, these projects needed to take a new approach to managing diverse and disparate data. Using NoSQL and semantic data models (instead of traditional relational databases), they were able to bring the data together. These projects used the architecture pattern called the Operational Data Hub ("ODH") to put that data into action and realize the promise of smart content.
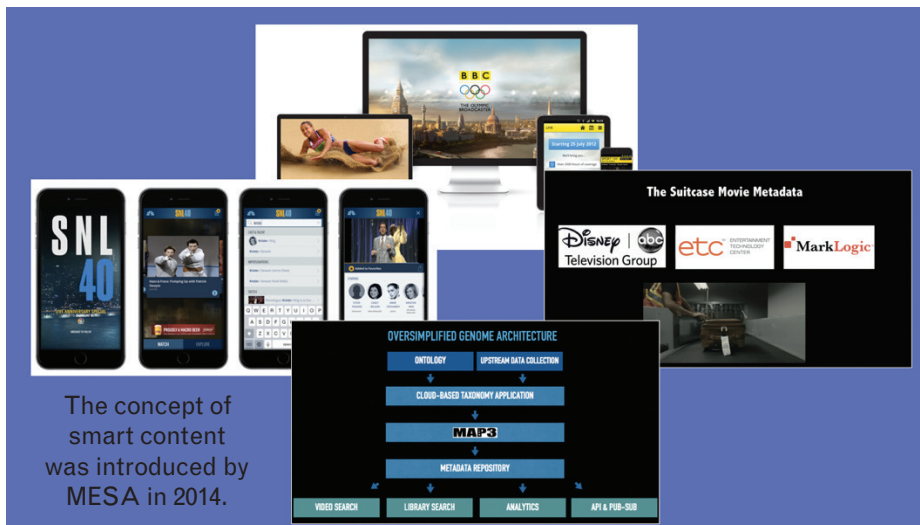
The concept of smart content was introduced by MESA in 2014.

Smart content in the industry

Since the start of the smart content concept, several projects around the industry have exemplified the value of the idea.

■ The BBC's digital delivery platform was created for the home country Olympics and is now in use to deliver metadata and content to digital audiences and external partners. This project brings together data from across the BBC's many news, content and sports sources and organizes that data with semantic ontologies. This data is made available for many uses including the BBC iPlayer, BBC Sports data and external partner feeds.

■ The NBC Entertainment Digital *Saturday Night Live* was launched to deliver a great experience and leverage the show's archive to celebrate the iconic program's 40th anniversary. The platform connected the production data with data describing what fans love about the show using NoSQL for the metadata and a semantic ontology for the themes and characters. This data made the SNL 40 app one of the most successful single show apps with Michael Martin, who led the project for NBC Entertainment Digital, saying at the Hollywood IT Summit that it allowed the broadcaster to "make a meal out of snacks" and realize the potential of this famous show's archive.

■ The Entertainment Technology Center at the University of Southern California's

project "The Suitcase" explored collecting and organizing production data for later use. Data generated during the production of a short film were then made available for use in delivery applications without having to rekey it. The project was called a "Mega Metadata Experiment" by The Broadcast Bridge and a whitepaper was presented at the Society of Motion Picture & Television Engineers (SMPTE) conference in 2016.

■ Disney's Creative Genome Project was created to collect and organize the descriptive metadata. With multiple sources of data within the organization and the rise of AI and machine learning creating new types of data, Disney's Creative Genome Project was designed to bring together all these different sources and make that data available across all applications. At the Smart Content Summit in 2017, Disney's Anthony Accardo said that the main objective was to "future proof their data" and "take a platform, not application, approach to collecting data."

Taking a new approach for smart data

These projects all shared the overall vision

of bringing together multiple sources of data from all around the content and then linking that data together, so it could be accessed for multiple uses.

They also didn't use a traditional relational data approach to manage this complex and varied data. In the traditional approach, all the possible uses for the data must be defined up front so that the resulting tables can be defined, and the data transformed to fit that schema. In addition, the categories and other reference data also need to be fixed in place with rigid hierarchical taxonomies. This is a good fit for applications that focus on a single type of data or a single use. For instance, many of the applications that generate the data around content.

However, this approach is a very bad fit for the Smart Content concept where the data from all multiple sources all around the content needs to be brought together and made available. The shortcomings include:

■ **Complexity:** The schemas to capture this kind of data can become very complex with sometimes hundreds of tables to handle just the metadata.

■ **Inflexibility:** If the data changes, all those tables need to be updated and the data changed to make use of the new information.

■ **Lost time:** All this adds up to technological complexity that keeps organizations from actually making use of the data and realizing the concept of smart content.

Because of these factors, very few traditional projects can deliver the smart content vision.

Matt Turner develops strategy and solutions for the media, entertainment and manufacturing markets at MarkLogic. He works with customers and prospects to develop MarkLogic enterprise NoSQL operational data hubs that enable clients to get the most of their data and deliver their products to the fans, audiences and customers that love them. Before joining MarkLogic, Turner was at Sony Music and PC World. matt.turner@marklogic.com, @matt_turner_nyc

Outside of the entertainment industry, the Operational Data Hub pattern is enabling many types of organizations to use the Smart Content concept to improve their processes with better data.

Bringing the data together

Instead of using rows and columns, these successful Smart Content projects used the flexible data model of NoSQL. This approach allows for flexibility in the data without having to design everything up front. This enabled these systems to bring together the data from multiple systems around the content, each with their own data models and schemas.

To link all this data together, these systems also used semantic data models. Semantic models enable organizations to model the entities around the content and define the many relationships between the entities without the limits of hierarchies or fixed taxonomies. This approach of using ontologies is key to realizing the smart content vision. They can describe complex data like characters, actors, places, and themes. They can also be used to map and link the teams and companies in the production, the finance and product data, and nearly every type of important data surrounding the content.

Semantic models also enable projects to take advantage of the new types of data being generated with AI and machine learning techniques. Computer vision can identify characters and actions in the frame. Text mining can identify the concepts, locations, dates and more. Other techniques are working on emotional, thematic and nearly every possible type of data that can describe content. Semantic ontologies can help make sense of this data and, combined with the NoSQL data about the content, create links that unify all the data about not just the work, but about any given scene or activity in the content.

Smart Content with the Operational Data Hub

These Smart Content projects all used a new architecture pattern called the Operational Data Hub (ODH) to put the flexible and rich NoSQL and semantic data into action.

This model lays out the process for creating smart content in three steps:

■ **Ingest** — Bring data into the system 'as-is'. This first step enables these systems to quickly bring the data from multiple systems together without changing the data or even defining the data models of the content

■ **Harmonization** — The process to create links between the data that enable universal access across the different data. This step includes the concept of an envelope pattern where new data is added to each record. For instance, identity information that is stored in slightly different schemas in multiple systems would have new data added to each record with the canonical schema for identity that would enable downstream systems to access that data with a single schema. This is also where the semantic links between data and ontologies are created.

■ **Access** — Once the data has been harmonized, the ODH pattern enables search, browse and data access with services. These services leverage both the harmonized schema as well as the original data schemas to enable the maximum flexibility in accessing the data.

This process for integrating content lets organizations quickly bring data together and get value from integrating that data without the upfront costs and time of traditional database approaches. It also enables organizations to increase the value of their data by continuously adding harmonization steps to generate richer and richer sets of data. The ODH harmonization step also enables organizations to leverage new AI and machine learning techniques to generate and link data to their content.

Finally, by establishing a pattern for data integration, organizations can focus on realizing the value of the data and using this same pattern to integrate any data.

Smart Content in action

The successful smart content projects use NoSQL and semantic data and the principles of the Operational Data Hub to go beyond simply integrating data, and deliver unique value and experiences using all the data around the content:

■ The BBC platform enables them to link content of any type together based on semantic concepts with sports scores appearing in news articles, and shows metadata for multiple partners being generated from the same system that powers their internal delivery.

■ The NBC SNL project can use the connections between the individual clips to deliver what Mashable called "addictive" content that you "won't be able to put it down."

■ "The Suitcase" project demo was able to show all the data generated during production for any frame of the final cut of the movie and let users explore the relationships between that data over the course of the film.

■ And finally, Disney's Creative Genome Project is enabling multiple sources of data to come together, evaluate AI and machine learning and providing data to drive insight and analysis that helps inform the creative, marketing and engagement around its shows.

Outside of the entertainment industry, the Operational Data Hub pattern is enabling many types of organizations to use the Smart Content concept to improve their processes with better data. These include insurance companies linking together customer data, aircraft manufacturers linking test and manufacturing data to plane configurations, and energy companies better understanding all the data around their plants.

Conclusion

The concepts of smart content continue to provide the roadmap for getting the most value out of the data around the content. The new approach of using NoSQL and the rich linking of semantic data enable organizations to bring this data together and the Operational Data hub pattern puts this data into action with a repeatable pattern. Together these approaches are helping organizations realize the value of their data and make good on the promise of Smart Content. ■