# Leveraging Time-Based Metadata to Enhance Content Discovery and Viewing Experiences

Ben Weinberger
Digitalsmiths

*Abstract*

*Today's consumers have more choices than ever before for video entertainment and viewing devices. But with this explosion of choice has come complexity. Finding engaging entertainment has become a time-consuming and frustrating endeavor, resulting in decreased engagement and satisfaction.*

*The key to overcoming this discovery challenge lies in rich, time-based metadata. By creating time-based metadata at the production level and leveraging metadata-driven solutions to build best-in-class search and recommendation applications, stakeholders can create additional value at every stage of the video content lifecycle.*

*This paper discusses the superior metadata technologies and how they can be applied to solve today's toughest discovery challenges.*

## MAKING DATA RELEVANT

To enable state-of-the-art video search and recommendation tools, you need state-of-the-art data. And you need the ability to access, integrate and normalize data from disparate sources.

## CREATING THE DATA SET

From dialog to set design, anything about a scene can be tagged. Efficiently and accurately creating this rich time-based data requires advanced algorithms for facial recognition, scene classification, speech recognition, natural language processing, closed-caption time alignment and ad break detection.

To understand the granularity of the resulting metadata, it is worth drilling down to the details. Each video (an asset) may have metadata associated with it. This asset-level metadata can be human authored or directly imported from 3rd party sources. Each asset in turn consists of a series of contiguous scenes. Each scene is named and is time-bound. This is a human authored process.

Metadata is tracked throughout an asset. Individual metadata elements, such as an actor, location, rights, score, etc., are associated with a scene (container) and specific frames of video (location within an asset). A metadata track may be subdivided into subtracks. For example, an objects track could be defined for tracking specific branded elements in an asset, such as cars. The process to create metadata tracks and subtracks is both automated and human authored. For example, once an actor is tagged by name, facial recognition software can tag the appearance of the actor throughout the asset.

A segment refers to a time-bound portion of the asset containing a metadata element. The segment can also have metadata associated with it (referred to as segment attributes). All metadata is automatically tracked to a frame-level timestamp, providing the ability to display the exact frame in which the metadata track or element occurs. The depth of metadata that can be associated within a single frame is shown in Figure 1, a frame from *Spiderman 3*.

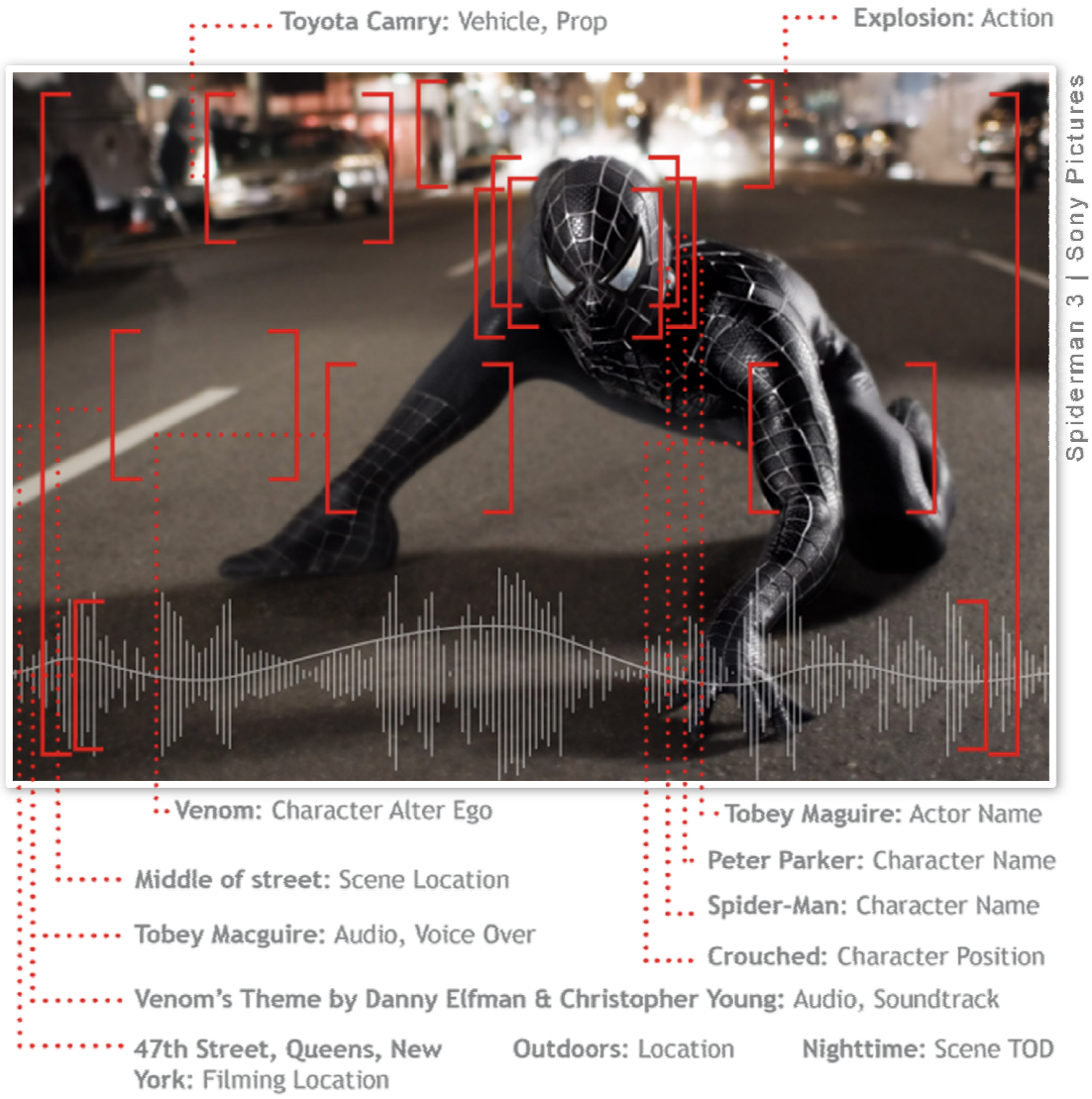


***Figure 1: All metadata created around individual frame***

Given the depth of information that can be created, it is preferable to tag the data at the time of production when much of the information is known by those closest to the creation of the asset. For example, in Figure 3, the filming location is tagged as 47th Street in Queens with a Toyota Camry in the shot. If the video were tagged by someone other than

the production unit, this level of information might be lost.

Information depth and granularity provides for much stronger ability to search and find the specific video segment you are searching for. While tagging at the production level is ideal, post-production tagging also yields deep, rich information that goes beyond the typical descriptions of title, major actors, and plot.

The ideal metadata solution should also support real-time tagging for live content. With the 92nd PGA Championship as an example, time-based metadata was married to the scoring feed to video-enable the Leader Board and the Scorecard. This enhanced viewing experience allowed fans to click directly on the golfer or hole to replay specific shots, increasing viewer engagement and creating new sponsorship and advertising opportunities.


***Figure 2: Metadata-driven live viewing experience***

The full potential of creating rich metadata sets is achieved by creating large libraries of tagged assets. This deeper level of intelligence opens the door to unparalleled search and recommendation functionality and accuracy.

For example, while it may be common knowledge that Tom Cruise danced in *Risky Business*, a metadata-driven search for "Tom Cruise dancing" will also deliver *Tropic Thunder*, a movie in which Tom Cruise briefly danced but is not even listed in the credits.

## ACCESSING THE DATA

Creating the dataset is the primary task; however, equally important is providing access to the dataset. As stated, the full power of a search and recommendation engine is found in the size of the libraries. But there

will be multiple libraries available as they are created by production teams, post-production studios, and third parties (well after post-production). The search engine must be able to interface with multiple libraries to maximize the value of rich metadata and to provide the ability to recommend videos across production houses, studios, and movie libraries.

Being able to create a unique look and feel for the user that meets the specific need of the licensee is equally important to the success of a video search engine. For example, the needs of the PGA PC application differs widely from the phone application (Figure 3) that leverages time-based metadata from *School of Rock*.

ENHANCING THE USER EXPERIENCE

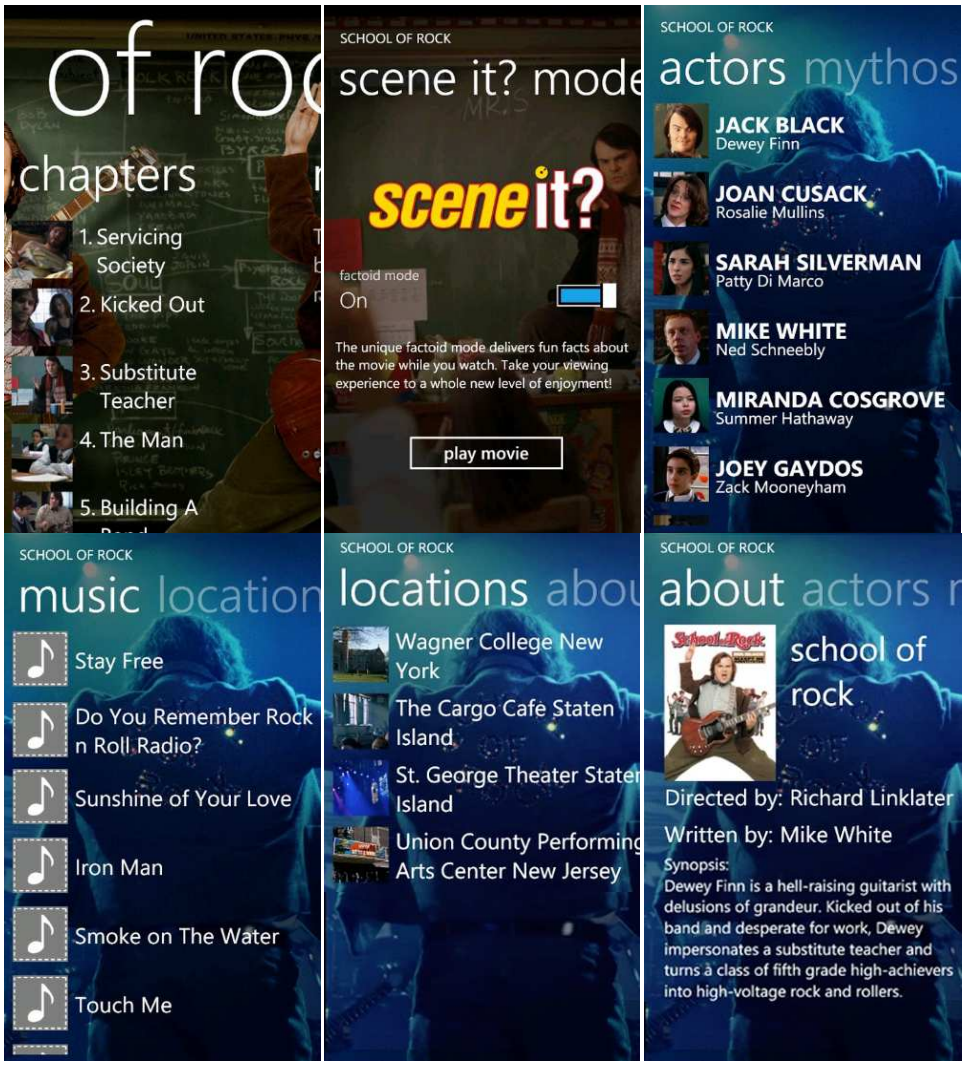


***Figure 3: Second screens***

With the growing popularity of connected devices, actionable, accurate metadata is needed to deliver the interactive applications that users have come to expect.

The same dataset could drive a connected television application, a set-top box application or a tablet application.

SUMMARY

Efficiently creating rich time-based metadata around each frame of a TV show, movie or live event requires advanced algorithms for facial recognition, scene classification, speech recognition, natural language processing, closed-caption time alignment and ad break detection.

The ideal discovery platform then integrates and normalizes scene-level metadata with rich 3rd party data from disparate sources to create a deeper level of intelligence around video content, enabling unparalleled accuracy and personalization in search results and recommendations.

With these time-based metadata solutions, stakeholders can develop best-in-class enhanced discovery and viewing experiences that drive engagement and better monetization of video assets.