

Enabling A Richer Video Experience With Metadata

A position paper for the [W3C Video on the Web Workshop](#)
12-13 December 2007, San Jose, California and Brussels, Belgium

[John Toebes](#), Chief Architect, [Cisco Media Solutions Group](#)

Introduction

In the Media Solutions Group, we see video as a necessary part of a true Web experience, but recognize that the Web allows for a much richer experience than just watching television. Furthermore, the proximity of a keyboard, mouse and other input device permits for a far richer interaction with the video content than is possible sitting back on a couch with a remote in hand.

To this end, we believe that enabling a true interactive experience with the end-user requires more descriptive data about the content they are watching, regardless of the format and location that they view the content. Lastly, the rise of a Web 2.0 technology sets the expectations for the end-user to be able to contribute, comment on and even correct the metadata associated with content.

Taxonomy of Useful Metadata

The need for metadata is well recognized with [W3C Metadata work](#) focusing on areas including [PICS](#) as well as dozens of standards in the industry for metadata associated with video content. In the realm specific to video, we can divide this metadata into four basic categories:

Content Attributes

This is basic descriptive information which describes the form of the video data. This includes things such as the length and format of the video as well as size and shape of the encoded content.

Content Description

This covers all information about the actual content. This includes areas such as Titles, listings of Actors, the Genre and Ratings for the content as well as many different forms of summary information for play lists and program guides.

Timed Metadata

This is information that applies to a subset of the content. At a basic level this is content such as Closed Captioning or even lyrics for music, but can

be much more sophisticated including listings of who or what appears in a scene or geospatial information such as where a scene was shot.

Hotspots

This is a more detailed level of information than [Timed Metadata](#) allowing identification of exactly which portion of a frame or sequence of frames an object, person or user interface element exists on.

The first two categories are well represented by many metadata standards including [Apple ITUNES XML](#), [Yahoo MediaRSS](#), and even the [Cablelabs VOD Metadata Content Specification 2.0](#). The [Mpeg-7](#) standard provides an XML form for the last two areas and even the [W3C SMIL](#) standard provides a framework for synchronizing information with presentation of content. There are also several next generation video on the web companies who are experimenting with video hotspot technology. While these standards serve different purposes, none has risen to the top as a common mechanism for holding the interactive information about video content on the web.

There is much growth on the web in [Timed Metadata](#) usage and implementation, from Adobe's captioning support in Flash CS3 to web-based scene detection (and, thus, chapter-ization), deep-links in [Google Video](#), and many others. Increasingly end users are allowed to contribute their own meta-information, in the form of timed comments or tags. However, as with the above, with the exception of captions, no standards exist for this burgeoning source of video metadata.

Source of Metadata

While the standard of representing the metadata is important to resolve, it is just as important to recognize where the metadata is being generated.

Content Creator

During the creation of the content a huge amount of valuable data is actually created, but rarely captured in a usable form. This includes all the work that has gone into clearing the rights for objects, images and people in scenes as well as knowing locations where content was created. Digital cameras today record much metadata about each photograph taken, including GPS location information with some newer cameras. This type of data is rarely made available for web video.

Publisher

When content is initially syndicated or released, someone has to go through the process of creating (often with a web form) at least the [Content Attributes](#) and [Content Description](#) for the content.

Professional Content

There are several service organizations which can go in after the fact and either by hand or algorithmically generate rich metadata. The best

example here are companies that do close captioning, but scene and image detection software for video is also introducing additional metadata.

Users

As we can see with the rise of many Web 2.0 environments where users are contributing volumes of detailed information (e.g. [Wikipedia](#)), dedicated fans are a major source of metadata for content. When multiple sources for this metadata is aggregated with peer-review, the quantity and quality of this data can easily exceed the other sources.

It is this last category which is key for deriving the true value of a Web based video experience. With the ability for users to contribute their knowledge to the content the engagement is higher and the experience becomes much more personalized.

Value from Metadata

Having rich metadata for video content provides many opportunities on which the web and other video applications can capitalize. The first obvious uses, driving many web companies today, are program guides complete with detailed program descriptions as well as mechanisms to search content in-depth -- finding not just the top level information such as one finds on [IMDB](#) , but even dialog and scenes contained within the video.

More sophisticated use of this metadata allows for better recommendations of content, and dynamic advertising where appropriate ads are displayed based on actual content of the video itself, or where merchandising opportunities are based on what has appeared in a video that a user has watched.

Impediments

Unfortunately there are several impediments to being able to gather and fully utilize metadata for a next generation Web application

1. The same video is encoded in multiple formats for delivery in different circumstances, to different devices, and for different data rates. While the [Content Attributes](#) might differ, the [Content Description](#) and [Timed Metadata](#) are shared across all forms of the same content.
2. Web applications which want to deliver video spend their time focusing on the players to deliver the video and the integration into the web browser - something which provides low value to them and less time is left to devote to the actual value producing part utilizing the metadata.
3. Each web site seems to create their own customized video player with its own controls and expecting the user to download the appropriate player or framework technology just to play the video. To avoid this download burden on end-users, many websites end up creating their players in [Adobe Flash](#) or limit playback of video to a particular browser on a particular operating system.
4. Existing metadata standards are focused around a single source of metadata

and do not take into account the multiple sources including users who create, enhance and correct metadata for a video.

5. There are several metadata formats in use, and absolutely no consistency in their application. Industry support for standards alignment, adoption and extension would positively impact the overall health of the content management and digital distribution industry.

Conclusion

This is not a new area for the W3C to focus on with the standardization of metadata around images covering hotspots, alternate text and layout information embedded in the XHTML for a web page.

By adopting the same standard for separating the video from the metadata in elevating video to a first class object and providing the primitives for applications to truly interact with the video the same way that the W3C has done for static images, we believe that the Web community will be able to focus on growing the real value of video in their applications.