



THE CURRENCY OF INTERNET VIDEO

*THE IMPORTANCE OF QUALITY METADATA IN
MONETIZING INTERNET VIDEO*

OCTOBER 2008

Gotuit Media Corp.
15 Constitution Way
Woburn, MA 01801
P: 781.970.5500
F: 781.970.5502
www.gotuit.com

Table of Contents

Executive Summary	3
Introduction: The Internet is about Data, People – and Metadata	4
All Metadata is not Created Equal	5
Metadata is the Currency of Internet Video	6
Applications of Video Metadata	7
Increasing Advertising Options for Advertisers and Publishers.....	7
Increasing Programming Options for Publishers and Syndicators	7
Increasing New Consumer Video Experiences.....	8
The Science – and Art – of Metadata Creation	9
Example – Manual and Automated Metadata.....	9
Video is More than the Sum of its Parts	11
Metadata has the Lowest Production Cost of all Video Attributes	12
Conclusions	13
About Gotuit.....	13

Executive Summary

Internet video is all the rage among consumers. Publishers, broadcasters and advertisers are all eager to catch this wave. Internet video is projected to be the majority of consumer Internet traffic in the coming years, and the Internet a significant distribution medium for video. Paradoxically, publishers are facing challenges in monetizing Internet video despite consumer demand.

The Internet differs from traditional means of distribution. Much of the value propositions of the Internet as a distribution medium have not been recognized and utilized towards the strategic goals of video publishers.

Metadata is the linchpin to unlocking this value. As the title of this paper states, metadata is the 'currency' of Internet video. With quality metadata, publishers can create video experiences integral to Internet audiences and new monetization schemes around these experiences, including advertising.

Metadata enables the following and more:

- Search – both at a file and scene level - that is a key function of the Internet
- Multiple navigation paths within or across different videos in a manner that users are accustomed to in Internet navigation
- Clip and playlist creation for exploiting viral sharing and social networking trends
- Dynamic and targeted programming to create higher user engagement
- Precise targeted advertising tailored to user behavior
- Accurate reporting and analytics critical to advertising and monetization

Publishers, therefore, need to consider metadata as the third key element of video production, in addition to video and audio. While metadata is critical to the success of Internet video strategies, costs associated with authoring metadata are insignificant to the overall costs of video production.

Metadata quality must be assured for publishers to deploy successful Internet video strategies. Quality metadata is human authored as opposed to automated. While automated schemes are neither sufficiently accurate nor reliable, they also do not allow the programming choices possible with human authored metadata. Moreover, human authored metadata is more efficient to create. At the same time, such metadata can be added to video even after video has been published, creating new use cases and programming options.

Such quality metadata cannot be an afterthought, or worse, overlooked. Publishers need to recognize that successful Internet video strategies may well rest on suitably authored metadata and metadata management systems.

Introduction: The Internet is about Data, People – and Metadata

Metadata is data about data – it is the connective tissue between people and the large amount of information on the Internet. Metadata provides usable context to data for users. For the foreseeable future, metadata will be used to harness the usefulness of the Internet as machines cannot bridge the semantic-gap between people and machines, nor understand the context of information that humans generally take for granted and understand intuitively.

As we enter the era where video is gaining importance on the Internet, the role of metadata becomes vital. Video on the Internet not only adds to the semantic gap between humans and machines, but also creates a visual gap that machines inherently cannot understand.

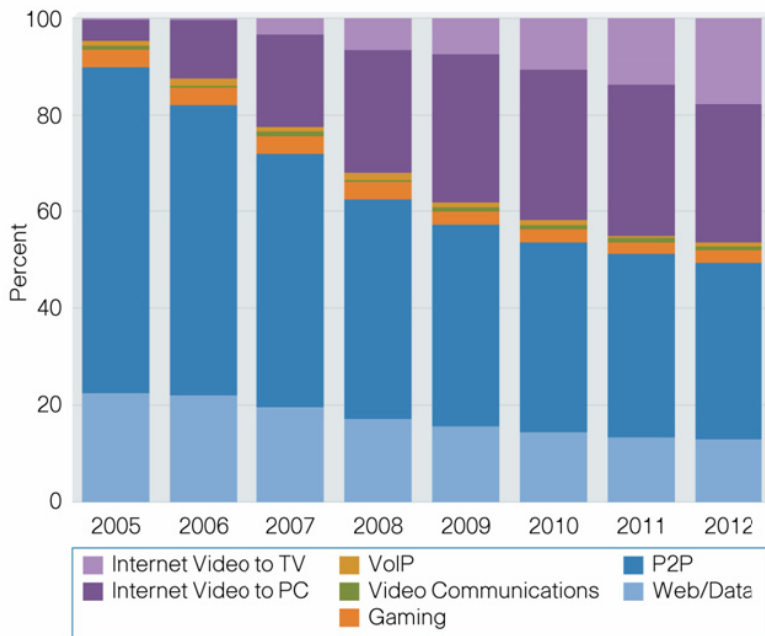


Figure 1

Internet video is forecast to comprise 90% of consumer Internet traffic by 2012ⁱ (Figure 1). Today, Internet video delivers 70% of the impressions of televisionⁱⁱ. At the same time, Internet video advertising in 2007 was around \$400 million compared to \$70 billion that television advertising generatedⁱⁱⁱ.

The industry is abuzz whether the Internet will create monetization of video the way traditional broadcast media has. Publishing to broadband with associated advertising like television is not driving desired returns for video.

This paper discusses the role of human-authored, quality metadata as an overlooked imperative of Internet video programming towards this objective.

Metadata is the connective tissue between people and the large amounts of information on the Internet

“Rich media is opaque to machine search and indexing.”

Mark Randall
Adobe Chief Strategist on Beet.tv

“Our challenge with all these (Internet video) ventures is to effectively monetize them, so that we do not end up trading analog dollars for digital pennies. This is the No. 1 challenge for everyone in this industry today.”

Jeffrey Zucker
President & CEO, NBCU

All Metadata is not Created Equal

Metadata itself is not a new concept. It is important, however, to recognize that use of metadata during the production process is very different from use of metadata as the ‘currency for video’ – a new construct that deserves attention. This construct is explored in the next section.

Overall, in the course of video production, a large amount of metadata is created for administrative, compliance, and operational purposes. Most of this metadata does not make it to the final publishing process owing to the various transcoding processes in the production workflow. Even if it did, metadata to build engaging consumer experiences and optimal monetization of video must be authored with this distinct objective.

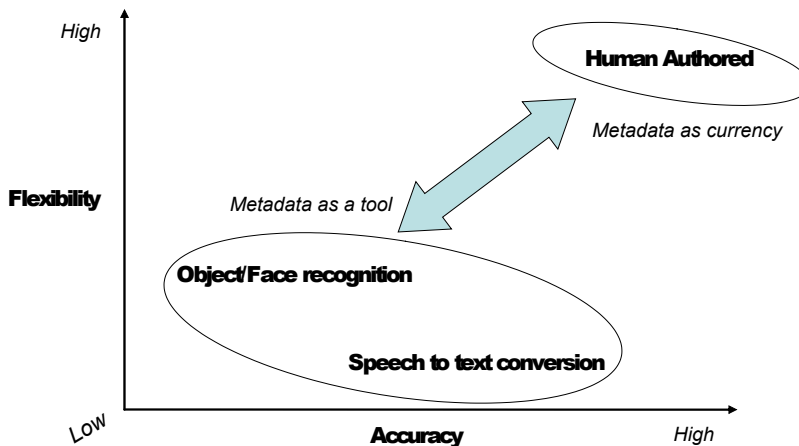


Figure 2

Such metadata needs a high level of accuracy and flexibility to serve its purpose (Figure 2). Amongst the growing need for metadata, the industry is also hoping for a silver bullet to address metadata creation through machine functions such as speech-to-text, facial, and object recognition. As we shall see, such technologies are not sufficiently developed to achieve the quality and accuracy required to engage and monetize audiences that human authored metadata can. Nor are such technologies, even when perfected, suited to entirely meet the needs of consumers, given the man-machine semantic gap discussed earlier.

For publishers, the following considerations are central to this discussion:

- 1) The limited accuracy of automated metadata can drive basic video search, but cannot be used for the purpose of improving the viewing or advertising experience.
- 2) Machine functions require significant training by a person, and therefore are not truly automated.

“The quality of a video’s metadata is as important as its video or audio quality and publishers should spend as much time improving the metadata as they spend on the video and audio”.

*Bill Hensler
Vice President & General
Manager, Dynamic Media
Engineering, Creatives
Business Unit at Adobe,
speaking at the Akamai
Customer Summit in Boston
in September of 2008.*

- 3) Machine functions are limited to the existing data set, and cannot add new information about the video like a person can.

Metadata is the Currency of Internet Video

Metadata based applications allow the inherent value of video to be unlocked and monetized on the Internet. These applications range from elementary search and discovery, to advanced use cases such as, scene level search, targeting, social networking, advertising and interactivity. (Figure 3)

Metadata enabled features
Video search (asset level)
Video search (scene level)
Seek and skip functions
Video packaging and presentation
Playlisting
Dynamic program updates
Multiple navigation paths within or across videos
Mashups/Remixes
Advertising (in-stream, overlay, banner)
Personalization and targeting
Sharing and social networking
Reporting and analytics
Recommendations

Figure 3

Video is intended to deliver an experience and engage us more than any other medium. People relate to, experience, and consume video on many levels. People develop affinities for characters, storylines, events, special-effects, and many other aspects of information captured on video. Bringing the vernacular of Internet experiences to video applications is extremely challenging in the absence of quality metadata.

Let us elaborate on some of the features listed in Figure 3 for the purpose of illustration. More in-depth implementations of such features are described later in the paper:

- **Advertising:** Metadata defines instream ad insertion points with accuracy. This can be used to create flexible ad logic, targeted advertising, and new forms of advertising supported programming such as playlists, mashups, and viral sharing consistent with Internet user behavior.
- **Search:** Scene metadata can be used to drive better and more granular search results. Metadata can capture intent of the video in addition to the content of the video to aid more sophisticated search. For example, a serene seascape has a very different intent with an accompanying music track from the film Jaws as opposed to a Largo from Vivaldi's 'The Four Seasons'.
- **Navigation:** Video programming, whether news or primetime sitcoms, is being produced in shorter, more discrete segments than before. Metadata allows scene and segment level search, virtual clips, sharing and playlisting without compromising the integrity of the original video asset. With metadata, advertising can be associated with such clips and playlists, in addition to the full length program.

We live in a time when metadata may actually be more important than data.

*Shelly Palmer
'Television Disrupted', 2006*

For video publishers and broadcasters, the absence of rendering video without the level of relational information – i.e., metadata – commensurate with the diverse associations people have with video and use cases on the Internet means that consumption, and therefore monetization, of video is compromised.*

Applications of Video Metadata

It is also important to understand that metadata is not a single attribute of video, but rather serves multiple purposes. Among these, metadata creates new avenues for creative expression by video publishers and new models of advertising. The uses of metadata are varied such that metadata creation and applications can continue to evolve much after a video title has been published.

Let us consider a few examples of video programming that has been tailored for Internet audiences using metadata.

Increasing Advertising Options for Advertisers and Publishers

- **Dynamic Ad Insertion and Flexible Ad Logic:** Implementations of metadata for dynamic advertising and flexible advertising logic enable broadband video publishers to enhance how they monetize their video libraries by authoring structured metadata describing each meaningful scene within the original source videos. This metadata defines the optimal in-stream video ad insertion points, allowing publishers greater control and flexibility with their advertising strategies. In addition, the ads served in the precise insertion points can be targeted by third-party ad providers based on the scene metadata such as Character Name, Player Name, Topic, Keyword, etc. Banner or overlay ads can also be targeted based on the rich metadata. As a result of this greater ad logic flexibility, rather than just pre- and post-rolls for each asset, publishers can set their ad logic to utilize the mid-roll insertion points. The publisher sets the ad timer, and the ad will play at the next available insertion point, no matter what asset or scene is being viewed, after the timer has expired. The result is that the viewer has the freedom to sample more assets and navigate directly to the most interesting scenes, while the publisher is able to monetize that experience in the most effective way possible. This capability has been applied to Flash and Move Networks' video formats.

Increasing Programming Options for Publishers and Syndicators

- **Search, Clips and Playlists:** Extreme Outdoor Network (www.xontv.tv) specializing in outdoor recreational activities such as hunting and fishing realized that regular thirty to sixty minute programming was not suitable for its Internet video audiences. In order to maximize the impact of their online video programming, XON choose to author metadata for its videos such that

* As an example, the abundance of user-created short clips on YouTube of prime-time shows, news, and talk-shows represents what consumers want and are willing to do in the absence of programmers offering this to them. A search for John Stewart on YouTube, for example, brings up more than 2400 clips at the time of this writing, most of them from The Daily Show. These are missed opportunities for programmers to serve audiences of this new medium and monetize their content.

- enthusiasts of individual sports could find video segments tailored to their specific interest within their genre of interest. Rather than create a large number of individual clips through editing the videos, XON applied metadata to their original thirty to sixty minute programs such that the original video assets can be rendered as virtual clips that can be searched, organized and programmed to meet individual user's needs.
- **Chapterization, Skip and Search:** When Fox Reality decided to broadcast their Fox Reality Really Awards show on the web, they indexed the entire award show in segments so that users could watch sections that were of interest to them. It was unlikely that most users would watch the entire program on the web. However, through metadata indexing, users can skip to sections based by award, show, presenter, musician and so on, as well as create playlists and watch them in linear fashion, thereby creating their own highlights of the awards ceremony.
 - **Dynamic Programming and Multiple Navigational Paths:** Sports Illustrated uses metadata to create dynamic programming for College Football fans through its FilmRooms™ video portals. Users can navigate through multiple paths to view highlights, which are updated as the games progress and rankings get updated. Users can search by team, player, position and other ways that give them easy access to create their own highlight clips that can be shared with others and posted on user websites and blogs.

Increasing New Consumer Video Experiences

- **Mashups, Personalization and Sharing:** Lifetime uses metadata to allow users to define their own virtual scenes within a video program. These scenes can be shared with others, and concatenated to create user-defined playlists (mashups). Users can also add additional metadata to make the scenes more meaningful and manageable. The scenes are organized by metadata and the underlying video assets are not changed, eliminating additional storage and management costs.
- **Reorganize and Collaboration:** Similarly, Carleton University uses video on demand to create lectures that students can tailor to their needs through indexing parts of video lectures and reorganizing them to their individual requirements. At the same time student notes and annotations make the videos searchable by other students.
- **Customize and Self-programming:** Sprint has used metadata to create a new application for fantasy football with the National Football League. NFL Fantasy Video is the first mobile application that allowed users to watch their own custom video highlight reels of just the players they want. Each week during the season, every play of every game in the NFL is indexed using

“Lifetime Movie Mash-ups fundamentally changes how we can engage our broadband video viewer and promote our Lifetime movies.”

*Kimberly Dobson
VP Business Development,
Digital Media, Lifetime
Networks.*

“VideoNotes has the potential to really add value to our existing services. Students have the ability to highlight their course lecture videos for review the same way they would use a highlighter to mark portions of their textbooks and written notes.”

*Jeff Cohen
Manager, CUTV, Carleton
University.*

metadata. Sprint customers can then set up their fantasy team, or favorite players, including a QB, 2 RBs, 2 WRs, TE, K, and a team defense. Once set up, users can see the video highlights of just their players, and even jump to a specific play. In addition, users can scout any other NFL player's video highlights and choose to add them to their team.

The Science – and Art – of Metadata Creation

Given the importance of metadata to Internet video, thoughtful consideration needs to be given to what constitutes good metadata. While enabling video search and interactivity required for quality Internet experiences, metadata is needed not only to maintain the original intent of the video, but also enhance the experience in ways not possible with traditional video distribution. Beyond that, metadata can enhance the Internet video experience in ways yet to be conceived.

Metadata authoring is a topic unto itself and the subject of many technical papers and industry initiatives, which are beyond the scope of this paper. In order to convey the essence of what constitutes good metadata, we'll rely on a simple illustrative example that addresses a common consideration in authoring metadata – should metadata be authored by people, as is most of the video production process, or should it be automated.

Example – Manual and Automated Metadata

Let's consider searching for a cameo appearance by Brad Pitt in an episode of the NBC sitcom, 'Friends'. Assuming a number of Friends fans have not already spent their valuable time to do this and post it on YouTube, this requires a few things. It requires metadata for the episode in which Brad Pitt appears. Given his celebrity, it is most likely that Brad Pitt is listed in the Content Description metadata that was created during production. This data was entered by someone on the production team.

Thereafter, it is possible that a user searching for this episode may watch the entire episode, and the metadata has done its job. More likely, the user may want to watch only the scenes where Brad Pitt appears in the episode. Since the final packaged video does not have the original time code the editors used to edit the video, this information is lost and must be recreated.

In order to search for Brad Pitt within the episode, advanced facial recognition software may be deployed that is trained to recognize Brad Pitt. Assuming it can do the job, it will identify the first frame and subsequent frames that Brad Pitt appears in. Scene change detection software may then be deployed to detect a scene change before the first Brad Pitt frame and mark that as the start of the clip. It may detect the next scene change to mark the end of the clip.

Theoretically, this seems to do the job – provided the technologies work reliably. The most well developed of such technologies – speech to text – works less than 100% reliably (generally considered to be 95% in the best case, but reportedly at 50% on broader scale^{iv}), so the first concern would be whether the technology worked in identifying Brad Pitt. Since he is a well known face, let's assume the system can be trained rigorously in this exemplary case, but it's still a less than perfect chance. Moreover,

training systems to perform voice, face and object recognition is time consuming, requiring tremendous upfront investment of time and resources. The second concern is whether the resulting clip or clips were watchable from a cinematic experience standpoint:

- 1) Are the scene changes correct, in addition to accurate?
- 2) Did the scene boundaries interrupt key dialog?
- 3) Do we know the context within which Brad Pitt is introduced into the show?

These are just some of the considerations. Conceivably, a better place to start the clip was the prior scene, or maybe further into the scene. A person can make this decision very quickly and intuitively, whereas automation can lead to not only a suboptimal result, but it may also be grossly inaccurate. Finally, a person would need to review and potentially edit the work of a machine.

To make a finer point of automated versus manual metadata creation, consider the following:

- 1) Were we trying to locate Ted Dansen in his Hellboy outfit, or Danny DeVito in his Penguin outfit, chances are facial recognition would be hopelessly lost, as even humans cannot sometimes recognize the faces behind the outfits. Nevertheless, a human is better suited to this task.
- 2) More dramatic contrasts between manual and automated metadata can be demonstrated in sports programming. Sports viewing is a combination of close ups, long camera angles, fast motion and fast camera transitions. The combination of this along with the fact that players are not always facing the camera makes it impossible to apply facial recognition technologies to create clips automatically. Creating clips of Lebron James' three-pointers or Tom Brady's touchdown passes can only be done by a person.

Metadata enabled features	Manually authored metadata	Automated metadata
Cue points	Infinite	Limited
Scene level search	Accurate	Approximate
Dynamic programming	Yes	Limited
Chapterization	Yes	Limited
Multiple navigation paths within or across videos	Yes	Limited
Ad insertion points	Infinite	Limited
Seek points	Accurate	Approximate

Figure 4

In any event, given the less than 100% accuracy of any automated systems, be they speech recognition, image or facial recognition, scene change detection and such, quality end results are derived through human authoring while using automation to facilitate the process (Figure 4). A second important consideration in authoring meta-data for Internet video discussed next is however impossible to automate.

Video is More than the Sum of its Parts

Beyond the obvious scene, object, face and speech recognition whether done automatically or manually, video is a complex communication medium. The creative combination of visuals, sounds, speech, emotions and storytelling inherent in any video makes it so. Inferring the intrinsic appeal of a video program on the Internet for different users can only be done by people.

In the earlier 'serene seascape' example, imagine that the music is from Jaws, but the video has a comic audio commentary lampooning the (irrational) fear of sharks. The emotion associated with the video is humor, as opposed to fear. The commentary could be educational about sharks, the intent being to inform as opposed to thrill. People can immediately establish such intent and capture it in metadata for their audiences.

Among the successful implementations of metadata listed earlier, alternate navigation schemes – including navigation across different video files – is one where human imagination can be applied to create new, lasting user experiences that are not possible with automated metadata schemes.

Consider multi-threaded programs such as, ABC's Lost, or reality shows with many participants and events such as, Fox's American Idol and CBS' Survivor, or sporting events – wherein users can aspire to recapture the experience of the original program in many different ways. Consider the following examples:

Lost:

- Sawyer + Kate + Romantic Scenes → creates a playlist across all episodes of scenes where Sawyer and Kate are together in a romantic setting

American Idol:

- Seasons 1-9 + Winners + Finals → creates a playlist of all the American Idol winners' final performance on the show

Sports:

- Tom Brady + Touchdown passes → creates a playlist of all of Tom Brady's touchdown passes in NFL games.

While the above examples are hypothetical, metadata easily allows users to essentially apply 'Boolean logic' (similar to what users do in web searches) to generate attachment through new experiences. In the absence of such metadata, programmers would need to actually edit and re-encode individual clips, which is a formidable task, if not an impossible one. It is also impossible to successfully create such dynamic playlists and alternate navigation schemes using individually encoded clips.

Human imagination remains ahead of technology. Making metadata choices by what automated technologies allow is inherently more limiting than generating metadata manually, wherein video can be tagged in many different ways, and metadata fields can be created and managed any way that a human operator conceives necessary, intuitive, probable, or even imaginable.

Metadata has the Lowest Production Cost of all Video Attributes

One of the underlying questions is the cost of authoring metadata and whether one approach is more cost effective than another. This boils down to the question of quality versus quantity. If accuracy and end-user experience is secondary to processing large volume of video for a basic search index, then automation is likely to help solve the problem better than a human. Automation, such as scene-change and speech-to-text serve well in the production stage of video. This is because there is a lot of raw footage and people handling the video are professionals. Their task is to manage the video production, not to consume or monetize the video. In the case of researchers looking to sift through large video libraries, the same argument applies – the video experience is secondary to the objective of locating a video or a clip within a video asset.

At the risk of being redundant, let's (re)visit some of the commercial applications of video:

- Search at a file or scene level
- Create, display and share virtual clips and playlists
- Create advertising insertion points and advertising logic
- Generate detailed usage tracking and reporting data

Automating metadata creation for each of these exemplary applications will require mostly disparate processes, in contrast to human authoring which allows all required metadata to be created in a single pass.

The cost of human authored metadata is, therefore, not only lower than automated metadata, but it is also insignificant relative to the overall video production costs. Human metadata authoring can typically be accomplished in much less time than the play-out duration of the video. People don't have to be trained to recognize speech or images like machines do, reducing upfront investment of time and resources. Lastly, human authored metadata allows for further human creativity and reasoning to be applied to video programming, bringing new elements of creativity to an already creative process with negligible incremental costs.

Conclusions

Metadata is a critical element to the success of video on the Internet. Publishers need to address metadata creation as an essential part of the video production workflow.

- Video as a complex medium requires human authored metadata to bring the vernacular of Internet experiences to video on the Internet.
- Quality metadata to create audience engagement and monetization should be authored with distinct objectives of creating such Internet experiences for video.
- Such metadata is best authored by people using authoring systems that allow 1) flexible and accurate metadata to be applied to video assets, and 2) additional creative expression to be brought to the medium of Internet video.

Publishers need to incorporate systems that author and manage metadata towards these objectives as they look to build audiences and advertising with their Internet video strategies.

The implementation examples described earlier in this paper are based on Gotuit's video metadata authoring and management system. These represent among the most advanced uses of metadata and Internet video implementations. The metadata in each case was human-authored either by Gotuit or its customer.

To learn more about how Gotuit can help implement solutions to create greater use and monetization of your video programming over the Internet, visit our website at www.gotuit.com, or contact our sales team at: 781.970.5414.

About Gotuit

Gotuit is the leading provider of premium metadata technology which optimizes the value of video libraries for professional content publishers. The Company's patented video metadata management system (VMMS) is an end-to-end system that unleashes the power of the metadata-defined scene to greatly enhance the presentation, discovery, advertising, and profitability of video libraries. Gotuit powers video for leading brands such as Lifetime, Fox, Sports Illustrated, Major League Soccer and more.

Sources

- i Approaching the Zettabyte Era, Cisco Systems, Inc., 2008
- ii Source: Display Impressions: comScore Ad Metrix, Jan -08, Television GRP's: comScore Estimates
- iii Source: IAB
- iv <http://www.techcrunch.com/2008/08/26/the-wb-rises-from-the-ashes-as-a-hulu-competitor/>