

Customized Video Playback: Standards for Content Modeling and Personalization

Michael Bush, Brigham Young University; Alan Melby, Brigham Young University; Thor Anderson, IMS Global Learning Consortium; Jeremy Browne, Brigham Young University; Merrill Hansen, ViewerFreedom.org

Introduction to Customized Video Playback (CVP)

Early in the cellular telephone revolution, the call from the promoters of that industry went something like “anything, anywhere, anytime.” Nicolas Negroponte, director of MIT’s Media Lab, later coined what can be considered to be an important slogan of the personalized content revolution, “nothing, nowhere, never unless it is timely, important, amusing, relevant, or capable of engaging my imagination” (Negroponte, 1994). Indeed, with hundreds of television channels available via satellite and cable, pre-recorded programs from VHS tapes and DVDs, as well as video files available on the Web, the variety of content available to us is seemingly endless. Not only do consumers have access to the arrival of an incredible variety of programs via these various means, but they also can time-shift those programs via VCRs and now digital personal video recorders (PVRs) such as TiVo and ReplayTV.

These developments in video technologies help to tailor video viewing to suit consumer tastes, expanding our options concerning *which* video programs we watch and *when* we watch them. Unfortunately, our ability to select *how* we watch these programs is mostly limited to our manual dexterity with the couch potato's best friend, the TV, VCR, DVD, or PVR remote control. Even playback software on most computers leaves much to be desired. Once a video program is selected from all those available, the viewer can only jump around from place to place within the chosen documentary, feature film, or perhaps recorded lecture. With respect to selectivity, however, the crudeness of this approach is apparent. Whether the viewer is an ordinary consumer, a student, or a professor as shown in our example below, the needed function is the same: To be able to select or even avoid segments as desired.

To enable two levels of selection, (1) not only selecting a desired program, but also (2) selecting the portions of the program to be viewed, creates the possibility of true content personalization, which we call Customized Video Playback (CVP). Once a specific program has been chosen, CVP enables the viewer (or presenter in the case of an instructor) to select only those portions of video that correspond to predetermined objectives or choices. The video program itself need not change, only the manner of viewing it.

In using software to accomplish this selection, an English literature professor might choose and show several scenes from a certain rendition of Hamlet to highlight Laertes' development as a character. Afterwards, to support student review in a computer lab, the professor could then make available the same presentation that had been made in the classroom. Using similar software, a foreign language educator might well wish to play a few lines of a conversation over and over again to focus on one communicative function, while a professor of film may wish to show only segments of a film that use certain cinematographic techniques.

Consumers could choose to use similar functionality to avoid content that they deem undesirable or to choose favorite segments of a film that they wish to view. A Kung-Fu fan could choose to view only the fight scenes from her preferred martial arts film, while a softhearted romantic could view the touching scenes from his favorite romantic comedy, or a watchful parent or discretionary viewer could avoid viewing specific scenes of offensive material during an otherwise appealing movie. In a similar vein, such an approach would provide a significant tool to classroom teachers who must concern themselves with the appropriateness of the segments they choose for in-class viewing (Gareis, 1997). In other applications going back to the early 1980s, feature movies are annotated in such a fashion so as to open them for use for students to learn language and culture (Gale, 1989).

In each of these cases, with CVP, viewers are seeing what they wish to see, when they want to see it. They are able to avoid watching a video program only sequentially from beginning to end, benefiting from a level of control that better meets their needs. The increased control enables them to tailor the playback of the video program for a particular group in the case of the English professor above, or even personalize the playback in the case of the student or the consumer.

CVP: Fine Tuning User Control and Personalization

There are at least two ways to achieve repeatable, customized viewing of a video program. One method, a file-based approach, is to encode the program digitally, place it on a hard disk, and use video editing software to create a new video program. A second method, a selective playback approach, amounts to making selections from a video stream based on a content model of the video, where the selected segments are not separate files but rather are “virtual” clips from a single video source such as a single title on a DVD.

The file-based approach has four major drawbacks:

- It is time-consuming,
- It requires specialized skills,
- It requires specialized, expensive, software and hardware tools, and
- It can violate copyright law.

In comparison, the selective playback approach, such as is implemented with CVP, involves the creation of a content model of the video program of interest. Based on viewer choices, playback software accesses that content model, and creates customized playback on-the-fly. To do this, CVP, as we envision it, depends on three coordinated technical components:

- The video program (henceforth called a video asset),
- A description of the content of the video asset (i.e., a content model), and
- A selective playback mechanism.

The video asset is any video stream that is accessible via time code. The Video Asset Description (VAD) is represented in some format — preferably based on a suitable standard

content description system — and segments the video asset using time codes rather than through any copying or modification of the original video material. In addition, end-user tools facilitate the creation of playback specifications from video asset descriptions. The resulting combination of software, data structures, data, and digital video stream amounts to an efficient approach to a content retrieval and processing system.

The advantage of using a standard content description system is that a significant portion of the development and delivery software would be reusable in the three different settings described above, from the classroom, to a student's workspace, to the home. The description could also be used with software packages from different vendors. In order for the system to work, someone must analyze the content of a video program and publish the analysis as a VAD. An end user interested in that video program acquires the copy of the video asset and its VAD and uses a VAD-aware tool to create a playback specification according to personal objectives. The video program is then viewed under control of that playback specification. Different end-users could potentially use the same video asset and the same VAD to produce a multitude of personalized views of the same video program, without needing to understand the inner-workings of the system and without actually copying or modifying the video program.

It is important at this point to differentiate between two approaches to establishing the playback specification for customizing the viewing experience. Our VAD-based approach, combined with suitable tools, allows end users to create their own playback specifications. In this case viewers examine the video asset description and then construct their personalized playback specification. Another approach is for viewers to purchase a pre-defined playback specification that has been created for a particular purpose, to select or avoid particular content. Such solutions employ a playback specification that is available to the public, rather than a publicly available video descriptor from which private, personalized video playback specifications are created. This distinction may have important intellectual property rights ramifications with respect to commercial films.

Technical Description of CVP

Although various CVP applications could each be developed independently from scratch, the purpose of this article is to argue for a standards based approach to CVP. We will present one standard for the description of a video asset and describe several existing or evolving mechanisms for the customized playback of video assets.

Every CVP application is based on one or more video assets that could exist in one of several forms. For example, a video asset could be an MPEG-2 file on a local hard disk, an MPEG-4 file on a Web server, or a commercial DVD, just to name a few possibilities. Part of the development process for a CVP application necessarily includes an analysis of the content of the relevant video asset and the identification of segments that can be played back for a particular purpose. We strongly feel that CVP developers should agree on a common Video Asset Description (VAD) format for representing the segmentation of a video asset and various types of descriptive information about the video asset as a whole, as well as about its various segments. One advantage of such a format would be to allow various developers to share the use of the same tools not only for the creation of VADs but also for their use. The focus of this

article is a description of the VAD format that we have begun to define and the methodology we are using.

Essentially, the proposed VAD format is an XML application conformant with the IEEE Learning Object Metadata (IEEE LOM) specification (IEEE LTSC, 2002; Duval. et al. 2002) and then extended using elements from MPEG-7 (Manjunath, Salembier, and Sikora; 2002). Once a VAD has been created for a particular video asset, it can be used as the basis for a number of CVP applications, where each application must include some selective playback mechanism.

As mentioned earlier, until now the most common mechanism for selective playback of a video asset has been a remote control. Unfortunately, this mechanism lacks a fundamental feature of CVP as we envision it: repeatability. There must be some way to recreate a customized visual experience without constantly having a thumb poised over the buttons of a remote control and an eye focused on a timecode display. In a way, the various mechanisms described below can be thought of as a fancy remote control with a memory of which buttons to "push" and when.

Suppose your video asset is a file on a Web server and suppose the client machine has a SMIL-compatible player installed. Then your CVP application could be a SMIL file (<http://www.w3.org/AudioVideo/>; (Kennedy & Slowinski, 2002). A SMIL tag for playing a 15-second clip from a video asset, starting ten seconds into the asset, might look like this:

```
<video src="toad-movie.rm" clipBegin="10s" clipEnd="25s"/>
```

Using SMIL is only one possible mechanism for achieving customized playback. Another mechanism is a DVD playback object that can be invoked using ECMAScript embedded in an HTML page. Some CVP applications use the Microsoft WebDVD playback object, MSWebDVD, which can be invoked from Internet Explorer using ECMAScript (Microsoft, 2002). Other similar DVD playback objects are available such as the one from VisibleLight (<http://www.visiblelight.com/>).

A third possible mechanism for selective playback besides SMIL and HTML is to use a DVD player (software and hardware) that has the ability to play a sequence of video files specified by the user. This is sometimes called a playlist; however, we are using the term playlist in a different sense. Some commercial software DVD players already include this ability. Within the context of CVP, we use "playlist" to refer to a sequence of segments and attributes within a single video asset. This ability to play a sequence of clips from within a single video asset is especially important when copyright restrictions do not allow portions of the video asset to be copied into separate files. It is also important when the video asset is segmented hierarchically, making the storing of all possible segments, some of which contain other segments, rather inefficient. Various playlist-aware DVD players are under development, and the DVD Association (<http://www.dvda.org/>) has formed a working group to define a standard format for video-clip playlists that define customized playback of a sequence of clips from within a DVD. Given a playlist-aware DVD player and a VAD, a CVP application might consist simply of a player running a particular playlist.

The DVD Forum (<http://www.dvdforum.org/forum.shtml>) is working on a specification for advanced consumer electronics DVD players called ENAV (Enhanced NAVigation). At this point it is unclear exactly what features will be included in ENAV and whether it will be based on an HTML+ECMAScript or a SMIL model, but it will hopefully include some support for CVP.

As personal video recorders (PVR) such as TiVo and ReplayTV gain popularity, the TVAnytime Forum (<http://www.tv-anytime.org>) is defining a specification for a second generation personal video recorder which has a customized video playback facility. One of this group's stated four-fold objectives is to "define specifications that will enable applications to exploit local persistent storage in consumer electronics platforms" (<http://www.tv-anytime.org/about/index.html>). More specifically the organization's working group on "Metadata Specification" seeks to establish "Segmentation Metadata" as a means "to edit content for partial recording and non-linear viewing" and to "navigate within a piece of segmented content" (<http://www.tv-anytime.org/workinggroups/wg-md.html>).

Given a widely accepted VAD format at some point in the future and a few widely used selective playback mechanisms, various tools should arise to facilitate the creation of VADs and the CVP applications for a particular playback mechanism. Then Customized Video Playback will be able to fulfill its potential. Of course, one key to the future of CVP is an adequate VAD format that gains wide acceptance. Given that standards are important to create critical mass and a propitious environment for development (Bush, 2002) and since our proposal is based on the IEEE LOM and MPEG-7, we will provide an introduction to these two existing standards.

An Introduction to the IEEE LOM and Related Activities (ARIADNE, IMS, and SCORM)

In 1996, the Learning Technology Standards Committee (LTSC) was formed within the Institute of Electrical and Electronics Engineers (IEEE). The committee is chartered to "develop accredited technical standards, recommended practices and guides for learning technology" (<http://ltsc.ieee.org/>). The LTSC is divided into various working groups that each hold quarterly face-to-face meetings as well as teleconferences to accomplish their standards work. The LTSC working group of most interest here is the Learning Object Metadata (LOM) group.

During the timeframe from 1995 to 1997, there were a few groups working on metadata issues. An early project that served as a foundation for work by other groups was the Dublin Core Metadata Initiative (<http://www.dublincore.org>). Two groups that took the Dublin Core metadata work to further expand it were the Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) (<http://www.ariadne-eu.org/>) and the Instructional Management Systems (IMS) project, later renamed the IMS Global Learning Consortium (<http://www.imsglobal.org/>). Both groups, ARIADNE and IMS, contributed their early metadata work to the LOM working group and all three combined their efforts to settle upon the specific elements that would comprise the final LOM standard. On June 12th, 2002 the LOM group finished version 1.0 of the Learning Object Metadata standard which will be known as IEEE 1484.12.1 (IEEE LTSC, 2002).

In the view of many individuals in this field, one of the benefits of having an IEEE committee working on learning technology standards is that the IEEE is an accredited standards body (in the United States, standards bodies are accredited by ANSI, the U.S. national member of ISO). This status can give IEEE standards more permanence and sometimes more credibility than specifications emerging from industry consortia such as the IMS. On the other hand, due to the careful deliberations and the need to take into account many different perspectives that are sometimes competitive in nature, accredited standards bodies are known to generally take long periods of time to produce a final standard. For example, the LOM version 1.0 standard has been roughly six years in the making, a very long time in the world of learning technologies, given the number of changes that can occur in that amount of time.

Recognizing the intentionally slow and deliberate nature of accredited standards bodies, other groups in the field of learning technology began taking the preliminary metadata element set and matching it with current technologies. The IMS produced the first XML binding of an early LOM metadata set in August, 1999. Since that time the IMS has updated its binding to reflect the progressively refined versions of the LOM information model that have occurred over time. This has served IMS members well, since many wanted to produce metadata records but had no specific guidance from the LOM group on how to represent the LOM information model using technologies such as XML. The Shareable Content Object Reference Model (SCORM) coming out of the Advanced Distributed Learning (ADL) initiative (<http://www.adlnet.org/>) was one such group. The SCORM references the IEEE LOM information model but utilizes the IMS LOM XML binding throughout its documentation and testing software. In June of 2002, the LOM working group formed an XML binding sub-group to work on an IEEE version of an XML binding for the LOM model. As of this writing, there are no estimates as to when this work will be completed.

One of the explicit design decisions made by the LOM working group with their information model and the IMS with their binding was to leave LOM records as open and flexible as possible. There is always a tradeoff between flexibility and interoperability. The more flexible the schema, the less interoperable it is and vice-versa.

In developing the VAD format, this tradeoff was taken into account as we explored ways to incorporate detailed descriptions of video assets with the more general metadata provided by the LOM information model. The thinking is that we can utilize the education related metadata that comes with LOM to describe the educational aspects of a video asset and we can also extend the <technical> element of the LOM schema to provide the extra detail needed for customized video playback. The additional metadata elements we will utilize for this aspect of the project come from the MPEG-7 working group.

An Introduction to MPEG-7

An ISO/MPEG working group was organized in October 1996 and began building the requirements for what is now called the “Multimedia Content Description Interface” or MPEG-7. This was created for the purpose of describing previously existing content to be compressed using current and future standards such as MPEG 1, 2 and 4. After the call for proposals in October 1998 and going through the initial standards development process, MPEG-7 entered the

final editing phase in September of 2001 and is well on its way towards being published as international standard ISO 15938 (Manjunath, Salembier, and Sikora; 2002).

The need for a model to describe multimedia content in this age of increased digital content production, expanded storage capabilities, and enhanced content consumption venues became readily apparent to a group of content creators, distributors, and technologists. As a result, they came together as the MPEG-7 community (<http://www.mpeg-industry.com/>), and developed MPEG 7 as **the** superset of multimedia content descriptions. It thus can then be used as a resource for efforts such as ours that are working to implement content description schemas. Where MPEG 1, 2, and 4 dealt primarily with compression standards, MPEG-7 exists as a descriptive mechanism for multimedia content. As Nack and Lindsay wrote:

MPEG-7 ... focuses on the standardization of a common interface for describing multimedia materials (representing information about the content, but not the content itself—"the bits about the bits"). In this context, MPEG-7 addresses aspects such as facilitating interoperability and globalization of data resources and flexibility of data management (1999, July-September, 69)

As can be seen from the final version of the "MPEG-7 Overview (version 8)" (Martinez, 2002), the specification itself covers a comprehensive range of functionality across audio-visual media. For example, one of the playback mechanisms mentioned earlier is the TV-Anytime technology being developed for personal video recorders, a technology that is developing with some level of mutual consideration of the work underway within the MPEG community. Indeed, "TV-Anytime uses the MPEG-7 Description Definition Language (DDL) to describe metadata structure as well as the XML encoding of metadata" (Cover Pages, 2002). On the MPEG-7 side of the developments we learn that:

MPEG-7 does not define a monolithic system for content description but rather a set of methods and tools for the different viewpoints of the description of audiovisual content. Having this in mind, MPEG-7 is designed to take into account all the viewpoints under consideration by other leading standards such as, among others, TV Anytime, Dublin Core, SMPTE Metadata Dictionary, and EBU P/Meta (Martinez, 2002, 4).

As a member of the United States Technical Advisory Group to ISO JTC1/SC29/WG11 (MPEG), Brigham Young University has reviewed the pre-publication MPEG documents for use in the VAD project. Since the MPEG 7 community is collecting information and sample applications showing how the standard has been implemented, we have been able to present our included sample VAD (see Appendix) at a recent MPEG 7 Awareness Event. Our presentation was embraced by the community and we are working on finalizing it as a "profile" for the MPEG 7 standard.

To insure applicability across a wide variety of types of audio-visual media and content types, MPEG-7 contains a broad range of descriptive power. It implements a Description Definition Language (DDL) based on XML that contains:

- The XML Schema structural language components,
- The XML Schema datatype language components, and
- The MPEG-7 specific extensions (Martinez, 2002, 9).

In addition, the MPEG-7 specification provides for the description of visual attributes, audio attributes, as well as Multimedia Description Schemes (MDS). MPEG-7 contains descriptors for such visual information as color, texture, shape, motion, localization, and face recognition. Its audio framework “deals with low-level Descriptors, for audio features that cut across many applications (e.g., spectral, parametric, and temporal features of a signal), and high-level Description Tools that are more specific to a set of applications” (Martinez, 2002, 9). These Description Tools from the MPEG-7 audio framework work in conjunction with the Multimedia Description Schemes (DSs) to provide multiple, descriptive, data streams for the media asset:

The DSs provide a standardized way of describing in XML the important concepts related to AV content description and content management in order to facilitate searching, indexing, filtering, and access. The DSs are defined using the MPEG-7 Description Definition Language (DDL), which is based on the XML Schema Language, and are instantiated as documents or streams. The resulting descriptions can be expressed in a textual form (i.e., human readable XML for editing, searching, filtering) or compressed binary form (i.e., for storage or transmission). (Martinez. 2002, 10)

With respect to the high-level Description Tools of interest to the VAD project, the MPEG-7 specification contains Spoken Content Description that provides a detailed description of words spoken within any given audio stream of a media asset. In addition, planned additions to the specification will allow for “linguistic description” in order to accommodate new Multimedia Content Entity tools. Basically it is possible to encode “the semantic structure of linguistic data as part of or associated with multimedia/multimodal content, such as scenarios, transcriptions, critiques, and so forth” (Martinez, 2002). This inclusion will definitely be a great help for the work required for the development of a VAD

Defining the VAD by Combining the LOM and MPEG-7

In the e-learning community, most efforts to create formats for shareable and reusable content objects are based on the IEEE Learning Object Metadata (IEEE LOM) described in earlier in this paper. Because the IEEE LOM does not include all the elements necessary to describe a video asset, such as a mechanism for describing a hierarchical segmentation of a film or other video asset, we had to look elsewhere. The forthcoming ISO MPEG-7 standard as described briefly above includes all the elements needed to describe a video asset, but MPEG-7 is not well known within the e-learning community, and thus a pure MPEG-7 object would not fit into the larger e-learning picture of sharing descriptions of various kinds of educationally-relevant assets using a common metadata framework. Therefore, we decided to combine the strengths of the IEEE LOM and MPEG-7 standards into a video asset description (VAD) format that both qualifies as an extended IEEE learning object and that can be converted to a pure MPEG-7 object (with some degree of information loss) as well as to multiple internal formats used in instructional technology applications.

In one sense and from the standpoint of solid principles of metadata development, our VAD format is an “application profile”, a necessary reality because “no single metadata element set will accommodate the functional requirements of all applications” (Duval, Hodgins, Sutton, and Weibel; 2002; April). Furthermore, since the IEEE LOM and MPEG efforts began about the same time, it is not surprising that there has been little contact between the two groups in the last few years, although the *MPEG-7 Overview* (Martinez, 2002) does reference the Dublin Core Metadata Initiative (DCMI, 2002). Despite the fact that MPEG-7 and the IEEE LOM share this early element of common heritage, it quickly became apparent to us in our work that our requirements for a VAD useful for our applications would have to draw on both these efforts.

The methodology we have pursued consists of the following steps:

1. Develop a summary of IEEE LOM,
2. Develop a summary of MPEG-7, Part 5, Sections 8-11,
3. Compare the IEEE LOM and MPEG-7 summaries to develop a map of correspondences,
4. Analyze the MPEG-7 elements that do not correspond to LOM elements and decide on subset,
5. Define formally the VAD as the IMS binding of the IEEE LOM, extended with a VAD namespace consisting of the subset of MPEG-7 elements identified in step 4

Once the VAD format has been defined, it will be tested in a variety of environments. If additional data elements are identified as needed, we will search diligently in the LOM and MPEG-7 and even suggest additions to them before creating our own data elements.

Conclusions

One way to understand the significance of this work is to understand the fundamental difference between the approaches used in IEEE LOM and MPEG-7. Our determination is that the first is well-suited to assure retrievability, while the second is useful for providing a technical description of the asset.

An analogy would be the Multiple Listing Service (MLS) widely used across the United States for cataloging available homes and other real estate. Using the MLS, it is possible to easily find a home that corresponds to a particular set of criteria. To build a copy of or to remodel that house, however, would require a lot more information, such as that contained in a set of architect’s plans. Where IEEE LOM plays the role of the MLS for educational content, MPEG-7 provides the equivalent of the architect’s plans. In other words, for certain applications, the IEEE LOM will serve the useful function of locating content, and MPEG-7 will open up the use of the content once it is located.

For other applications, our sense is that there will potentially be a great deal more content described in terms of MPEG-7 than will be described using IEEE-LOM. Yet, there will be

software that is set up to use the information contained in IEEE LOM. The results of our work will enable us to create software to allow the conversion of content descriptions from one system to the other.

In addition, CVP is important not only to education but also to home use. In either case, it will be impossible for any benefit to be derived without appropriate specifications for content modeling as well as tools for personalization of the content. The development and adoption of the VAD format would facilitate CVP. Indeed, there is a good chance that intellectual property restrictions with respect to various approaches to CVP are such that VAD will be the only viable means to achieve CVP within a legal framework for content distribution.

Finally, although we have moved a long way down the road to developing a viable approach to VAD development, there is a great deal of work to be accomplished. We have begun the process for making the results of our work available within the MPEG-7 and IEEE LOM communities, thus allowing for the continued evolution of the VAD specification. It is in fact not out of the question that either of these two specifications will be extended based on real-world use of the VAD format. So to insure that our work is available as widely as possible, we have set up a Web site that contains various files that document the VAD format: <http://zola.byu.edu/vad>. Input is solicited from any and all interested individuals.

REFERENCES

- Bush, M. (2002, November/December). Connecting Instructional Design to International Standards for Content Reusability. *Educational Technology*, In press.
- Cover Pages (2002). TV-Anytime Forum Publishes Metadata Specification for Network Broadcast Content. <http://xml.coverpages.org/ni2001-10-02-c.html> .
- Dublin Core Metadata Initiative (2002). Web site: <http://www.dublincore.org/>.
- Duval, E., Hodgins, W., Sutton, S. and Weibel, S. L. (2002, April). Metadata Principles and Practicalities. *D-Lib Magazine*, 8 (4).
<http://www.dlib.org/dlib/april02/weibel/04weibel.html>
- Gale, L. E. (1989). Macario, Montevideodisco and Interactive Digame: Developing interactive video for Language instruction. In W. Flint Smith (ed.), *Modern Technology in Foreign Language Education: Applications and Projects*. pp. 235-248. Lincolnwood, IL: National Textbook Company.
- IEEE LTSC (2002). *Draft Standard for Learning Object Metadata*. IEEE Learning Technology Standards Committee (LTSC) P1484.12 Learning Object Metadata (LOM) working group. http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf
- Kennedy, T. and Slowinski M. (2002) *SMIL: Adding Multimedia to the Web*. Indianapolis, Indiana: Sams Publishing. (Synchronized Multimedia Integration Language.)
- Manjunath, B., Salembier, P., and Sikora T. (2002) *Introduction to MPEG-7: Multimedia Content Description Interface*. West Sussex, England: John Wiley & Sons, Ltd.
- Martínez, J. M. (Ed) (2002). *Coding of Moving Pictures and Audio, MPEG-7 Overview*. ISO/IEC JTC1/SC29/WG11N4980. MPEG Alliance Web Site.
http://www.mpeg-industry.com/mp7a/w4980_mp7_Overview1.pdf.
- Microsoft (2002). MSDN Web Site.
http://msdn.microsoft.com/library/default.asp?url=/library/en-us/directshow_sp1/htm/writingascripbaseddvdapplication.asp (if this URL is no longer available, go to <http://msdn.microsoft.com/>; then search for "mswebdvd activex control".)
- Nack, F. and Lindsay, A. T. (1999, July-September). Everything You Wanted to Know About MPEG-7: Part 1. *IEEE MultiMedia*, 6(3), 65-77.
<http://lass.cs.umass.edu/~shenoy/courses/spring01/papers/ieeemm-mpeg7.pdf>
- (1999, October-December). Everything You Wanted to Know About MPEG-7: Part 2. *IEEE MultiMedia*, 6(4), 64-73.
- Negroponete, N. (1994, August). Prime Time Is My Time: The Blockbuster Myth. *Wired* 2.08, 2(8). <http://www.wired.com/wired/archive/2.08/negroponete.html>

Appendix (sample VAD document)

A VAD file is an XML document with the root element <lom> from the IMS binding of the IEEE LOM, extended with MPEG-7 elements. The high-level LOM elements used in this example are **general** (information about the video asset), **lifecycle** (versions, etc.), **metametadata** (information about the Video Asset Description), and **technical** (detailed information about the video asset, using MPEG-7 elements).

```
<?xml version="1.0"?>
<lom xmlns="http://www.imsglobal.org/xsd/imsmd_rootv1p2p1"
  xmlns:mp7="urn:mpeg:mpeg7:schema:2001"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=" http://www.imsglobal.org/xsd/imsmd_v1p2 http://www.imsglobal.org/xsd/imsmd_v1p2p2.xsd
  urn:mpeg:mpeg7:schema:2001 ../mds-2001.xsd">
  <general>
    <identifier>963853ef726b5104</identifier>
    <catalogentry><catalog>ISBN</catalog><entry><langstring>0-7888-2700-6</langstring></entry></catalogentry>
    <title><langstring>Chocolat</langstring></title>
    <language>en-US</language>
    <description><langstring>The impact of a chocolate shop on a French village</langstring></description>
  </general>
  <lifecycle><version>1.3</version></lifecycle>
  <metametadata>
    <identifier>{43EC1762-4820-4293-B19E-804E08B8AE42}</identifier>
    <contribute>
      <role><source><langstring>VADv1.0</langstring></source><value><langstring>creator</langstring></value></role>
      <centity><vcard>BEGIN:vCard
        FN:Alan Melby
        ORG:Brigham Young University
        END:vCard
      </vcard></centity>
      <date><datetime>2002-09-28</datetime></date>
    </contribute>
    <metadatascheme>LOMv1.0</metadatascheme>
    <language>en-US</language>
  </metametadata>
  <technical>
    <format>video/mpeg</format>
    <duration><datetime>01:57:20:00</datetime></duration>
    <mp7:Mpeg7>
      <mp7:DescriptionMetadata><mp7:LastUpdate>2002-09-28</mp7:LastUpdate></mp7:DescriptionMetadata>
      <mp7:Description xsi:type="MediaDescriptionType">
        <mp7:MediaInformation><mp7:MediaProfile><mp7:MediaFormat>
          <mp7:Content href="urn:mpeg:mpeg7:cs:ContentCS:2001:2">
            <mp7:Name>Audiovisual</mp7:Name>
          </mp7:Content>
          <mp7:Medium href="urn:mpeg:mpeg7:cs:MediumCS:2001:1.3">
            <mp7:Name>DVD</mp7:Name>
          </mp7:Medium>
        </mp7:MediaFormat>
      </mp7:MediaInformation>
    </mp7:Description>
  </mp7:Mpeg7>
</lom>
```

```

</mp7:MediaFormat></mp7:MediaProfile></mp7:MediaInformation>
</mp7:Description>
<mp7:Description xsi:type="ContentEntityType">
  <mp7:MultimediaContent xsi:type="VideoType">
    <mp7:Video id="MainTitle">
      <mp7:TemporalDecomposition>
        <mp7:VideoSegment id="chapter1">
          <mp7:TemporalDecomposition>
            <mp7:VideoSegment id="chapter1scene1">
              <mp7:TextAnnotation type="description">
                <mp7:FreeTextAnnotation>opening credits; music; village aerial view</mp7:FreeTextAnnotation>
              </mp7:TextAnnotation>
            </mp7:VideoSegment>
            <mp7:MediaTime>
              <mp7:MediaTimePoint>T00:00:00</mp7:MediaTimePoint>
              <mp7:MediaDuration>PT1M24S</mp7:MediaDuration>
            </mp7:MediaTime>
          </mp7:VideoSegment>
          <mp7:VideoSegment id="chapter1scene2">
            <mp7:TextAnnotation type="description">
              <mp7:FreeTextAnnotation>entering church; bells; Count introduced</mp7:FreeTextAnnotation>
            </mp7:TextAnnotation>
            <mp7:MediaTime>
              <mp7:MediaTimePoint>T00:01:24</mp7:MediaTimePoint>
              <mp7:MediaDuration>PT0M20S</mp7:MediaDuration>
            </mp7:MediaTime>
          </mp7:VideoSegment>
          <!-- remaining scenes go here -->
        </mp7:TemporalDecomposition>
      </mp7:VideoSegment>
      <!-- Remaining chapters go here -->
    </mp7:TemporalDecomposition>
  </mp7:Video>
</mp7:MultimediaContent>
<mp7:MultimediaContent xsi:type="AudioType">
  <mp7:Audio id="english_full">
    <mp7:MediaTime>
      <mp7:MediaTimePoint>T00:00:00</mp7:MediaTimePoint>
      <mp7:MediaDuration>PT01H57M20S</mp7:MediaDuration>
    </mp7:MediaTime>
  </mp7:Audio>
</mp7:MultimediaContent>
</mp7:Description>
</mp7:Mpeg7>
</technical>
</lom>

```