

## Standards, Specifications, and Paradigms for Customized Video Playback

**McDonald, Jarom Lyle**

jarom\_mcdonald@byu.edu  
Brigham Young University, USA

**Melby, Alan K.**

melbyak@yahoo.com  
Brigham Young University, USA

**Hendricks, Harold**

harold\_hendricks@byu.edu  
Brigham Young University, USA

---

Culture is fully inundated with video--from the ubiquitous DVD and succeeding optical media formats, to locally stored digital bits passed between DVRs and video iPods, to over a billion Internet-streamed videos a day. Unfortunately, while those involved in humanities education and research know how widespread video usage is and are attempting to integrate such a rich medium into what they do, they are more often than not struggling, fighting against the medium and associated baggage rather than using video for their own purposes.

For all of the ways in which video differs from other forms of media, perhaps the most challenging obstacle to effectively utilizing video assets as objects of teaching and research is their inflexibility. Because of the complexity of video technologies and the pressure of external interests, video is an incredibly closed medium, especially when compared to text, image, or even audio. In many ways video resists fundamental activities of digital humanities inquiry such as metadata, structural, and segment analysis and annotation. What's more, video also is, technologically speaking, a linear medium; it is (as much as if not more than other media) architected to proceed continuously from point A to point B, serving up bits in order and only responding to very limited, legacy interface controls. Even the "interactivity" touted by content holders (such as DVD "extras") is a rigid, linear interactivity, designed to keep the control of playback under the stewardship and

limited scope of the video producer rather than the needs of the learner, the desires of the scholar, or the tastes of the consumer. To encourage collaborative, reusable approaches to video (while avoiding legal pitfalls or isolationist tendencies that come with an extracted clip approach), we need to incorporate a more thorough, flexible, and widespread method of *customized* video playback.

The papers in this panel will focus on data-driven customized video playback (CVP), from theory and methodology to real-world use cases that are evolving and practical implementations that are both already in use as well as under development to meet the needs of the Humanities today. The first presentation will make the case for the fundamental groundwork for video asset analysis and eventual customized video playback, the Multimedia Content Description Interface (also known as MPEG-7). This XML standard for describing (both globally and in timecode-associated ways) video assets offers a markup solution that is complementary to common video encoding containers (such as MPEG-2 and MPEG-4), and, as XML, can be easily coupled with other relevant data and metadata standards as well. The second paper will present an argument for ways to take these video asset descriptions and use them to enable both people and technology to better facilitate customized video playback using a videoclip playlist specification (serializable as plain text, as XML, as JSON, or as any other data exchange format). With the segment descriptions of a thorough video asset description, a videoclip playlist can then define custom playback operations. The final presentation will demonstrate several use cases of customized video playback, along with working models for achieving the type of interactivity we desire with the technologies we have today, including demonstration of a CVP system in use at several university campuses.

The panel as a whole will seek to argue a unified justification and methodology of customized video playback, and invite future collaboration from the Digital Humanities community who can, if they desire, push these ideas further towards making our proposed standards, specifications, and paradigms as widespread, useful, and effective as possible.

## Finding the Best in Approaches to Video Asset Description

**McDonald, Jarom Lyle**

jarom\_mcdonald@byu.edu

Brigham Young University, USA

---

From Google's Web Services to Wikipedia's DBpedia project to the underlying architecture of modern digital libraries, our notion of how to make data more semantic is moving (slowly but persistently) towards ideal principles that the W3C lays out for what is commonly called "the Semantic Web." This is even true for the subject of my study, video data, albeit with much less of a semantically-inflected critical mass. There are a few solid, innovative investigations (such as the BBC's video portal and the many incarnations of the Joost video platform) that are or have been working to bring technologies such as metadata, RDF/RDFa, and SPARQL to the storage and dissemination of video (especially online video); but there is still a lot of work that needs to be done in order to make today's video assets truly useful in a way that Tim Berners-Lee would approve, a world "in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Scientific American).

While the Semantic Web includes a large number of topics too broad to cover in this proposal, I will focus on one particular aspect of semantic markup that does apply to video data. It is vital to underscore the unique nature of video as an object of *perception*--that is, video is meant to be played for a viewer with the linear, temporal nature in the forefront of experiencing the video. Thus to describe the data of a video asset, as a whole, in a useful way would necessarily require a structured analysis of more than just the metadata *about* the video that you might be able to achieve with Dublin Core, IEEE-LOM, or RDF; the most significant need is a system that can connect such semantic vocabularies to a thorough, analytic description of the video content itself in as close an approximation to the playback act as might be reasonably able to achieve--in other words,

a workable time-coded markup language. This isn't to say that a video must be necessarily viewed chronologically; rather, given that video exists as bits served from time point A to time point B, it must be described that way in order to make use of the data encoded there. If a video asset has the right description of its segmented, time-coded parts (of which, naturally, there may be many versions based on who is doing the markup or who is using the materials), it will eventually allow for more than just watching the video; a segmentation model of video markup is essential for enabling a system of interactive, customized video playback.

Several options for such a language to use are available and have been somewhat explored both commercially and academically, but none are completely satisfactory. Naturally, given the success of the Text Encoding Initiative, it makes sense to consider its ability to function as a time-coded video markup system. In fact, Reside (2007) and Arneil and Newton (2009) have presented just such an idea at recent Digital Humanities conferences. The flexibility and thoroughness of the TEI makes it an attractive option; however, while the speech transcription models can potentially provide time-coded descriptions of spoken elements of a video (and even be retrofitted to other elements of video content), because the TEI is a text-encoding framework, it lacks a temporal segmentation scheme designed specifically for existing models of video encoding and playback (for example, referring to multiple video or audio tracks, multiplexing metadata with the binary streams, etc.). Most projects exploring video markup descriptions also mention the W3C's Synchronized Multimedia Integration Language (SMIL). Since version 3.0, SMIL integrates a temporal segmentation model with one for spatial fragmentation, allowing semantic relationships both within and between video elements. What's more, SMIL is a W3C recommendation, offering the potential for tighter integration with web delivered video as it continues to mature. Several commercial endeavors (including the streaming platform Hulu) have incorporated SMIL into their playback process, allowing for a sophisticated combination of video annotation (for example, Hulu uses it for their advertisements and upcoming social viewing features) and search/retrieval (combining the time-coded markup

with RDF metadata). Yet SMIL provides no scope for how to implement various temporal segment references, instead leaving that task up to playback mechanisms (of which there are currently very few for SMIL). More significantly, SMIL provides no way to refer to described segments out of the context of the document, making it difficult to design a URI scheme for accessing the various clips, something integral to having true semantic web functionality or building video players that can interpret the descriptions consistently. Video streaming servers such as those provided by Apple, Microsoft, and Adobe have all developed their own model for segmenting (either in markup or in actual bits) video data and serving it with instructions for playback, but in these cases the systems very heavily limit the metadata that can be included, and the resulting descriptions are completely coupled to the proprietary vendor technologies (for example, a set of Flash Streaming Media Server cue points is not portable to other systems without intervention).<sup>1</sup>

For the past 10 years, the Moving Picture Experts Group has defined and refined what they've formally titled the "Multimedia Content Description Interface," also known as MPEG-7. MPEG-7, an XML-based, ISO/IEC standard, is an expansive, far-reaching specification that does many, many things (including defining itself and defining the language by which it defines itself and its various parts); what is of particular interest to this proposal is Part 9, "Profiles and Levels." Recognizing that there are many approaches and viewpoints surrounding video asset description (those mentioned above, plus such systems at TV-Anytime, the SMPTE Metadata Dictionary, and even extensions to the Dublin Core standard), MPEG-7 seeks to be a superset of video markup, and the concept of "profiles" as laid out in Part 9 of the spec offers various focused schemas and methodologies that conform to the MPEG-7 spec but serve unique needs. Seeing a need for a general purpose, video-specific description language, Brigham Young University has collaborated with Motorola and the Japanese National Broadcasting Corporation to publish a "Core Description Profile" (CDP), a framework that utilizes MPEG-7 descriptions and provides all the necessary tools for time-coded, segmented, video annotation.

Every file that conforms to the CDP schema (a schema now included directly as part of the MPEG-7 specification and which has been released as open-source by ISO) must also conform to the MPEG-7 super-schema. In addition to the MPEG-7 root element and any necessary header information, a CDP document has a series of <description> elements that contain <MultimediaContent>; a simple example of such an element might look something like this:

```
<MultimediaContent xsi:type="VideoType">
  <Video id="MainTitle">
    <TemporalDecomposition>
      <VideoSegment id="chapter1">
        <TemporalDecomposition>
          <VideoSegment id="chapter1scene1">
            <TextAnnotation type="description">
              <FreeTextAnnotation>opening credits;
                music; village aerial view
              </FreeTextAnnotation>
            </TextAnnotation>
            <MediaTime>
              <MediaTimePoint>T00:00:00
            </MediaTimePoint>
              <MediaDuration>PT1M24S
            </MediaDuration>
            </MediaTime>
          </VideoSegment>
          <VideoSegment id="chapter1scene2">
            <TextAnnotation type="description">
              <!-- other types of annotations are
                possible as well -->
              <FreeTextAnnotation>entering church;
                bells; Count introduced
              </FreeTextAnnotation>
            </TextAnnotation>
            <MediaTime>
              <MediaTimePoint>T00:01:24
            </MediaTimePoint>
              <MediaDuration>PT0M20S</MediaDuration>
            </MediaTime>
          </VideoSegment>
          <!-- remaining scenes go here -->
        </TemporalDecomposition>
      </VideoSegment>
      <!-- Remaining chapters go here -->
    </TemporalDecomposition>
  </Video>
</MultimediaContent>
```

Having been designed as a general-purpose video asset description schema, the CDP is the most promising format for defining video clip boundaries, including metadata and annotations. On the surface it may not seem much different from other markup schemas such as SMIL; however, the real power of the approach lies in combining a CDP-conformant description with other parts of the MPEG-7 specification. First of all, because MPEG-7 is the data description framework for the same

group behind the MPEG-2 and MPEG-4 video encoding containers and codecs, MPEG-7 can be easily multiplexed directly into a video container. Additionally, as XML, a CDP video asset description can incorporate any other relevant data through namespaces, including TEI, Dublin Core, RDF relationships, or future information schemas. And finally, because it is a general purpose description framework, it can also be serialized into any needed format such as SMIL, IEEE-LOM (a Learning Object Metadata standard) or CMML (the Continuous Media Markup Language), an XML schema defined by the organization behind the Ogg media formats and promising tight integration with emerging HTML5 video technologies. To project this even further, imagine a robust, RDF-aware repository (such as FEDORA) full of digital video objects that connect video streams to valid MPEG-7 video asset descriptions, making videos easily discoverable, easily searchable, and ultimately, truly semantic. And finally, this approach to video asset description lays the foundation for, to return to Tim Berners-Lee's comment, enabling us to make video playback from computer to person customizable, flexible, and just what we need it to be.

---

## References

**Berners-Lee, Tim, James Hendler and Ora Lassila.** *The Semantic Web. Scientific American May, 2001.* <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.

**Bush, Michael D., Alan K. Melby, Thor A. Anderson, Jeremy M. Browne, Merrill Hansen, and Ryan Corradini** (2004). 'Customized Video Playback: Standards for Content Modeling and Personalization'. *Educational Technology*. **44.4** : 5-13.

*MPEG-7 Overview.* <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.

*W3C Semantic Web Frequently Asked Questions.* <http://www.w3.org/2001/sw/SW-FAQ#what1>.

---

## Notes

1. It's notable to also mention MIT's Cross Media Annotation System (XMAS), a project developed over the last decade to incorporate time-aligned video annotation and commentary into Shakespeare classes at MIT. The approach that XMAS

takes is much more in line with what we see as the proper way to annotate video; however, theirs is a closed system very tightly coupled to their specific Shakespeare needs, so it isn't known what technologies they're using or how portable those technologies might be.

## Videoclips as Playlists in Customized Video

**Melby, Alan K.**

[melbyak@yahoo.com](mailto:melbyak@yahoo.com)

Brigham Young University, USA

---

The video interface that we are all familiar with has garnered universal acceptance. This is true both of the iconic symbols of playback control, as well as the actual functions of control that are allowed. Since the days of very early analog playback, media consumers have been allowed to do pretty much just the following:

- Insert (or open) media
- Play
- Pause
- Stop
- Fast-forward
- Rewind
- Volume control (including muting)

There have been a few additions to the list as media (especially video) technology has evolved; for example, with the introduction of laserdisc and subsequently DVD, "return to menu," "next/previous chapter," and "go to title" controls have entered the collective interface. Streaming media has also invoked the need for a "fullscreen toggle" control. And many playback systems are now allowing for the display of certain content-provider defined metadata or even a "scrub" bar for more flexible time placement. But such change comes slowly, and all of these playback controls are very simple, require a good deal of user intervention, and truly limit the video asset's effectiveness as a learning object.



Image 1: A current video playback interface, offering only standard, simple user controls (source: <http://commons.wikimedia.org>)

To provide an example of the limitations of commonly accepted playback, let me briefly describe what happened this semester in a Hebrew class at our university. The instructor wanted to use several clips from a commercial Israeli movie to help the students with their vocabulary. With current US laws and technological learning curves making it unreasonable to rip the clips from the DVD and edit them for her class's needs, she instead instructed them to check out the DVD from the reserve library, start playing it, jump to chapter 14, watch for 4 minutes and 17 seconds (rewinding if necessary to fully understand the dialogue), jump to 2 minutes into chapter 31, watch for 3 minutes, pause the DVD and go look at some of the resources she'd posted on her course website, return to the DVD, etc. etc. This is an extremely ineffective way to use the video in a learning environment, but unfortunately, it's really the only methodology that is widely available right now.

Historically, this hasn't always been the case. Laserdisc technology provided unprecedented control for both users and instructional designers over the video asset, and a large number of rich learning experiences were created and shared on laserdisc. More recently, the concept of "WebDVD" seemed poised to recover some of the lost functionality when laserdisc didn't emerge outside of a niche market; but WebDVD didn't catch on, either. Technologies of streaming media (such as Youtube annotations or bookmark-driven systems in place at CNN, ABC, Hulu, and other commercial streaming media institutions) have some potential, but they are still in their infancy,

aren't widely used, and don't allow for anything other than an editorial overlay; the content itself is still played back under very strict control that the user can only pause, stop, rewind, etc.

What we propose, then, and what is so desperately needed, is a completely different playback system that allows for true interactivity for instructional designers, teachers, and viewers. Bush et al. propose two models for customizing video playback that seek to alleviate the sort of haphazard "playlist" illustrated earlier. One, of course, would be to strip the digital bits onto a local filesystem, edit the content as needed, add in annotations (subtitles, links, external info), and share the new video with all students to view in traditional playback systems, repeating the process when the video content might need to be viewed in a different way. But this (as the authors point out) is time consuming and expensive, not to mention the unfortunate copyright implications of such an approach.

The other model for enabling customized video playback, and what we are currently developing, is to combine a data-driven, "descriptive" approach with a "selective playback mechanism", software specifically built for customized video playback according to a robust, standardized specification for delineating the different actions that the designer/instructor might want to have students experience. The most promising form of video description is the Video Asset Description (VAD), an XML encoding of clip boundaries, video content, and other metadata that is isomorphic with the MPEG-7, part 9 core description profile and which is described in another paper on this panel. When a video asset is associated with a full VAD (or several of them), an instructor--or in many cases even automated software--can use that description to generate a playlist. When I say "playlist," I'm not referring to the common usage of the term as a description of a media collection (a list which describes what assets to play and in what order), but it is similar; what I propose is a notion of a video clip playlist (VCP), a description of timecoded clips within a video asset. A collection of these clips, along with the actions to take for each clip, could be fed directly into the queue of selective playback software, software that would be programmed to

know how to read the instructions and present the new playback session.<sup>1</sup>

Each instruction in a videoclip playlist would be a triple that would consist of a framecode number, an operation, and an operand. The framecode is not a direct representation of either human perception of time or of frame count, but is instead a convenient fiction that allows for the most effective and standardized accuracy in calculating either time or frame. The operation, most easily represented as a numeric opcode, would be one of a number of operations that a playback system might encounter. Of course there would be the standard "play, pause, stop, mute" controls, but they would be under the stewardship of the instructional designer who is authoring the videoclip playlist. There would also be codes for jumping to a new timecode (not just a chapter, not just an approximate location on a scrub bar, but a frame-accurate location), for displaying annotations (subtitles, scholar-composed notations, instructor comments, etc.), displaying "wrap" data (for example, material retrieved from web services and displayed in an additional pane of the selective viewer at the precise moment the playlist instructs), and so forth. The operand would be a piece of data that makes the op code intelligible; if a command instructed the player to jump to a new time code, the operand would be the time code to jump to. If a command instructed the player to start playing a clip, the operand might be the number of frames to play.

We have designed a simple RNG schema for encoding these clips in an XML file that is both machine and human readable. The file would have some header information that identifies the videoclip playlist, associated video asset descriptions, and references to wrap data and other annotations, as well as an instructionList of the commands that the player would need to perform the custom playback. The instructionList looks like this:

```
<instructionList>
  <instruction trigger="0" opCode="68"
operand="60">show 'skipping' message for 2
seconds</instruction>
  <instruction trigger="0" opCode="0"
operand="0">pause before seeking</instruction>
  <instruction trigger="0" opCode="65"
operand="52164">seek to frame 52164</
instruction>
  <instruction trigger="52164" opCode="75"
operand="32">new clip [32] begins</instruction>
```

```
<instruction trigger="52164" opCode="85"
operand="0">show wrapData #0</instruction>
  <instruction trigger="52164" opCode="85"
operand="1">show wrapData #1</instruction>
  <instruction trigger="52164" opCode="85"
operand="2">show wrapData #2</instruction>
  <instruction trigger="53753" opCode="68"
operand="60">show 'skipping' message for 2
seconds</instruction>
  <instruction trigger="53753" opCode="0"
operand="0">pause before seeking</instruction>
  <instruction trigger="53753" opCode="65"
operand="99926">seek to frame 99926</
instruction>
  <instruction trigger="99926" opCode="75"
operand="57">new clip [57] begins</instruction>
  <instruction trigger="99926" opCode="85"
operand="3">show wrapData #3 </instruction>
  <instruction trigger="99926" opCode="85"
operand="4">show wrapData #4 </instruction>
  <instruction trigger="99926" opCode="85"
operand="5">show wrapData #5 </instruction>
  <instruction trigger="100911" opCode="75"
operand="58">new clip [58] begins</instruction>
  <instruction trigger="100911" opCode="85"
operand="6">show wrapData #6 </instruction>
  <instruction trigger="100911" opCode="85"
operand="7">show wrapData #7</instruction>
  <instruction trigger="100911" opCode="85"
operand="8">show wrapData #8 </instruction>
  <instruction trigger="103033" opCode="99"
operand="-1">end of playlist -- indefinite
pause
</instruction>
</instructionList>
```

A player, of course, would only need to be passed the triples represented by the integer values of each instruction's attributes, and in fact a videoclip playlist could be serialized in any necessary data exchange format, whether it be JSON (for building a browser-based player for customizing streaming media playback), plain text (that might include just tab-delimited integers easily consumable by an appliance with an embedded selective player), and so forth.

Obviously, one key to such an approach to facilitating customized video playback is the creation of the selective players themselves. We are currently undergoing development on specifications that would allow anyone to build such a player. With the combination of robust video asset descriptions, shareable, thorough videoclip playlists, and intelligent, VCP-aware players, customized video playback is once again a reality.

---

## Notes

1. Some might ask why it's necessary to go to the trouble of having a markup layer associated with time-aligned video segments at all; why not use extracted clips? However, the legal and technical obstacles involved in extracting segments of video are far greater problems than those experienced

through the copy/paste of small snippets of text, making clip extraction unfeasible for most cases (not to mention the fact that clips themselves are just as rigid, and must be re-extracted if the use cases change). Moreover, extracted clips are difficult to share and collaborate on, and would still need some sort of annotation layer associated with them for editorial commentary, additional subtitles, etc. Our proposed methodology can handle annotation, collaboration, modification, re-use, and legal restrictions all with one approach.

## Customized Video Playback; Where We've Come From, Where We're Going

**Hendricks, Harold**

harold\_hendricks@byu.edu

Brigham Young University, USA

When we talk about customized video playback, it's important to recognize that the actual playing back of the video is paramount; theories of how to mark up the video's content or describe the desired playback are significant only insofar as they can lead to actual implementations that satisfy some of the use cases that we might envision for a customizable video playback system. These use cases generally fall into three types: one-on-one interactivity, classroom lecture, and large audience presentations, such as annotated cinema. I hope to move the locus of attention from CVP theory to practice in three ways: by discussing some of our historical attempts to achieve such an implementation, by demonstrating a current, working system for customized video playback in use in several academic institutions today, and by outlining where our work is moving next (and how we envision collaborating with others outside our project who have so much to contribute).

The introduction of videodisc technology in the mid 1970s provided the first practical method for inexpensive video storage and random-access playback, especially in the realm of academic instruction. *Macario*, a repurposed Mexican motion picture, was issued as a custom videodisc pressing with interactive menus and annotations coded to the linear playback of the video. When a student would pause the video, the interactive materials would appear, allowing

for commentary, instruction, thought-questions tied directly to individual scenes being watched, and even replay of selected video with choice of audio track.

*Macario's* model of annotated video playback offered some innovative learning opportunities to, for example, intermediate Spanish language and culture classes, but it was still a fairly simplistic model, one built upon and improved over the next few years. Projects such as the German Video Enhanced Learning, Video Enhanced Teaching (VELVET) program empowered students with more custom tools, such as the ability to filter out or select particular types of annotations (both text and image), highlight keywords in accompanying transcriptions, or even perform intricate searching through accompanying materials to narrow in on particular scenes (replaying them as needed) of use to the student. These types of activities demonstrated how useful customized video playback could be, focusing the viewing experience and tailoring it to particular educational needs.

In the late 1980s, Junius Bennion and Glen Probst modified some of these previous models of interactive video to allow more control of video assets within targeted learning experiences. Having first created a methodology for an Apple II-controlled videodisc of *Raiders of the Lost Ark*, Bennion, Probst and James Taylor reprogrammed the content to work with Hazeltine's Time-shared, Interactive, Computer-Controlled, Information Television (TICCIT) System at BYU. Within this modal the motion picture is divided into scenes with light-pen interactivity with annotations, transcriptions, questions, text and audio commentary, and instructional drills. Examples of these TICCIT programs include versions of *Black Orpheus*, *The Seventh Seal*, and *C'eravamo tanto amati*. With the TICCIT modified model, customized, interactive video instruction moved from research projects to the language lab.

The ideas underlying some of these models of customized video playback are the same principles expounded upon in the other sessions of this panel, implemented in the best possible way using the technology available at the time. But they were all inextricably linked to the technology itself, needing custom

produced videodiscs or a complex networked computer system to run. When videodisc technology never caught on (for a number of reasons), these innovative products were made obsolete. Likewise our attempts to achieve robust customized video playback through such frameworks as HyperCard, ToolBook, and "WebDVD" have struggled for much the same reason. Recently, however, we have been able to achieve quite functional and effective CVP through an existing system entitled "Electronic Film Review" (EFR), a methodology for controlling video playback of DVDs.

The EFR approach, demonstrated as a poster session at the 2006 Digital Humanities Conference, is based on the MPEG-7 and VideoClip Playlist open standards discussed earlier, and is designed to be implementable in any media player for time-coded video that supports playing a segment of video based on time codes. The current implementation of the EFR approach runs on Windows XP computers that have decoders suitable for watching DVDs through shared, DirectShow DVD decoders (the current EFR software does not include its own DVD decoders). For individual language study, each user-defined clip of a film can be annotated with vocabulary, culture, and other notes. The EFR player itself includes the video window, the common media interface controls, custom "playlist controls" (for navigating between pre-defined clips), and areas to display the annotations. The EFR system also includes an authoring tool, EFR Aid, as well as a compiler to generate the playlist format, to ease the definition and annotation of various segments of a particular video.

Because the EFR system is based on open standards, any learning materials created for particular video assets are shareable; video asset descriptions and playlists can be transferred from one user to another. What's more, these resources that the EFR system helps create are not coupled to the video asset itself; they are a form of meta-annotation (hence the title of "film review") that do not interfere with a single bit of the video data, thus respecting any copyright laws that might exist. As plain-text (serializable as XML), they are also fully searchable, allowing for discovery of relationships between videos that may not have been previously known. Most importantly, the EFR system makes video

much more than just watchable; it makes video useable.

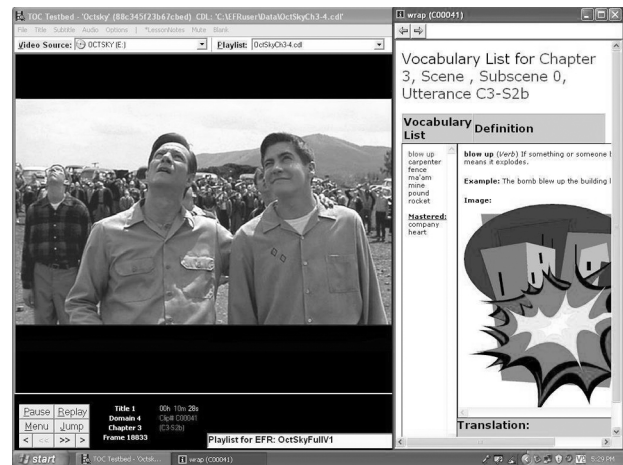


Image 2: A screenshot of the EFR video player

By useable, once again we mention the three primary use cases: individual interactivity, classroom interactions, and annotated cinema. Both our historical efforts and the current EFR implementation have focused primarily on a single user interacting with a computer, with some efforts made to enhance classroom presentation. However, the EFR program has successfully been used in all three of these cases, providing the means for enhanced comprehension, vocabulary building, speech modelling, and cultural awareness for individuals, a means to integrate these same video-based activities into the classroom, and also a tool for modifying the playback of full-length feature films with content filtering and additional subtitles without modifying the copyrighted and encrypted video.

Now, as mentioned earlier, several universities (Brigham Young University, the University of Hawaii at Manoa) have used or are currently using a Windows-XP based EFR system for DVD playback. However, once again the evolution of technology is forcing change in our approach to customized video playback. With the introduction of Blu-Ray, the explosive growth of online, streaming video, and constant legal and political fighting between content providers and content consumers, we see it necessary to broaden the scope of the EFR project to allow for all possible use-cases that we might imagine. We are currently undergoing a project, a collaboration between academic institutions and commercial enterprises, to define open specifications for building a CVP



system for any technology. Our goal is to create content and meta-data that can be used on any machine without worry that the necessary technology might not be available. Though still in their nascent stages, these specifications will build on the principles outlined throughout this session--reusable, robust XML markup of video assets, clip divisions, annotations, and playback instructions. If we are successful, we hope to end up with standards that can be used regardless of the video encoding format or delivery system. We invite suggestions and participation from the community as we move forward from the historical and current availability of customized video playback towards an approach that works for all time-based media now and in the future.

---

## References

**Hendricks, Harold.** (1993). 'Models of Interactive Videodisc Development'. *CALICO Journal*. **11.1**.

**Melby, Alan** (2004). 'The EFR (Electronic Film Review) Approach to Using Video in Education'. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004*. L. Cantoni & C. McLoughlin (ed.). Chesapeake, VA: AACE, pp. 593-597.