

# **Building a Multimedia Digital Library of Audio/Video Resources Using Open Source Software and Open Digital Library Standards**

---

M.G. Sreekumar

**Abstract:** The emerging information landscape portrays a complex domain before the library fraternity with multiplicity of physical as well as digital formats in which they manifest in diverse publication types. Libraries today operate on a truly multimedia environment. In the present academic and special library setting the educational videos, instructional visual aids and audio learning resources form a significant part of the collection. Multimedia collection, especially the non-textual knowledge resources in the library at the same time pose a variety of problems not only in the provision of their services to users, but with their sole dependency on a host of gadgets also such as media players, viewers, etc. The current information environment unequivocally prompts libraries to leverage on the latest digital technologies towards building multimedia digital libraries and in setting up dynamic electronic information systems. These newer breeds of information services offer lots of power and visibility to libraries and the user community without compromising on their quality and performance. Yet, they also pose problems in the course of their preparations while submitting them in a digital library. Swift developments in media-related technologies, rapid obsolescence and their redundancy, need for time-to-time media and format migrations, resource intensive maintenance, etc. adds to the complexity and complications of these collections. Leading edge technical skills and technology backing for these facilities demand sophisticated human resource requirements and high end IT infrastructure. Multimedia digital libraries also indeed require the strong footings of Resource Description Framework (RDF) vision supplemented with descriptive metadata standards such as Dublin Core, Metadata Object Description Schema (MODS) or Metadata Encoding and Transmission (METS). They also need the strength of Extensible Markup Language (XML) encoding schemas, related Document Type Definitions (DTDs) and Extensible Stylesheet Language (XSL) transformations between the non-traditional datastreams and the Hypertext Markup Language (HTML) front-end. This paper reports the experience of IIMK in setting up a multimedia digital library using the Open Source Greenstone digital library software.

## **Introduction**

Multimedia knowledge resources as document surrogates and as significant knowledge sources in a library needs no emphasis. In the present academic and special library setting the educational videos, instructional visual aids and audio learning resources form a significant collection. Time is fast catching up the world over that the traditional forms of collection development techniques and collection maintenance strategies need replacements with the upcoming trends in the profession. Consequently, the traditional information service options are no longer acceptable to a large majority of the users or that there is a strong demand for newer forms of services which are not so familiar to the libraries. Libraries therefore are now forced to be friendly and familiarized themselves with all relevant and current popular multimedia formats.

As technology marches past at a tremendous pace, the traditional meaning and definitions of a library's collection range also undergo a great deal of change. Interactive and multimedia learning resources are one of the most rapidly changing and exciting areas of education in the world today. The recent entrants are computer-based training (CBTs) materials/Web-based training (WBTs) materials, especially interactive multimedia programs that run on personal computers. These new technologies offer students, teachers and researchers access to materials as never before. Multimedia can deliver large amounts of information in ways that make it manageable, approachable, and useful. And by making it possible to access illustrations and photographs, sound and video, as well as large amounts of text, interactive multimedia programmes present learning information to teachers, students, and scholars in newly engaging and meaningful ways. The integration of multimedia programmes into libraries and classrooms promises not only to change the kinds of information that is available for learning, but the in which that learning takes place.

### **Multimedia Collection: Features, Problems and Prospects**

Predominantly multimedia collection of a library comprise video resources (educative as well as entertainers), audio recordings, computer-based training materials (CBTs), Web-Based Training materials (WBTs), illustrations, photographs, etc. Depending on the nature of the parent institution's goals, academic focus, research thrust, and above all, its resource allocations and financial capabilities, their collection strength may considerably vary.

In most cases, the collection-building process of the multimedia resources may follow the normal traditional mode. In the case of e-Resources the purchase may be restricted to licensing and only online access to the product may be available to the library. Publishers usually enforce stringent copyright and IPR restrictions to multimedia resources.

Multimedia collection, especially the analogue knowledge resources, poses a variety of practical and operational constraints to libraries. One of the major limitations is their sole dependency on a host of gadgets such as media players, viewers, etc. They also suffer from reach and visibility limitations as they can be viewed or listened to by a single or fewer number of viewers/listeners simultaneously. Further more, there are shortcomings with respect to resource discovery possibilities other than the basic indexing offered by their catalogues (traditional as well as digital). Collection maintenance and long-term preservation of these special collections are always a concern for librarians.

Multimedia resources in the newer breeds of media formats such as the CBTs, WBTs, digital videos/audios etc. have the added advantage of being dynamic and mobile with respect to their flexibility, portability and efficiency of access/dissemination. The current information environment also unequivocally prompts libraries to leverage on the latest digital technologies towards building multimedia digital libraries and in setting up dynamic electronic information systems. These newer information services offer lots of power and visibility to libraries and the user community without compromising on their quality and performance. Yet, they also pose a multitude of problems in the course of their preparations while submitting them in a digital library. Swift developments in media-related technologies, rapid obsolescence and their redundancy, time-to-time media and format migrations, resource intensive maintenance, etc. add up the complexity and complications of these collections. This paper attempts to share the IIMK experience in developing, implementing and maintaining the multimedia collection as a part of its digital library development activities which comprise the entire publication types, formats and media.

*Digital Information Resources and Networks on India***Open Source Philosophy of IIMK**

IIMK is an ardent proponent of the Open Access and the Open Source Software philosophies which is sweeping the world now. It was a chosen decision not to go in for a proprietary software for its digital library applications. Accordingly we evaluated some of the popular Open Source Software for digital libraries, which were in use internationally. 'Dienst', 'Eprints', 'Fedora', 'Greenstone', etc. were the candidates for the preferred software. Obviously Greenstone outscored the group and we formally adopted the software for creating the IIMK digital library (Greenstone). The Greenstone Digital Library Software (GSDL) is a top of the line and internationally renowned Open Source Software system for developing digital libraries, promoted by the New Zealand Digital Library project research group at the University of Waikato, headed by Dr. Ian H. Witten, and is sponsored by the UNESCO [Witten]. Greenstone software (versions starting from 2.30) along with Java Run Time Environment (JRE) were deployed for the purpose. The software suite is available at the open source directory 'Sourceforge' [Sourceforge].

**Digital Library at IIMK**

The primary objective of the IIMK digital library was to enhance the digital collection in a substantial way, by strategically sourcing digital materials, conforming to copyright permissions, in all possible standards/formats so that scalability and flexibility is guaranteed for the future and advanced information services and are assured to the user community right from beginning. The digital library was planned in such a way that it would integrate and aggregate the existing collections and services with an outstanding user interface. Accordingly, necessary strategies were adopted towards working out the digital library system. This implied that the digital library system should have a strong collection interface capable of embracing almost all the popular digital standards, digital formats and software platforms, in line with the underlying digital library technologies in vogue. This was crucial in the case of multimedia integration, which was again important as it was planned to host a digital audio and video library as part of the core library collection.

The Digital Library Project at IIMK took off during September 2001. The first few months were mainly devoted to infrastructure build-up such as hardware and software, digitization strategies, strategies regarding collection building, and finalization of the variety of content categories. Since the campus intranet and the Internet connectivity were robust, there were no worries on those fronts. The IIMK DL was slated to be populated with e-Books, ('born digital' as well as digitized), e-Journals, management cases, technical reports, working papers, faculty publications, teaching notes, CBTs, WBTs, educational videos/audios, value adding the existing OPAC, speeches, presentations etc. It took around two/three months for us to install and get proper command over the software. The collection comprises multiple digital formats such as HTML, XML, PDF, PS, RTF, JPG, CIF, MPEG, etc. The DL has now got a sizable collection of over 500 e-Books covering almost all areas of business and management, literature, classics, and Information Technology. Cross collection pointers were given to project Gutenberg (classics and literature) and to other GSDL example collections (social science, environmental science) and hence the virtual collection strength is over 15000 e-Books.

IIMK has plans to host the IIM Consortium (led by the six IIMs in India) digital archive and studies are underway on the feasibility of putting the IIM Consortium digital content under Greenstone (IIM Consortium). A rigorous collection development drive is being worked out with most of the popular electronic publishers specializing in business and management. One of the major bottlenecks to the speedy collection development is the non-availability of proprietary and trade literature in Greenstone

acceptable formats such as HTML, Word, or PDF. Most of these publishers put their materials in their own proprietary e-Book reader formats, from which the text extraction becomes almost impossible.

### **Resource Discovery**

Leading edge technical skills and technology backing on campus demand sophisticated human resource requirements and high-end IT infrastructure. The quality and performance of a digital library is measured in terms of its information retrieval logistics. Multimedia digital libraries also indeed require the strong footings of Resource Description Framework (RDF) vision supplemented with descriptive metadata standards such as Dublin Core, Metadata Object Description Schema (MODS) or Metadata Encoding and Transmission (METS). They also need the strength of Extensible Markup Language (XML) encoding schemas, related Document Type Definitions (DTDs) and Extensible Stylesheet Language (XSL) transformations between the non-traditional datastreams and the Hypertext Markup Language (HTML) front-end.

### **Multimedia Digital Library Workflow/Methodology**

One of the salient features of this collection is that the multimedia digital library is carved exclusively out of open source software (except for the media conversion utility) and open digital library standards. The untiring support and hard work of the IIMK's Digital Library team in the Library and the Computer Center need special mention.

#### *Collection Strategy*

The videos/audios multimedia collection was planned as a 'metadata' collection. The digital objects (video file) were stored in a separate media server so that the digital library server is not overloaded or overburdened. Hyperlinks were given to the Streaming Server where the audio/video objects were kept in JPEG format.

#### *Software and Standards*

As discussed earlier, the Greenstone Digital Library Software (GSDL) was deployed for the digital library application. "PlayTV-USB-Pro" software was used for converting the analog videos to digital files. Videos were first converted to AVI files and subsequently made to JPEG format. "Unreal Media Server" software was used for streaming the videos from the media server. "Streaming Media Player" software was given with all the retrieved pages for the users to download and stream/view the file. Each Video file was built as an HTML embedded page and XML was used for metadata encoding purposes. "Dublin Core" was the metadata standard used.

#### *Dublin Core Metadata and XML Tagging*

To create indexes for section and sub-section (document hierarchy and structuring) in Greenstone the document should be in HTML format. Obviously now the Source File has to be edited as a HTML file structure. For the section and sub-sections, the source file has to be edited appropriately, giving XML tags passed as comments in the body of the HTML file.

Unqualified version of Dublin Core Metadata standard was deployed for resource description of the multimedia digital library. The Section Tagging pattern used for the video collection is given in Figure 1.

```

<Section> <Description>

<Metadata name="Title"><a HREF="ums:\\TCP:streamserver:5119\video\V83.mpg">l. 20
  Steps to better management : Solving problems and thinking creatively(NEIF)
  V201 </a></Metadata>
<Metadata name="Subject" mode="accumulate"><font size=1>
  Creativity</font></Metadata>
<Metadata name="Subject" mode="accumulate"><font size=1 >Strategic
  Management</font></Metadata>
<Metadata name="Subject" mode="accumulate"><font size=1>
  Strategy</font></Metadata>
<Metadata name="Subject" mode="accumulate"><font size=1> Innov
  ation</font></Metadata>
<Metadata name="Subject" mode="accumulate"><font size=1 ></font></Metadata> <Metadata
  name="Publisher" mode="accumulate"><font size=1> NEIF</font></Metadata>
<Metadata name="Resource Identifier" mode="accumulate"><font size=1> V 201
  </font></Metadata>
<Metadata name="Source" mode="accumulate"><font size=1>
  </font>NEIF</Metadata>

</Description> </Section>

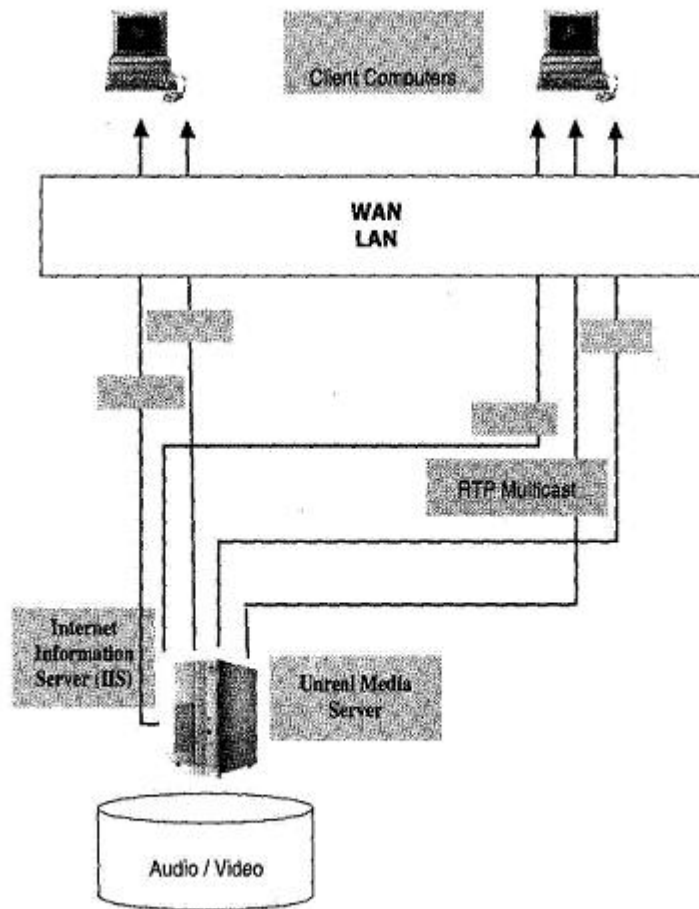
```

**Figure 1:** XML based Section Tagging of Dublin Core Metadata

Unreal Media Server is an open source server software which ensures efficient multimedia deliver over the network, LAN as well as WAN. UMediaServer is a streaming server for Windows operatin systems. Unreal scores high over its proprietary counterparts such as Microsoft and Real, with it efficiency in streaming high quality content.

### System Architecture

The Unreal Media Server software is downloaded from "http://www.umediaserver.net/bin UMediaServer.zip", unzipped, installed and configured in the Windows operating environmen Installation and customization tips are available in the software site and it is encouraged to follm them. Transmission Control Protocol (TCP), Hypertext Transfer Protocol (HTTP) and Real Transpo] Protocol (RTP) and Multicast transport protocols are used for streaming content from Media Server t clients. Incoming content is not stored on the client computer's hard disk and the user is not allowe to save media locally, thus author's rights are fully protected. User authentication and access restriction enable trusted access to media resources. User logging provides the means to track user activity. Fc the client workstations, Unreal Media Player client software was provided. The System Architectur of the multimedia digital library is given in Figure 2.



**Figure 2: System Architecture of the Multimedia Digital Library**

### Creating Links to Media Resource

While setting up the Streaming Media Player it registers a custom URL protocol on the user's machine. This is a UMS protocol, allowing launching Streaming Media Player directly from the hyperlink on the web page. Though the most comfortable way for the users to access remote media is to click on a link in the Web browser, for this application we gave the hyperlink embedded in the Dublin Core metadata element 'title' as follows:

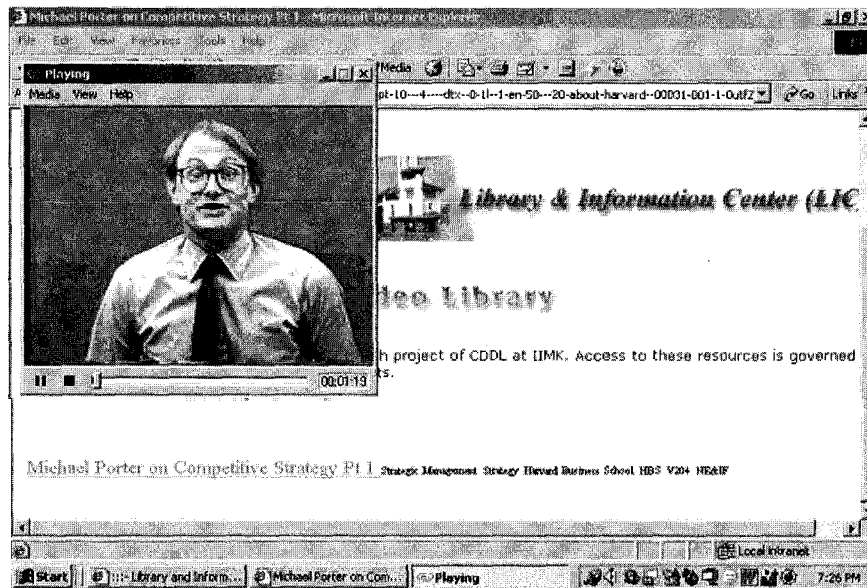
```
<Metadata name="Title"><a HREF="ums: \\ TCP: streamserver:5119\video \V83.mpg">1. 20 Steps to better management: Solving problems and thinking creatively (NEIF) V201</a></Metadata>
```

In the above Metadata Tag, 'UMS' is the custom protocol of the UMediaServer Software. Streaming is achieved using the 'TCP' protocol. 'Streamserver' is the media server holding the multimedia digital objects. TCP's default port '5119' is the port utilized. For HTTP, default 80 is the port specified. Using Media Server Configuration tool, you can create virtual folders and map them to the different physical locations on the server machine. Alternatively, the physical location can be anywhere in the server machine's LAN. 'Video' is the folder where the digital videos are located and 'V83.mpg' is the object being streamed to the client.

*Digital Information Resources and Networks on India*

**Access Methods**

The multimedia collection is one of the important and appreciated services of the IIMK Library Portal. Users are able to visit the multimedia digital library and select the videos/audios of their choice by browsing alphabetically or by searching the author, title/free text, keyword, source or publisher. The Dublin Core metadata description facilitates efficient subject approach to information retrieval of the important resources with fanciful titles too, such as 'Eye of the Beholder', 'Just in Time' etc. Retrieval of cross collection resources is another special feature which is seldom achieved through traditional methods. A sample query frame of the streaming digital video is shown in Figure 3.



**Figure 3:** Video Digital Library through Streaming Server

The digital library at IIMI has been able to strike great impact amongst the user community on campus. The academic quality of the resources coupled with flexibility, convenience and efficiency of access across the entire campus are the major reasons for this scenario. User education and product orientation sessions held at regular intervals by the library, both online as well as live, have paid rich dividends in the increased awareness and resource utility. The usage and usability of the digital resources have increased extensively as per server logs and monitoring statistics. The multimedia collection now streams out over hundreds of world-class educational videos, as well as audio resources. The distinction of this work is that it has used only open source digital library softwares and open digital library standards across the entire workflow and operations. This has significant impact and implications as the proprietary softwares and standards are exorbitantly expensive and beyond the reach of most of the academic and special libraries in developing countries. A mirror site of the IIMK digital library is available at "<http://intranet.iimk.ac.in/cgi-bin/library>" [IIMK Digital Library], as

an example collection of Greenstone digital library software (<http://www.greenstone.org/cgi-bin/library?e=p-en-home-utfZz-8&a=p&p=examples>).

### Acknowledgement

The author acknowledges his thanks to the Director, Indian Institute of Management Kozhikode (IIMK) for his encouragement and support. Thanks are also extended to the Greenstone Software team at Waikato University, New Zealand, and to the Digital Library team at the IIMK.

### References

1. Greenstone <<http://www.greenstone.org>>
2. Witten, Ian H. et al. 2001. Greenstone: Open-Source Digital Library Software *D-Lib Magazine*, 7 (10): 1-16.
3. Witten, Ian H. 2003. Examples of Practical Digital Libraries : Collections Built Internationally Using Greenstone *D-Lib Magazine* 9 (3): 1-15.
4. SourceForge.net (world's largest Open Source software development website) <<http://www.sourceforge.net/>>
5. RDF. <<http://www.w3.org/RDF/>>
6. XML. <<http://www.w3.org/XML/>>
7. METS. <<http://www.loc.gov/standards/mets/>>
8. MODS. <<http://www.loc.gov/standards/mods/>>
9. IIM Consortium. <<http://intranet.iimk.ac.in/libintra/iimconsortium.htm>>
10. UMediaServer <<http://www.umediaserver.net/bin/UMediaServer.zip>>
11. IIMK Digital Library <<http://intranet.iimk.ac.in/cgi-bin/library>>