Developing Standards and Practices for Archiving Multimedia Material

Jasmin Rivers, Illiana Thomas, Tiwana Walton
1704 Weeksville Road, Box 672
Elizabeth City, North Carolina 27909

***Abstract*-The Center for Excellence in Remote Sensing Education and Research (CERSER) at Elizabeth City State University is an umbrella for several funded projects. These projects include: The Northeastern Chapter of the IEEE-GRSS Society, The Center for Remote Sensing of Ice Sheets (CReSIS), NOAA's Educational Partnership Program at ECSU, The Undergraduate Research Experience in Ocean and Marine Science Program (URE OMS), as well as the Undergraduate Research Experience Program supporting undergraduate research at ECSU. These programs entail numerous special events that generate both video and photographic images. Reports and other documentation are also created to record the success and progress of each program.**

A standard method of digitizing images, movies, and documents and documenting the attributes of these digitized files was developed during this project. Standard procedures utilizing the equipment currently in place were developed, tested, and refined in order to convert VHS formatted tapes and developed photographs into digital files. Software such as CapWiz and Hewlett-Packard flatbed scanning software were used to complete this process. These files were then annotated in a Microsoft Access database utilizing the Dublin Core Element Set as a standard. A combination of Microsoft Access and Active Server pages was then utilized to provide access to the digitized images and the attached metadata via the web. This paper will discuss the digitizing procedures developed and the metadata elements used to build the URE Multimedia Library.

I. Introduction

A. Overview

The Center of Excellence in Remote Sensing Education and Research (CERSER) website currently maintains a library of images and videos utilizing the Multimedia Server (http://mmt.cs.ecsu.edu/nia_photo_library/). This library is not catalogued and gives very little information on the files within. While this may be suitable for storing large amounts of information, it makes retrieving the needed material at a given time very cumbersome and time consuming.

This year's Multimedia team took on the task of establishing procedures for digitizing images not currently in the library and documenting the contents of the library utilizing an established set of metadata standards and practices. Digitizing was completed utilizing hardware and software currently in place within the CERSER labs while documentation was completed with Active Server Pages and Microsoft Access.

B. Digital Library

A "digital library" is a collection of files that are stored and accessed electronically. Files that are stored inside the library should have a common topic for all of their data. Digital libraries can be designed for computer graphics, operating systems and/or other networks. All of these libraries can be combined into one common interface dealing with computers. However it is important that the information contained in each library remain separate. The purpose of a digital library is to make available a central location for accessing information on specific topics. It makes searching faster and more convenient for users.

C. Digital Library Examples

There are many digital libraries three of which are: The Everglade Information Network & Digital Library, The University of Washington Digital Collections, and AmericanSouth.org.

The Everglade Information Network & Digital Library stores information relating to the South Florida environment. This includes collections of scientific and technical reports, natural history writings, datasets, photos and a directory to other sites relating to the Everglades. This digital library can be found at: http://everglades.fiu.edu/index.htm

The University of Washington Digital Collections is primarily composed of pictorials that accompany essays, audio and video clips, newspaper reports and maps. Within the library you can browse the collections and search easily by creating a (phrase) to look up a specific essay or picture. This digital library can be found at: http://content.lib.washington.edu/index.html.

The AmericanSouth.org library contains resources that document the history and culture of the American South. This project was originally started with a grant from the Southeastern Universities Research Association (SURA) and is currently being funded by a grant from The Andrew W. Mellon Foundation. It is currently being developed and maintained by Emory University. The URL can be found at: http://www.americansouth.org.

D. File Types

The URE Multimedia Library currently documents several file types depending on the type of media entered. For still image files, the Joint Photographic Experts Group format commonly referred to as JPEG was used. The name comes from the group who created the standard. Audio Video

Interleave (AVI) and Windows Media Video (WMV) formats were used with movie files. Text files were converted to Portable Document Format (PDF) in order to retain the formatting used in each document.

E. Building on Past Research

The 2006-2007 Multimedia Team utilized the research from the 2005-2006 Multimedia Team research to enhance this project. Research on database use, Active Server Pages, and JavaScript programming was taken from past research and applied to the library ontology. This allowed the team to concentrate on the structure of the library versus the programming aspects.

II. Digitizing Procedures

A. Introduction

Digitizing involves the conversion of printed material or analog video to a digital format. Various hardware components and software applications were used for these digitizing procedures. Throughout the project several procedures failed and hardware was found to be no longer supported by manufacturers.

B. Hardware

ADS Technology's DVD Xpress DX2 Video Converter allowed the team to transfer video from VHS tapes, digital or analog, into a digitized format. The video converter provides high quality video with smaller file sizes. This hardware captures directly to video formats including DVD (MPEG-2), SVCD, VCD (MPEG-1), MPEG-4, and DivX, all through the encoder inside the DVD Xpress DX2. More information about the DVD Xpress DX2 Converter can be found at: http://www.adstech.com/.

Hewlett-Packard's HP Scanjet 5530 Photosmart Scanner was used to convert packs of 24 photographs (3x5 or 4x6) in less than five minutes into digital images. The scanner's Automatic Photo Feeder (APF) allows the user to scan these multiple photographs. The scanner also has the ability to scan 35mm slides, transparencies, negatives, and 3D objects. The image resolution is 2400x4800 dots per inch (dpi) with a 48-bit color depth. These photos can be scanned into a photo album or email format. More information about the HP Scanjet 5530 Photosmart Scanner can be found at: http://h10010.www1.hp.com/wwpc/uk/en/sm. Panasonic PV-D4745 Combination DVD/CD player + HiFi VCR is used to play VHS tapes. It inputs analog video and audio through RCA plugs into the DVD Xpress Converter which outputs through USB cords digital audio and video to be captured by the CapWiz software.

C. Software

ADS Technology's Capture Wizard Software was used to capture video. It works in conjunction with DVD Xpress and the VCR to capture video and export it. The captured files were combined and edited using Windows Movie Maker and converted to WMV files. The software utilized for scanning photographs in this project was the Microsoft Scanner and Camera Wizard. This program was part of the Microsoft Windows XP operating system, which was used on the Gateway Laptop and is automatically called up by the scanner.

D. Video Conversion

The DVD Xpress Capture Wizard software was used to convert our video from VHS to digital through the following steps:

1) Open the Capture Wizard Software by START> PROGRAMS> ADS TECH DVD XPRESS> CAPWIZ
2) Choose the worldwide video standard based on country- North America (NTSC) then select NEXT
3) Select Video Input: RCA or S-VHS Video then select NEXT
4) Create file name and select file location then select NEXT
5) Choose Video Format
6) Select FINISH
7) Press PLAY on VCR
8) Click the PREVIEW button on Cap Wiz to begin previewing the video
9) Click the RED DOT (on Cap Wiz) when you are ready to RECORD video to your computer.

E. Photograph Scanning

The following steps were used for scanning photos taken by a 35mm camera into a digital format:

1) Turn the scanner on.
2) Place a maximum of 24 photos into the APF, face down
3) Press SCAN PICTURE button on the scanner
4) Select the application to scan the photos, Microsoft Scanner and Camera Wizard
5) Enter the wizard and select Next
6) Select CUSTOM SETTING (default is 100 dpi, optimum is 300 dpi)
7) Select Color Picture 300 resolution (dpi) then select OK
8) For Paper Source select DOCUMENT FEEDER then select Next
9) Page size should be A6 and size of the picture (4.1 x 5.8) then select Next
10) Enter a name for the pictures
11) Select a File format (JPG)
12) Choose the location for the saved images and select Next
13) Browse scanned images
14) Scanner will now process photos
15) Once the scan is complete, select "Nothing" and Next
16) The scanner will display how many images were copied and provide a link to their location.

III. Metadata

A. Defined

Metadata can be simply defined as data about data. It can be used to describe various types of data from images in a digital camera to the fields found in a mailing address. An example can be the terms 37920, 070422, and 8155744559. All three have no value until you know that 37920 is a zip code, 070422 is the date April 22, 2007, and 8155744559 is a telephone number with the area code. This is an example of data describing data.

Metadata in a computer file can be the author, creation date, modification date, file size, or file type. Metadata in a digital camera could be flash on/off, f-stop setting, macro on/off, or creation date.

B. The Dublin Core Metadata Initiative (DCMI)

The Dublin Core Metadata Initiative (DCMI) is a group promoting the adoption of transferable metadata standards and developing dedicated metadata vocabularies for describing resources that allow for more intelligent information retrieval systems.

The DCMI offers simple standards to make possible the finding, sharing, and management of information. DCMI develops and maintains international standards for describing resources while supporting a worldwide community of users and developers. It also promotes the use of Dublin Core solutions. The development and maintenance of a core set of metadata terms labeled the Dublin Core Element Set is the main activity of DCMI.

C. Dublin Core Element Set

The Dublin Core Element Set is a word list of fifteen properties for use in describing resources. The term was coined at a workshop in Dublin, Ohio in 1995. The term "Core" is used as the fifteen elements used are broad and generic which allows them to be used in describing an extensive range of resources. The Dublin Core is only part of a larger set of descriptors and specifications maintained by the DCMI.

D. Title

The Title element is the formal name for the resource. Examples of formal Titles are "The Mona Lisa", "The International Polar Years 2007-2009 Report on U.S. Federal Agencies' Planning", and "Meridian Uninhabited Air Vehicle."

E. Description

A resource can be described in several different ways. It could be an abstract, an image, a table, or simply text about the resource.

F. Contributor

The Contributor element is the person, organization, or other body responsible for making contributions to the resource. Some examples are an individual such as Dr. Tynum Drake, or possibly an organization such as Spelman College.

G. Coverage

The Coverage element includes two possibilities as broad descriptors: time and location. Time can encompass a set date or a range of dates. Location could be geographical coordinates or it could be the proper name of a location. The current practice is to use the Thesaurus of Geographic Names [TGN] when naming a location. The more proper scientific terms to be used in this element are Spatial (geographic) and Temporal (time).

H. Creator

This element is close to the element Contributor, but the distinction is that this entity actually makes the resource. Once again, the creator can be a person, organization, or other body responsible for making the resource.

I. Date

Date is a point in time that can be associated with the resource. This can be at various levels of detail such as just the year or down to the tenth of a second. Formats for this element are referenced to the World Wide Web Consortium (W3C) guidelines, which can be found at: http://www.w3.org/TR/NOTE-datetime. Examples of this format are:

Year: YYYY (e.g. 2005)
Year and month: YYYY-MM (e.g. 2005-07)
Complete date: YYYY-MM-DD (e.g. 2005-07-16)

J. Format

The Format element is the physical medium, or dimensions of the resource. Examples of this element include size and duration or the file type such as JPEG or AVI.

K. Identifier

The Identifier element is a string of characters or numbers given to a resource to identify it and only it. An example of this would be the International Standard Book Number (ISBN) given to published works.

L. Language

The Language element is used to note the language of the resource. A listing of languages and their two letter identifiers can be found at http://www.iana.org/assignments/language-subtag-registry.

M. Publisher

The Publisher element is the individual or organization that makes the resource available. Examples of this would be a publishing company, an individual, or a university.

N. Relation

The Relation element indicates how and to what resource the current resource is related. The specific sub-elements of Relation are Relation.Identifier and Relation.Type. Relation.Identifier holds the metadata of the related resource and Relation.Type describes the relation.

O. Rights

The Rights element describes the intellectual and property rights of the resource. This could be in the form of a copyright or just a general statement.

P. Source

Source is used to note where the resource was taken from. This could be utilized when the resource is the particular page of a book or a short clip of a full length movie. It is recommended that a formal naming convention such as an ISBN be used for this element.

Q. Subject

This element is used to describe the topic of the resource utilizing keywords, phrases or classification codes. It is best to not use the time or location in this element, as the Coverage element will hold that information.

R. Type

The Type element is used to describe the resources type according to a set of terms set for by the DCMI. This is not the element to be used for the file type, size, or dimensions as they should be recorded in the Format element. A list of DCMI Type terms can be found on the DCMI website at: http://dublincore.org/documents/dcmi-type-vocabulary/.

IV. Results

A. Accessing the Library

The URE Multimedia Library was placed on the Center of Excellence in Remote Sensing Education and Research (CERSER) server. The URL for the start page is http://cerser.ecsu.edu/mmt0607/ure_mm.asp and a link can be found on the NIA server at http://nia.ecsu.edu.

The library was placed on the CERSER server as it currently runs the Windows IIS server which runs Active Server Pages (ASP) natively. The CERSER server currently runs the CERSER satellite image archive which made it a tested platform for the development of the multimedia library.

B. Active Server Pages

ASP pages were developed by Microsoft and allow a developer to build interactive and dynamic (changing) web pages. In this projects case, the ASP allowed the team the ability to interact with a database containing the elements of the metadata pertaining to each entry into the library. The knowledge of ASP used in this project was gained through the 2005-2006 Multimedia Team project. For a more in-depth explanation of ASP pages and their uses see the 2005-2006 Multimedia Team project located at http://nia.ecsu.edu/onr/05-06/researchteams/mmt/.

C. Main Page

The main page, or start page of the multimedia library (ure_mm.asp) allows the user to immediately enter a search parameter, browse the entire library, or view collections of images previously established. This page was developed with the intent of proving the concepts and refining the interaction between the user and the library. The "Browse All Entries >>>" link returns the entire content of the library ordered by the date element.

D. Searching

The library search is a simple single word or phrase search. Multiple terms such as "CWM AND Bindschadler" that are not found in a phrase will yield a null return. This entry searches the subject, description, and format fields of the database and returns files in a results window (results.asp). The results are displayed in the order contained in their Date element. Currently there are no means of reordering the results page.

E. Collections

Collections are common components of established multimedia libraries, which allow a user to access an analogous group of files quickly. The library currently contains three collections: Celebration of Women in Math, Distinguished Lecture Series (IEEE-GRSS and CERSER), and Undergraduate Research Experience Team Reports.

Each of these collections uses a keyword search applied to the results page to return a group of files associated with that topic. For example, The Celebration of Women in Math collection link is coded to send the search phrase "Celebration of Women in Math" to the database. This phrase should always be the entry for the subject element in the database when a file is associated with this event.

As more files are added to the library, more collections should become available to the user. This determination will be made by the library manager and the program director.

F. Details Page

Once the search results are displayed, each file will have a "Details" link below it. This link opens the page ure_mm_lib_details.asp in a new browser. This file accesses the database to retrieve the detailed information for that file. In the case of an image file, a thumbnail is displayed of the file on the left side of this page. The location of this thumbnail is part of the record in the database along with the location of the original file. Files other than images will use icons of their file types such as a PDF or a movie icon. Below this thumbnail or icon will be a link to the original file which will open with the associated software for the computer the user is currently accessing the library from.

The main feature of the details page is the element list containing the description of this entry. The fifteen DCMI elements are listed here for the record being accessed. In Windows Internet Explorer a description of the element can be found by moving the curser over the magnifying glass located next to the element name. The elements listed may be pared down as the library grows and it is found that they are not used.

The final feature on the details page is the "Close Details Window" link which returns the results page to the forefront of the screen.

V. Future work

A. Add Library Items

Among the future work needed to sustain the multimedia library will be the continuous addition of files. These will include videos and photographs of events that have taken place in past years as well as current events. The need to be thorough and document the language utilized is paramount to the continued growth and use of this library. The database entries should be codified so that future developers can continue in a consistent manner to add entries and ensure that they meet the standards set down by the DCMI.

The Dublin Core Elements currently displayed in the details window of the library should be tracked to record usage. Refinements should be made in the library to reflect elements used and possibly those elements that should be added to ensure an accurate reflection of the metadata needed. Those elements not utilized should be removed in order to eliminate extraneous fields in the database.

B. Expand Field Searches

The current search engine explores only the subject, description, and format fields when a search parameter is entered. Future work should also include an advanced search capable of searching all fields that are contained within the database. This addition could be developed in a short manner of time to increase the capabilities of the library.

Another refinement would be the ability to reorder the search results by a chosen element such as date, reverse date, titles, and format. This would allow the user to interact with the database in a more efficient manner in obtaining the file needed.

C. Convert Videos to MPG

The current videos are in AVI or WMF file formats leading to large file sizes. As the library is added to, a procedure for converting the files to MPG, the standard web format, should be developed. This could entail the use of Adobe Premier or other video editing software capable of converting the original analog video. A backup digital file in DVD format should also be created to preserve the original content of the event without reduction in physical size.

REFERENCES

[1] "Active Server Pages 3.0", maranGraphics, Inc. (2000)

[2] "Design and Implementation of a Digital Library", Retrieved October 5, 2006 from the World Wide Web, http://www.acm.org/crossroads/xrds5-2/diglib.html

[3] "Dublin Core", Retrieved November 3, 2006 from the World Wide Web, - http://dublincore.org/

[4] "World Wide Web Consortium", Retrieved November 3, 2006 from the World Wide Web (W3C), - http://www.w3.org

[5] "Sussex-Lisbon Area Historical Society, Inc", Retrieved November 17, 2006 from the World Wide Web, http://www.slahs.org/dublin_core.htm

[6] "Database Guide", Retrieved from the World Wide Web November 17, 2006 http://www.geekgirls.com/menu_databases.htm

[7] 2005-2006 Multimedia Team Project, Retrieved January 8, 2007 from the World Wide Web, http://nia.ecsu.edu/onr/05-06/researchteams/mmt/