Automated Workflow for Archiving Video Art in DVD Media

Maria Esteva, Karla Vega; Texas Advanced Computing Center, The University of Texas at Austin; Austin, TX; Bethany Scott; Charlotte Mecklenburg Library; Charlotte, NC; Summer Gunnels; Texas Advanced Computing Center, The University of Texas at Austin; Austin, TX; Keerthana Kumar; Department of Computer Science, The University of Texas at Austin; Austin, TX

Abstract

From 2000 to 2010, the great majority of artists submitting video art work to the Blanton Museum of Art at the University of Texas at Austin did it in DVD format. To archive these works in an HPC storage resource where they will be actively managed and preserved, the DVDs have to be converted into files that can fulfill both preservation and access functions. A survey of the technical lineage of the pieces as well as best practices in digital and in video preservation were important to decide the configuration of an automated workflow and the components of a SIP to submit the video files to storage. Quality video metrics were used to achieve a losseless conversion of the DVDs and were included in the workflow to verify the integrity of the conversions. Metadata about the files is gathered to dynamically customize the conversion processes according to the characteristics of the input DVD. Process information and technical metadata are recorded in a METS document that forms part of the SIP. Built with open source software, the workflow is integrated to the iRODS implementation of the storage resource as part of the ingest process.

Introduction

We present an automated workflow that produces a submission information package (SIP) for archiving video art acquired originally in DVD format. For each DVD submitted to the workflow, the resultant SIP contains an ISO disk image, production quality and access files, and a METS record with technical metadata about the files and about the conversion processes involved. The project was elicited by the need of the Blanton Museum of Art at the University of Texas at Austin to find a file-based preservation solution to archive their video art collection in a high performance storage (HPC) resource at the Texas Advanced Computing Center.

DVDs are an optical storage medium in which video data is encoded in MPEG 2 lossy format [1]. Popular due to their high storage capacity and portability, a little more than a decade after their release in the early 2000s DVDs are becoming obsolete. Currently, the preference in cultural institutions is for file-based preservation and access strategies within high-density storage resources that enable centralized management, replication, and diverse modes of distribution of data [2]. As institutions move in this model of digital archiving, they need to make decisions about legacy DVDs.

We started by studying the options for conversion of video art that had been compressed into DVD media. A survey of the technical lineage of the works of art explained the choice of DVD in context with the time of creation and acquisition of the videos. While some works were acquired in other formats, the majority of the videos in the collection had been delivered by the artists in DVDs labeled as original or master. Considering those DVDs as the input we created a workflow to generate files that would fulfill preservation and access functions required in the museum. The workflow was built entirely around open source libraries and formats. The design was also influenced by the large storage capacity and the configuration of the HPC resource in which the collection will be managed and preserved.

The experience highlighted the importance of metadata and of video quality control methods to achieve the best possible transcoding of the video and audio components of the works. While we created a generic workflow, we do not use default settings. Instead, the conversion process adjusts dynamically in relation to existing parameters in the original DVD. We also introduced possibilities for user customization of the number and kind of digital conversions. In our design and implementation development we assumed that the museum holds full rights for preservation, education, and display for the DVDs in the collection. However, curators and registrars at the museum will have to determine on a case-by-case basis which works will be submitted to the workflow to comply with intellectual property rights and permissions.

Survey of the Video Art Collection

We conducted a survey of the video art collection by inspecting the museums provenance files. We recorded how each work was originally created and any reformatting already done; formats and storage media owned by the museum; and availability of artists statements with instructions regarding the works technical characteristics, intentions for the display of the piece, and authorization for future conversion. Data points were recorded and organized by decades for analysis. Table 1 shows the results of the survey, which records only the original media owned by the Blanton.

The museum has a collection of thirty three video art works; of these, twenty three were acquired in DVD format. For ten of the pieces, the collection includes other original storage media such as VHS, U-matic, Betacam, reel-to-reel videotape or as files on computer hard-drives. For each work, the Blanton has at least one DVD exhibition copy that has been either provided by the artist or was created at the museum to fit their exhibition equipment. Both creation and accessioning dates are important to understand the medium type in which the works were acquired. We learned that with the exception of two, all the works were acquired in the last decade and almost all in DVD format.

The survey revealed the different levels of technical awareness of the artists in relation to their work. Of the ten videos acquired in formats other than DVD, two artists left instructions regarding future reformatting. Anima by Bill Viola is a color video triptych acquired as a set of three digital Betacam tapes and a set of DVDs for exhibition purposes. In the instructions left to the museum, Viola refers to the DVDs as a temporary platform for the work while the Betacam masters represent the essential form of the piece. He further instructs that the quality and aspect ratio of any future versions should match that of the master tapes. According to provenance records Liliana Porters Drum Solo of 2000, sold as number 53 of a series of 100 in 2003, was created in 2000 in 16mm film and transferred to digital video and to VHS. The latter is the format in which the piece was acquired by the museum. The museum has documentation about the works production, including the instructions sent to the company that produced the film and invoices from the one that transferred the film to digital video. Some years later the artists sent a DVD to the museum for exhibition purposes with the clarification that the VHS was to be considered the master from which other copies could be made when needed. Artist Jeremy Blake showed technical awareness in relation to the DVDs he produced for his Winchester Trilogy. The artist describes the three works as time-based paintings containing digital animations and mash-ups of retouched images. The master DVDs are signed, dated, and numbered and come with a corresponding viewing disc. The Blantons files contain instructions on displaying the piece in 16:9 aspect ratio with a 42-inch plasma screen. The artist also indicated that this aspect ratio should be preserved in any future conversions.

We found technical information in the exhibit catalogue and clippings kept in the provenance files. The importance of sound in the video Film de Film by Jorge Macchi and Edgardo Rudinsky is highlighted by art historian Andrea Giunta in the essay Jorge Macchi: The Anatomy of Melancholy in which she wrote, In Film de Film, the closing credits from several films rice on the screen, out of focus and illegible. The text is an image that interprets the sound: each line of text appears at the bottom edge of the screen in correspondence with the tempo of the music. The lines of text establish the moment of the notes, a new writing system for music [3]. Created in 2007, the video was acquired in 2008 in DVD with stereo sound. This text is informative in relation to aspects of the original DVD such as number of audio channels that needed to be considered during conversion.

None of the video works in the collection contained the technical documentation that is currently considered standard for digital preservation purposes. This is changing as the museum has recently adapted The Documentation and Conservation of the Media Arts Heritage (DOCAM) and Matters in Media Art documentation models and templates and is updating their own documentation [4][5]. The policies for future acquisitions are also changing to request high quality masters in media independent sustainable formats.

Our observations also included how the DVDs are used at the museum. DVDs are stored in the storage vaults under climate-controlled conditions. When needed for exhibits or educational purposes the exhibition copies are carried and handled only by authorized users to the playback device in the exhibit room or the classroom. In considering their conversion to files and the ways in which they can be accessed within and outside the new storage environment, there are authenticity, integrity and security measures that must be implemented.

The majority of the video collection was acquired in the last decade, coinciding with the prime time of DVD usage. Thus, the DVD was a natural choice for most artists to deliver their work. Only one master DVD was sent to the museum by artist Alejandro Paz of his video El Guardaespaldas from 2003; and the same is true for eight more works. Nine artists submitted a master DVD along with an exhibit copy in consideration to the tear and wear of the media. While for the majority of the videos in the collection the DVD is the second format iteration, even artists that sold pieces created thirty years earlier in magnetic tape sent DVDs as masters to the museum. While the DVD is a convenient format for storage and display with the Blantons current equipment, the video quality and file size are limited compared with high-definition video now possible with high-density storage like TACCs resources. The possibility to observe the collection as a whole was important to understand the role of media at different times and specifically the relevance of the DVD in the collection.

Workflow Requirements

After studying the video collection and literature on digital preservation best practices, we determined that converting the video that was only on DVD format to media independent files would help mitigate risks of disc error and data loss, and allow for easier management of the collection. Based on the survey results we established the following guidelines for the transcoding workflow:

- The workflow will only be implemented for works that the Blanton owns in DVD only. The workflow does not address videos stored in other analog and digital media.
- Preservation and access files in the SIP must fulfill education, exhibition, and promotion functions.
- The workflow should be flexible to allow changes in the selection of file formats and number of steps.
- The DVD will be preserved as a disk image.
- Transcoding should not involve further compression and should allow preserving original parameters such as aspect ratio, video and audio channels, and number of streams.
- The workflow should include mechanisms for subjective and objective video quality control.
- Checksums will be generated from all the components of the SIP to establish the objects authenticity.
- Metadata and provenance documentation should be extracted and generated automatically.
- The transcoding processes and results of quality control assessment should be documented for clear identification of our intervention in the technical history of the videos.
- The workflow should be implementable in the storage system.
- The workflow should be fully automated and easy to implement.

Video and Digital Preservation as a Roadmap

The workflow was designed using best practices in video and digital preservation as a roadmap. We adapted recommendations from different sources, considering the characteristics of DVDs containing video art as the input source.

The DVD is an optical disc storage and access format widely used in the last twelve years for access and delivery of audiovisual media. Its life expectancy has been studied, and despite different manufacturing qualities, DVDs do not constitute a long-term archival storage solution due to their vulnerability and risk of obsolescence [1] [6]. DVDs are also not recommended for digital art

Table 1. Survey of the acquisition media of the video collection distributed by decade.

Dates of creation	Works acquired	Reel-to-reel tape	U-Matic	VHS	Digibeta	Hard Drive	DVD	Sold as series	Exhibition copy
1960 - 1969				1					
1970 - 1979	2	1	1	1		1	2		
1980 - 1989				1					
1990 - 1999					1		3	4	
2000 - 2009	31			1	2	1	18	10	9
Total	33	1	1	3	3	2	23	14	9

preservation due to their high compression rate [7]. In addition, from a digital assets management perspective, DVDs physicality is inconvenient in contrast to the benefits of file-based archival solutions now used in cultural heritage and commercial settings [8] [9]. These considerations justify moving the media-dependent collection to a file based system. With this framework, we considered the kinds of digital files that would be part of the SIP.

ISO 9660 disk image is a standard, uncompressed file, considered archival because it constitutes an identical copy of the content and structure of an optical disc media. Among many preservation projects that use disk-imaging technologies, we highlight the Tangible Media Project and the Veterans Project at the Library of Congress. The first one captures the contents of CD-only historical material, as ISO images, and the second captures oral histories submitted in DVD format [10] [11]. Like our project at the Blanton, the ultimate goal of the Tangible Media Project is to create a generic workflow for the aggregation, bagging, transfer, inventory, access, verification and long-term storage of digital materials that can be used by multiple Library divisions [10]. Their workflow captures the CDs as ISO disk images and as individual files. In our project we capture a disk image of the DVD and use it as the source for transcoding to preservation and access formats.

Most references in the literature address best practices for video reformatting projects so we started by looking into those to map our workflow. To decide what files would go in the SIP we modeled the format options from the Stanford Media Preservation Lab [12]. The ISO image of the DVD is the equivalent to SMPLs preservation master. To generate the highest possible quality access file, we chose the same path as the Vancouver City Archives, a Matroska container with FFV1 codec to create a lossless conversion with respect to the lossy MPEG2 [13]. Optionally, the workflow can generate online access and streaming files, in which case a lossy codec would be acceptable for the transcoding process. The choice of wrappers and codecs in different institutions is decided by emerging standards and usage needs. We are aware of recommendations to use Material eXchange Format (MXF) wrapper and JPG2000 codec [14]. However, our choice of open source software for transcoding is a starting point for creating other file formats. Like MXF, the Matroska container allows encoding of audio and any accompanying materials, and both are good candidates for this work. Matroska is an open standard free container format for general use, while MXF is a Society of Motion Picture and Television Engineers (SMPTE) standard container format that is typically used in professional digital video and audio media. We expect to provide the MXF and JPG2000 combination in the future.

Visual quality control is routinely used to determine the differences between the analog original and the digital file in digitization projects. In the same manner, we use a full reference video quality metric, which assesses the quality of a transcoded video using the original video as reference, to assess the quality and successful completion of the transcoding step. The metric selected is the Structure Similarity (SSIM) index [15] [16]. The DVDs are already a highly compressed format, the video quality metric allows us to determine if the conversion was lossy or lossless by providing a measurement of quality between the original and the converted file.

Metadata

During the workflow we gather two types of metadata. Technical metadata is extracted from the preservation master disk image and the reformatted files using MediaInfo [17]. Provenance metadata, containing information about the kinds of conversion and their outcomes as results of the video quality control metrics, are generated by the workflow script as each step begins and ends. All metadata is encoded in a METS record that adheres to the video profile developed by the University of Texas Libraries [18]. Provenance metadata, such as the type of conversion, is encoded as PREMIS events as shown in Table 2 below.

Table 2. Mapping of conversion steps to PREMIS events.

PREMIS Element	Metadata
Event Type	Disk imaging
Event date/time	Time stamp
Event detail	ISO966
Event type	Conversion_production
Event date/time	Time stamp
Event detail	Matroska/FFV1
Event outcome information	Average SSIM
Event outcome information	Standard Deviation
Event type	Conversion_streaming
Event date/time	Time stamp
Event detail	MPG4
Event outcome information	Average SSIM
Event outcome information	Standard Deviation

During the implementation phase we learned that certain technical metadata elements were needed for the conversion process. For example, the number of audio streams is an important element to consider for the integrity of the audio transcoding. Our implementation uses this information dynamically in the transcoding process to correctly map all the streams from the input file to the production quality access file. Otherwise, the transcoder would only encode the best stream of each kind, resulting in loss of data. Bit rate information is also required, as it represent a measurement of video quality that should be specified in the conversion process.

Workflow Step-by-Step

The following steps convert the input DVD to preservation and access files. Metadata and checksums resulting from the workflow are incorporated in the final SIP.

1. Metadata is extracted from the input DVD. Technical metadata includes the container, frame size, frame rate, overall bit rate, duration, aspect ratio, video and audio format and types and number of streams. Through this step the script verifies the exact technical characteristics of the content so that these are not changed during the transcoding process.
2. A disk image of the DVD and its checksum are produced. This step ensures the preservation of the original DVD as it was created. The disk image contains all the contents of the optical disc and can be used to reproduce a new copy of the master DVD and to convert it to other files. It does not allow for straightforward content playback.
3. The production quality file and its checksum are created. In order to create this file, we use the metadata of the original file to ensure that the technical characteristics of the video are preserved. We use the FFV1 codec, a lossless video codec, to allow original data to be reconstructed from the compressed DVD. If available, other streams such as audio and subtitles are copied in their original format. Furthermore, we selected the Matroska multimedia container due to its ability to combine unlimited number of media files, as well as its ability to combine different video and audio codecs [19]. The Matroska file is not fully supported by all media players, but it can be converted to other formats. Next, the metadata of this file is extracted.
4. The streaming file (optional) is created to give the museum the option to stream the content electronically. This file can be created using any container or codec provided by the transcoding software. Typically, an MP4 or Quicktime movie container is used, along with a lossy codec such as H264.
5. In order to ensure that the preservation master has been transcoded correctly, a video quality metric is used to compare the original disk image to the preservation files and optionally to the streaming file. The SSIM implementation compares a single image file and produces a measurement between -1 and 1. If the SSIM Index is 1, the reference file and copy are identical, which is what we aim for in this step. For lossy conversion, anything below 1 indicates a loss in quality from the original file. In our workflow, the measurement is averaged between the SSIM index of 100 frames. We tested this algorithm to ensure that the quality was preserved (See Tables 3 and 4). The output of the SSIM algorithm is recorded as metadata.
6. All metadata is formatted into a METS record.

Workflow Software Selection

The workflow was implemented on a computer running Mac OS X 10.7.5. The disk image creator [20] and checksum creator (md5) are available through the operating system. Invisor and MediaInfo were considered for the metadata extraction tool. Invisor does not offer a command-line tool, and there is a small price for the full version of the software [21]. In contrast, MediaInfo offers a free command-line tool and allows the extraction of technical and tag data of video and audio files. MediaInfo is also a cross-platform tool [22]; for these reasons, we chose it for the workflow.

In order to select the transcoding tool, we experimented with common software and libraries including Handbrake, Mencoder, and FFmpeg. Although it offers the Matroska container, Handbrake only offers two lossy codecs, namely MPEG-4 part 2 and MPEG-4 part 10. Mencoder and FFmpeg are very similar, and both offer the codecs that we chose for this work. In the workflow, we tested both tools and ultimately chose FFmpeg due to its extensive documentation, ease of use, and ability to handle single audio files. FFmpeg is also a complete cross-platform free software [23].

We found reference to several video quality metrics in the literature and determined that full-reference quality metrics would be best for this implementation. Full-reference quality metrics assess the quality of a transcoded video using the original video as reference.[1] The two most common full-reference quality metrics are Peak-Signal-to-Noise-Ratio (PSNR) and Structural Similarity (SSIM) Index. PSNR is an error-based method, while the SSIM index method uses the structural distortion measurement instead of the error. According to Hor and Ziou, the values of the PSNR can be predicted from the SSIM and vice-versa [24]. Thus, either metric would be appropriate for this work. The SSIM Index metric was chosen since the algorithm is available in several programming languages. We used the python implementation of SSIM, which has the following dependencies: Python, NumPy, SciPy, and Matplotlib [25][2]. All the dependencies are free and relatively easy to install.

Each software component of this workflow is available as a separate package. In order to implement the workflow, we built and tested each piece of software. We created a shell script to execute the workflow. This shell script runs automatically by clicking on it and only requires the user to input the name of the work to be transcoded. In addition to the shell script, we wrote python scripts to interface with the SSIM algorithm, to extract the technical characteristics of the work of art to provide the correct flags to the transcoding tool and to format all the metadata into a METS record.

Workflow Testing Results

As a result of the workflow we obtain a SIP that contains: an ISO disk image of the DVD; a Matroska file using FFV1 encoding; an access file (optional), and a METS record with technical metadata associated with the files and metadata about the

[1] In contrast no-reference metrics do not depend on the original video to estimate the quality. Such a metric will take one video and output one quality score. This kind of metric would be useful to automate visual quality control in traditional digitization workflows.

[2] This python implementation has the ability to display frames, although this is optional and not used in our workflow.

transcoding process. The workflow was tested on nine video art works. Table 3 shows the output of the video quality metric for one of the test cases. The table shows the average SSIM index for the production quality file (MKV) and the streaming copy (MP4). A SSIM Index of one means that the two images being compared have the same quality, and thus that the encoding was lossless. In the conversion from disk image to production quality file we aim to for a result of one. A SSIM Index below one indicates a loss of quality. However, in our experience we still have to determine an acceptable threshold for lossy conversion. As seen in the table, the streaming copy has a small loss in quality due to the lossy encoder used to create it. When visually inspected, the loss in quality is imperceptible.

Table 3. Compares SSIM Index (quality index) between ISO disk image and MKV/MP4.

DVD	100 FRAMES		500 FRAMES	
La Flecha	MKV	MP4	MKV	MP4
Average SSIM	1	0.974301	1	0.968107
Standard Dev.	0	0.000034	0	0.000104

In addition, we tested the differences in quality between an MP4 file created from the ISO image and one created from the production quality Matroska file. Table 4 shows that the quality is nearly the same when transcoding the movie from either source, which demonstrates that the production quality file can also be used to generate access files without an additional loss in quality.

Table 4. Comparison of the quality of a streaming file generated from disk images or Matroska input files.

DVD	100 FRAMES		500 FRAMES	
La Flecha	MKV	ISO	MKV	ISO
Average SSIM	0.97249	0.97430	0.96789	0.96811
Standard Dev.	0.00003	0.00003	0.00010	0.00010

During the development of this work we learned that some of the metadata associated with the original file must be incorporated to ensure the integrity of the production quality file. First, DVDs may contain more than one media stream. For example, a DVD might contain audio in several languages, or video that is shot at different angles. FFmpeg, and other video transcoders, only process one media stream of each kind by default. Therefore, in order to correctly map all the streams and preserve all the information in the DVD, the stream information metadata needs to be acquired prior to the transcoding step in order to be incorporated in the FFmpeg transcoding command. Second, re-encoding of audio is more complex than we originally thought. The audio and video streams should be separated and encoded individually. In addition, other more advanced audio encoders might be required to handle the audio files, depending on the audio codec requested. This process is not straightforward and using encoding defaults, such as allowing FFmpeg to guess the number of audio channels, will usually result in loss of conversion integrity. We did not re-encode the audio but rather copied it, ensuring that it was identical to the original. This also allowed us to avoid audio-video sync issues. Finally, the streaming file requires a specific combination of audio/video codecs since containers like MP4 support limited combinations. Separation and individual encoding of each stream is recommended, since the original file does not always provide audio codecs that work with video codecs like H264.

Archiving and Video Art Management

The collection will be archived in the iRODS [26] instance in Corral, the storage resource managed by TACC [27]. Corral consists of 6 petabytes of online disk and a number of servers providing high-performance storage for all types of structured and unstructured data. Through a parallel file system the data is accessible from all of TACCs computational resources and display systems, and can be moved through multiple serves and services at full speed simultaneously. The workflow will be integrated to the ingest process for this particular collection in iRODS. Museum staff will be able to submit the SIPs through different web, client, and command line interfaces and the metadata will be registered in iRODS database allowing search and access of the works of art. Once in iRODS, authorized museum can transfer and organize the collection, add descriptive metadata to the different files, verify their integrity over time, and manage their off-site and geographical replication depending on museum policies and the intellectual rights allowed to the museum for actions such as preservation and distribution of videos. Corral is monitored 24/7 and has security measures in place to avoid data loss.

Conclusions and Future Work

Our project started with the investigation of the technical lineage of the video collection to design a workflow to convert video art acquired in DVD format for which there are no better quality versions available for reformatting. The study indicated that the conversion had to preserve the integrity of the video art. We introduced a quality metric that helped us identify the metadata and the order of the workflow steps needed to achieve a lossless conversion. In addition we did not find significant differences in quality between the conversions in which the input file is a disk image or a Matroska file, so both may be used as sources for further conversions. Integrated to the ingest process the workflow will facilitate ingest of files and documentation to the storage resource where museum staff members will be able to manage the video art for access and preservation. Next we plan to test and implement the workflow for a larger DVD collection of video at the University of Texas Libraries that interfaces with a different repository system.

Future work will include creating a single software package with a Graphical User Interface. Through a GUI, we may abstract the different tools used for the workflow, allow the use of other transcoding software if different audio/video codecs are desired, and facilitate user customization of the overall workflow. We will also explore no-reference video quality metrics. To address very large video collections at the University of Texas, we will work on parallelizing the workflow, taking advantage the availability of large HPC systems that combine storage and computing resources.

References

[1] NDIIP. Sustainability of Digital Formats Evaluation, Factors and Relationships, Media Dependency. (2012). Retrieved February 4 2013 from:
http://www.digitalpreservation.gov/formats/fdd/fdd000348.shtm

[2] Esteva, Maria, Jordan, Christopher, Urban, Tomislav, Walling, David. Cyberinfrastructure Supporting Evolving Data Collections. In Proceedings of the 8th International Conference on Preservation of Digital Objects, iPRES 2011, Singapore, (2011). Retrieved February 4 2013 from: http://ipres2011.sg/conference-procedings

[3] Giunta, Andrea. Jorge Macchi: The Anatomy of Melancholy, The Jack S. Blanton Museum of Art, Texas, USA (2008). Retrieved February 4 2013 from: http://www.jorgemacchi.com/es/textos/286/jorge-macchi-anatomy-melancholy

[4] Documentation and Conservation of the Media Arts Heritage (DOCAM). Documentation Model. Retrieved February 4 2013 from: http://www.docam.ca/en/documentation-model.html

[5] TATE, Matters in Media Art, Templates for Acquisitions. Retrieved February 4 2013 from: http://www.tate.org.uk/about/projects/matters-media-art/templates-acquisitions

[6] Wijers, Gaby, Ethics and Practices of Media Art Conservation, A Work in Progress, Version 0.5 Part 2. (2010). Retrieved February 4 2013 from: http://www.packed.be/en/resources/detail/ethics_and_practices_of_media_art_conservation_a_work-in-progress_version_0/P75/

[7] Netherlands Media Art Institute, Video Art Preservation: Resources. Retrieved February 4 2013 from: http://nimk.nl/eng/video-art-preservation-resource

[8] Archivematica. Retrieved February 4 2013 from: https://www.archivematica.org/wiki/Main_Page

[9] Atempo Digital Archive. Retrieved February 4 2013. http://www.atempo.com/products/digitalArchive/index.asp

[10] Lazorchack, Butch, Rescuing the Tangible From the Intangible, The Signal, Digital Preservation, Library of Congress (2012). Retrieved February 4 2013 from: http://blogs.loc.gov/digitalpreservation/2012/07/rescuing-the-tangible-from-the-intangible/

[11] American Folklife Center, The Library of Congress. Veterans History Project, (2012). Retrieved February 4 2013 from http://www.loc.gov/vets/

[12] Stanford Media Preservation Lab. Moving Image Digitization Services at SMPL. Retrieved February 4, 2013 from: http://lib.stanford.edu/stanford-media-preservation-lab/moving-image-digitization

[13] McLellan, Cindy. BC to KY and Back Again: Crossing Borders to Discover New Processes for Digital A/V. Presentation at Beyond Borders: SAA Annual Meeting Session 607, San Diego, CA. (2012).

[14] Fleischhauer, Carl. Format Considerations in Audio-Visual Preservation Reformatting: Snapshots from the Federal Agencies Digitization Guidelines Initiative. ISQ Information Standards Quarterly. Spring, 2:2, pp. 36-40, (2010).

[15] Wang, Z. W. Z., Lu, L. L. L., & Bovik, A. C., Video Quality Assessment Using Structural Distortion Measurement. Proceedings International Conference on Image Processing. (2002). Retrieved from February 4, 2013 from: http://linkinghub.elsevier.com/retrieve/pii/S0923596503000766

[16] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. Image Quality Assessment: from Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 13:4, pp. 600-612. (2004). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15376593

[17] MediaInfo. Retrieved January 31 2013 from: http://mediainfo.sourceforge.net/en

[18] University of Texas Libraries. UT Video METS Profile, (2010). Retrieved February 4 2013 from: http://www.lib.utexas.edu/schema/UTVideoMETS.xml

[19] Matroska Media Container Home Page. Retrieved February 4 2013 from: http://www.matroska.org/

[20] Diskutil (8) OS X Manual. Retrieved February 4 2013 from: http://developer.apple.com/library/mac/#documentation/Darwin/Reference/ManPages/man8/diskutil.8

[21] Invisor: Inspect your Media Files. Retrieved February 4 2013 from: http://www.pozdeev.com/invisor/

[22] MediaInfo. Retrieved February 4 2013 from: http://mediainfo.sourceforge.net/en

[23] FFmpeg Documentation. Retrieved February 4 from: http://ffmpeg.org/ffmpeg.html

[24] Hor, A., & Ziou, D., Image Quality Metrics: PSNR vs. SSIM, ICPR 20th International Conference. (2010). Retrieved February 3, 2013 from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5596999

[25] Vacavant, Antoine. Structural SIMilarity. Retrieved February 3, 2013 from: http://isit.u-clermont1.fr/ anvacava/code.html

[26] iRODS: DataGrids, Digital Libraries, Persistent Archives and Real-time Data Systems. Retrieved February 3, 2013 from: https://www.irods.org/index.php

[27] Texas Advanced Computing Center (TACC). Corral User Guide. Retrieved Feburary 3, 2013 from: http://www.tacc.utexas.edu/user-services/user-guides/corral-user-guide

Author Biography

Maria Esteva holds a PhD in Information Science from the University of Texas at Austin. She is a researcher and data curator at the Texas Advanced Computing Center.

Karla Vega holds an MS in Mechanical Engineering from the University of California, Berkeley. She works as a research engineer at the Texas Advanced Computing Center.

Bethany Scott holds an MS in Information Science from the University of Texas at Austin. She is Cataloging and Metadata Librarian at the Charlotte Mecklenburg Library in Charlotte, NC.

Summer Gunnels is an undergraduate student at the Department of Mechanical Engineering at the University of Texas at Austin, She works as a student assistant at TACCs Vislab.

Keerthana Kumar is an undergraduate student at the Department of Computer Science at the University of Texas at Austin.