The VRA Bulletin, Spring 2005, v. 31 no.3, pp.50-55.
ISSN: 1046-9001(print) 1097-8020 (online)
http://online.vraweb.org/
http://www.vraweb.org/publications/bulletin.html

Digitizing the KET Archives: A Brief Case Study

Lisa R. Carter

Kentucky Educational Television (KET) began broadcasting original programming statewide in September 1968. As the state's public broadcasting network, KET provides Kentuckians with PBS programming in addition to their own locally produced arts, cultural, documentary, and public affairs productions. The network is a leader nationally in the distribution of adult education programs, specifically video instruction in adult basic skills and workplace education. In addition, the station delivers KET-produced professional development seminars and fully accredited high school classes in foreign languages, the humanities, and physics to Kentucky schools.

It is KET's mission "to serve the unmet needs of Kentucky's homes and families, educational institutions and students and its workforce." Their values include generating products that "educate, inform and inspire," using their resources to "provide access to lifelong learning opportunities" and exploring and developing "more effective techniques and technologies." Their vision of themselves is that of a "resource for lifelong learning," an "electronic town hall," a "cultural showcase celebrating the heritage of the Commonwealth," and a "leader pioneering the effective implementation of emerging technologies in the delivery of information resources to all Kentuckians."[1]

With this perspective of their purpose and thirty-five years of programming generated in the service of these goals, and in light of the Federal Communications Commission's 1997 mandate that all television stations be broadcasting digitally by 2006,[2] KET decided to convert their vast collection of analog video to digital. Several converging factors motivated KET to take action. Like many prolific content producers, KET had accumulated more videotape than they could effectively store. Not only were they out of space, but the network's staff also struggled to keep track of what had been sent to the archives. By digitizing their collection, the people at KET hoped to condense the physical size of their collection while creating a searchable and useful mechanism for finding content. In addition, they knew that their old magnetic tapes were reaching the end of their expected longevity. Considering the age of its tapes and the short shelf life of magnetic media, KET knew that they had to migrate their content to a new format to ensure its survival. Most importantly, KET felt that in order to better reuse and repurpose their assets in future initiatives that would further the educational mission, the assets needed to be

digital. Bringing legacy content into the digital environment would enable faster location of historic footage, enhance the editing process, and allow the distribution of this educational media on new and innovative platforms.

In January 2003, the network received $600,000 in grants from two private donors to undertake a pilot project that would digitize KET's video archive. The purpose of the project was to "convert and preserve its entire locally created program archive in digital format. This will ensure that 3[5] years of KET programs are preserved and not lost or damaged by transferring the program archives to the new industry standard format. The programs will be maintained on state-of-the-art digital storage retrieval platforms and with indexing, be accessible via the Internet."[3] As the project began to take shape and was discussed as an archival activity, it became clear that the project would benefit from additional expertise from outside the station. The University of Kentucky Audio-Visual Archivist was brought in under contract to manage the project full time for the first year-and-a-half. Service providers were brought in to provide a basic understanding of what potentially useful products existed and to share information about similar projects.

In discussing their project with vendors and surveying the public television and archival communities, KET discovered that, at the time, there were no similar projects to use as models. No one vendor had a solution that would meet all the goals of the project, and their proposed project was one of the first to address the digital conversion of legacy video materials on a collection-wide scale. There were few standards, studies, or best practices relevant to digitizing local television. Because of this lack of precedent, KET undertook this effort as a pilot project and as an opportunity to learn and contribute to the dialogue about standards and best practices for digitizing television materials, while creating digital surrogates for their soon-to-be-obsolete tapes.

The Place to Start: How Many Tapes?

The first thing to be determined was the scope of the project. While KET had been working from an estimate of 7,500 hours of local programming, they had no way of knowing exactly how many tapes they had to ingest or how many actual hours existed on those tapes. The available database of tapes represented a combination of the tapes that existed in the "for-air" tape traffic library, as well as the tapes that no longer air, which had been physically transferred to a large storage room. The latter was included in what KET was calling its archives, and these tapes had not been documented in a way that allowed them to be distinguished from the currently airing tapes. An inventory was taken to determine how many tapes existed in the locally produced program archive.

Every tape identified as an archival program tape was organized according to format and tape number on the archives' shelves. Tapes that were not program tapes (had not aired on the main channel, such as GED, professional development, and distance learning) were moved to another physical area. All of the 1-inch tape, which was thought to be the oldest, was placed first in tape

number order, followed by all of the 3/4-inch, which was considered the next oldest format, followed by the Beta SP, which is the current broadcast format. This work was quickly done, since this was generally the existing order of the tapes.

The existing FileMaker Pro database was sorted by tape number and the inventory staff went through each record and each tape in order, matching database records to tape. Inventory staff added information to each record that described a tape on the shelf. Through this process, the information that was verified and gathered included: Tape Number, Location, Series Title, Program Title, Part/Reel Number, Record Date, Length, Format, and Version. If the information that existed in the database conflicted with the tape's label or with visual observation, the information was changed in the database. Since Location information had not been previously entered into the database, each tape that existed on the shelves was given a new, unique location number. If a tape was not on the shelf, this field was left empty and a sort at the end of the inventory project allowed for the elimination of those records from the database. At the end of this stage of the project, each tape in the archives had a record, each record described an actual tape, and each record was updated and corrected with the information existing on the tape's label. This resulted in a count of 9,133 local programming tapes in the archive. Of these, 5,453 were 3/4-inch, 1,101 of those were 1-inch, and 2,579 were Beta SP tapes. While this process documented the number of tapes, the number of minutes on each tape was only occasionally recorded on the label and/or in the database, so the number of hours to be digitized remained unconfirmed.

Selecting a Format

If the first question the service providers asked was "How many tapes, how many hours?" then the second question they asked was "What is the digital format you want to use?" While this question seemed straightforward, the answer really went to the heart of the project. The project had been defined in the grant proposal as a preservation project. But "preservation" held many different meanings for each of the members of the project team. The archivist's understanding from professional forums was that there was no concretely settled preservation standard at the time, and digitization was not perceived of as a preservation destination. If video was to be digitized for preservation, then ingestion should result in uncompressed files. The production staff was primarily concerned that the resulting files be easy to find and use. Since KET had not yet decided on a digital media acquisition format, this implied the selection of a format that could be used by most editing platforms. The resulting files needed to have at least the same visual quality as the originals and to be at least as easily editable as tape. While an uncompressed file format would have suited the producer's needs and matched the current archival best practice at the time, the technologists pointed out that if KET bought enough storage to store these videos as uncompressed digital files, there would be no project money left to actually do the ingestion. Given the timeframe and the budget of the grant project, the digitization had to result in compressed files.

In order to employ a diverse storage strategy, KET chose to store the resulting essence files on a RAID and make backups of those files on DVDs. The project team determined that if the essence files were to be backed up as identical copies on DVD, meaning without transcoding, the essence files should be created in a DVD-compliant format. While this matter was under discussion, PBS set its contribution standard for standard video at MPEG-2 8 mbps 4:2:0.[4]

Given the criteria of the project, the limitations of the budget, and emerging standards, KET decided that the files resulting from this project should be MPEG-2 files with a 4:2:0 color profile using a 2 pass variable bit rate process. The 3/4-inch tapes were to be digitized at an average of 4.5 mbps and the 1-inch tapes with a ceiling of 9 mbps. The Beta SP tapes, which were likely to have a longer life span, exhibit less playback difficulties, and require a higher digital resolution, were not to be digitized in this initial project. The essence files resulting from the digitization of the 3/4-inch and the 1-inch tapes would also be transcoded down to Windows Media Player proxies for delivery over the Internet to match KET's current delivery specifications.

Handling the Original Tapes

Since the resulting files would be compressed and their versatility for reuse somewhat unknown, KET was convinced that they needed to keep the original tapes indefinitely. During the ingestion workflow, basic preservation work was to be done on the magnetic tapes, not only to stabilize the original media for indefinite, long-term storage, but also to ensure the best transfer. Tape preservation handling, inspection, and cleaning guidelines for the project were developed based on several archival resources, most specifically, the AMIA Videotape Preservation Fact Sheets.[5] In addition, KET was concerned about sending the tapes off-site for digitization. As a television station, they had the technical infrastructure to support the project and space for the work to be done, but they did not have the in-house staff with the expertise and time to focus on the project. Therefore, the network sought an in-house solution that would be managed and implemented by an outside vendor.

Metadata

Additionally, in support of the original goals for this project, KET wanted the digitizing process to include some enhancement of the administrative, technical, and descriptive metadata about the tapes. At the very least, technical information about the files was to be recorded as the videos were digitized, but most importantly for KET's concerns, condition and descriptive information about the original tapes needed to be added to the FileMaker Pro database. Enhanced data, based on information shown on the video, such as accurate title information, production credits, participant identification, and subject description, was desired. This metadata gathering process was envisioned as partially manual, based on what the digitizing staff heard and saw, and partially automated, utilizing software to electronically gather closed captioning information, detect changes in scenes, recognize on-screen text, and convert speech to text.

Selecting a Vendor

Based on this vision for the project, KET issued a Request For Proposals.[6] Since few service providers were prepared to undertake the work as outlined in the RFP, the response was low (five responses). Only one service provider could approach the requirements in the RFP for an estimated cost that was close to the budget on hand.

As the selected vendor, the PPS Group's Scene Savers division hired a team of four people to achieve the goals of this project. In addition to a project director and technical support based in Cincinnati, Scene Savers relocated a Project Manager and locally hired two ingestion technicians and one quality control assistant. As required by the project specifications, all of the equipment to be used for tape inspection, cleaning, playback, and ingestion was supplied, maintained, and owned by Scene Savers. The PPS Group and the KET project team worked closely to further specify the project parameters and processes to ensure that the project's goals were met within the available budget.

Project Workflow

In order to achieve the goals of the project, Scene Savers constructed a workflow and documented the process in a project manual. This workflow included the following steps: retrieval of the tapes from the archive, inspection, climatizing, cleaning, encoding, copying digitized video in folder to a storage server, quality control, metadata entry, relabeling and bar coding the original tape, and returning the original tape to the archives.

Retrieval

The archive tapes are stored in a large open room that is not secure, minimally monitored, and with marginal environmental control. In spite of this, the tapes remain relatively undisturbed, since few people have access to a database describing the content. However, the project had to account for the occasional use of tape by KET staff and needed to include tracking so that the encoding team could find each tape and could account for all of the tapes they handled at all times. The corrected archives database was sorted by format (3/4-inch or 1-inch) then tape number, and a list was printed, placed on a clipboard, and used to manually check the tapes out of the archives space.

The digitization workstations were set up in the broadcasting area of the facility, so the tapes were removed from the storage area for ingestion. As the videos were placed on book trucks, the cases were dusted and then the tapes were transported down a floor and across the building. The Scene Savers personnel placed a check mark on the list and a date noting which tapes were removed, when, and by whom.

Inspection

Each tape was inspected before being placed in any machine, since a damaged tape can damage a playback machine and be further damaged during playback. Tapes were smelled to determine if there was any off-gassing occurring that indicated binder deterioration and warned

of potential oxide shedding problems. A visual inspection of the tape included looking through the cassette windows and opening the cassette door. This inspection was aimed towards finding mold or fungus that might be on the tape, which then required quarantine and specialized treatment. In addition, any oxide shed was detected at this stage if any brown or black flakes were observed loose in the cassette. Pack problems, such as stepped packs, gaps, folds, etc., due to uneven tensioning and poor previous playback conditions, were seen during this inspection. Any tape that looked or smelled like it might be a problem was set aside for more specialized treatment at a later date.

Climatizing

After inspection, the tapes were placed on shelves in the digitizing area to climatize for twenty-four hours. The storage conditions in the archives were considerably different than those in the ingestion area. The tapes were allowed to stabilize to the temperature of the room in which they were played back, stopping any expansion or contraction that may have been occurring before the tape was placed in a machine.

Cleaning

Both the 3/4-inch and 1-inch tapes were run through refurbished and recalibrated RTI cleaners. In the cleaner, the tape ran past a cleaning tissue to remove any loose oxide and debris from the tape surface. This pass also gave the tape a smooth and evenly tensioned wind. This prepared the tape for the best transport through the playback machine, resulting in less clogging of the tape heads and better playback results when encoding. While the cleaners arrived from RTI with a burnishing blade, on the advice from the company, these were removed since all of the materials being cleaned were over ten years old. This was to avoid the possible stripping of particularly bad tapes.

The previous steps were taken to avoid significant damage to the original tapes, but were also implemented to protect the playback equipment and to provide a better signal during the encoding. Because the tapes are machine dependent, the success of a project that requires converting a signal on magnetic tape to any format relies on the best playback conditions, including tapes that are in the best shape they can be; equipment that is cleaned, calibrated, and maintained by professionals; and equipment operators who are well trained in video transfer and problems that may occur during playback.

Encoding

Two encoding machines were set up for monitoring by one technician at a time. This allowed Scene Savers to run two encoding stations for two 8-hour shifts or sixteen hours a day. Accounting for the time required for set up, deck cleaning, and database entry, this resulted in encoding six hours of programming per workstation per shift in real time. Therefore, approximately twenty-four hours of programming was digitized each day.

Each station contained a refurbished playback machine, a computer with a monitor, speakers, a broadcast quality monitor, a waveform monitor, vector scope, and audio control boards. A Digital Rapids hardware encoding card generated the MPEG-2 file.

During encoding, the video signal was logged by Virage software. The Virage software controlled the playback decks and the Digital Rapids encoder, allowing the technician to start and stop playback, encoding, and logging by clicking one button. However, the software's main function was as indexing software. Virage generated key frames and detected scene changes, segmenting the video within the database so that portions of a program could be found, documented, and used more easily. In addition, as the software analyzed the video, onscreen characters, such as titles, slates, and lower thirds were recognized and recorded as text in the database. The Virage SoftSound engine translated speech to text (with varying success), creating some searchable text for those programs that did not have closed-captioning. For those videos that had closed-captioning, Virage captured it as a text file within its database.

Copy Digitized Video to Storage

At the end of an encode, the file was saved to the desktop of the computer on which it was digitized. After checks to verify that the ingestion occurred properly, the MPEG-2 files were saved to one of nine 1.75TB NAS servers. Initially, these servers were backed up to one of the other nine servers, but in January 2005, DVD backups of each file were burned, to free the servers for the storage of original essence files and proxies.

Quality Control

Within a day of saving the files to the NAS server, the quality control assistant and project manager played the digitized files to ensure that they were captured completely and that the files were not corrupt. The digital files were scanned to discover any problems that might have been missed by the encoders. It was also at this stage that significant descriptive information was added to the FileMaker Pro database records.

Metadata Entry

During the digitization process, Scene Savers enhanced the data recorded in the database about the video, thereby meeting KET's goal of having more descriptive, searchable data available about their video content. This documentation of the tapes was not full or standardized cataloging; the encoders were primarily video technicians and to increase data entry time would have increased the cost of the project. Yet the minimal manual data gathering that was done improved retrievability of KET content substantially.

As was discovered during the initial inventory phase, KET content had been previously described haphazardly. In most cases, information about the tapes was recorded minimally, usually limited to as few completed fields as Tape Number, Series Title, Program Title, Episode Number, Record Date, and Length. The data in these fields was applied inconsistently, the series

title often showing up in the Program Title field, series titles written in several different forms, producer information in the description field, descriptions in the notes field, etc.

Key to the success of the project was Scene Saver's normalization of the data, including the creation of regular forms of titles, creators, lengths, as well as verification of the accuracy of the information in these fields and enhancement of inadequate data. In addition to description guidelines developed and implemented specifically for this project, KET chose to use the recently released metadata schema, PB Core, to structure its database and data entry activity. Element definitions were analyzed and enhanced for KET's specific purposes.

At the time KET initiated use of PB Core, the schema consisted of fifty-eight elements. For the purposes of this project, only forty of these elements were used. Twenty-two of the technical elements such as Data Rate, Aspect Ratio, Frame Rate, Bit Depth, etc., were constructed to either be calculated fields or auto-enter fields. This left eighteen PB Core fields that needed to be filled out during the encoding and quality control process. These included:

- Title
- Alternative Title
- Series Title
- Program Title
- Episode Title
- Description
- Tape Number
- Identifier (digital file)
- Notes
- Date Created
- Format Duration (length)
- Audio Channels
- Color
- Language
- Audio Versions
- Format File Size

In addition, KET and Screen Savers added the following fields to gather information needed for the project, resulting in twenty-two fields that the encoders and/or the quality control staff needed to check for accuracy or fill out[7]:

- Tape Condition
- Tape Location
- QC'd By
- QC'd Date
- Encoder Name
- Encoding Station

Relabeling and Barcoding the Original Tape

Once the quality control and metadata gathering steps were completed, the tapes were rewound and relabeled for permanent storage back in the archive. The tapes were rewound in a playback machine that was designated specifically for that purpose, so that the playback machines being used for encoding were not interrupted. In addition, since the rewind deck was also regularly maintained, this final step ensured an even-tensioned wind for the tape as it was returned to storage.

The cases were stripped of their former labeling. A new label containing Title, Location, Tape Number, and Identifier was printed with a barcode that maps to the Identifier number using the Code 39 standard. A label was placed on the spine of the tapes case and an identical label was placed inside on the tape's cassette. The labels are made out of polypropylene and chosen for their potential durability and inertness. They have a synthetic adhesive that should not dry out as the glue on typical tape labels does. The labels were printed using a resin heat transfer process that should slow fading. It is expected that the labels will have longevity beyond that of the magnetic media.

Return Tape to the Archives

Finally the tapes were returned to the archives and checked back in on the list originally used to check them out. Occasionally, when a problem tape was set aside, it was not returned to the archives. This was documented on the list and in the tape's database record.

Backing the Files Up to DVD

The DVDs were burned from the initial backups made on the NAS servers. The media selected for this process was Mitsui MAM-A silver DVD-Rs with a Phthalocyanine dye data layer and a white inkjet printable layer on the top surface. The DVD cases ordered were tested to make sure the discs were suspended by and rotated freely in the hub. KET selected an Amtren FlexWriter SE4 automated DVD duplicator and printer which uses the Padus DiskJuggler recording software for controlling the robot remotely. KET staff wrote a script that instructed the robot which files to write to DVD from the NAS server. This same script also told the robot what to print on the white printable surface of the DVD and told a paper printer what to print on a wrapper that was placed in the DVD's case. The script retrieved this data from the FileMaker Pro database, which was updated by Scene Savers as they normalized and enhanced the videos' metadata. This label information is equivalent to the labeling on the original tapes including Title, Tape Number, Identifier, Description, and a barcode.

KET staff then matched the labeled discs with the printed wrapper that was placed in the DVD's case. A second inspection occurred to make sure the right disc was in the right case before the disc was filed. The barcodes were scanned as the discs were shelved to mark the DVD as shelved, and a random percentage of the discs were checked to ensure that the discs were fully readable.

More Questions than Answers

Throughout the project, practicality challenged KET's original goals as well as archival theory and best practices. Due primarily to a limited budget and secondarily to minimal previous archival effort, this project served to expose more issues than it resolved. In conclusion, some of these issues are outlined briefly here.

As public television stations address this digital transition, how can collaboration with information professionals enhance the results and cost-effectiveness of such projects? KET's experience demonstrated that bringing an archivist to the table helped them to examine and come to consensus on the meanings of terms like "archives," "archival," "preservation," "retrievabilty," and "accessibility." While these definitions were rarely resolved to meet with any archival ideal, the archival perspective increased expectations and offered options focused on longevity. The earlier an archivist is brought in to advise on the process that governs an asset's life-cycle, the more efficiently selection, preservation, and accessibility guidelines can be applied resulting in more cost-effective digital projects.

Digitizing an archive involves a great deal more activity than anyone would expect. Based on the KET experience, television stations currently underestimate their preparedness to take on a project such as this. Underlying conditions necessary for a successful project, such as thorough documentation of the number of tapes, number of hours, and the value of the footage, or a clear understanding of the expected uses of the content, generally are not in place. While television production staff has not anticipated a need for documentation and analysis of assets, the archival community assumes that these conditions are in place. Both communities need to have a better understanding of the current state of television archives to better plan for digital transition projects.

The digital transition for television and their programming archives requires fundamental cultural change. Broadcasting has been based on pushing carefully constructed, time-based content out to an audience. When a program was finished with its airing cycle, it no longer needed to be accessed. As broadcasters look to the digital future, they see an essential transformation of their business where content provision is driven more by individual user need than ever before. The digital environment requires a content producer to employ metadata standards, longevity planning, and accessibility as significant pre-production activities.

While information professionals can provide a helpful advisory role in television's transition, compromises will have to be made to traditional practices and standards. The archival ideals that frame appraisal, preservation, and descriptive practice can be dismissed out of hand or viewed as unrealistic in light of minimal budgets and looming deadlines. Ensuring the preservation and accessibility of television materials will require archivists and broadcasters to meet halfway on their expectations for digital projects. Further analysis of projects like this one at KET can better prepare all of us for realistic forays into our digital future.

Notes

1. "Our Mission." KET 2005. Kentucky Educational Television. 17 Jan. 2005 <http://www.ket.org/agency/about/mission.htm>.
2. Federal Communications Commission. "Advanced Television Systems and Their Impact Upon the Existing Television Broadcast Service, 5th Report And Order on MM Docket No. 87-268." 23 April 1997 Federal Communications Commission. 17 Jan. 2005 <http://www.fcc.gov/Bureaus/Mass_Media/Orders/1997/fcc97116.pdf>.
3. Hopkins, Shae. Grant Proposal to the J. Graham Brown Foundation. Kentucky Educational Television. 2002: 1.
4. Butler, Jerry. "Bit rates, etc." E-mail to Paul Stackhouse. 27 May 2004.
5. Wheeler, Jim and Peter Brothers, ed. Hannah Frost. "Videotape Preservation Fact Sheets." [AMIA] Guidelines, Manuals, Q&As and Fact Sheets. 2002. Association of Moving Image Archivists. 17 Jan. 2005 <http://www.amianet.org/publication/resources/guidelines/videofacts/about.html>.
6. Request for proposals available upon request from author. E-mail: lisac@uky.edu.
7. Field definitions and instructions available upon request from author. E-mail: lisac@uky.edu.