Rescue and Preservation of Videotape Collections

Sun PASIG Event
June 25, 2009

Brian Campanotti
CTO
Front Porch Digital
The Reality

- Video tape and film archives are in a constant state of decay
- Every year valuable content (monetary, cultural, historical, etc.) is being lost to deterioration
- Access to this legacy material requires human intervention and is naturally plagued with difficulties in locating content
- Legacy video tape machines are becoming harder to find and maintain adding to the risk of access loss
- It cannot be debated that migration of this content into a digital content repository is the necessary solution
Difficult Decisions

• Significant perceived barriers exist in the large scale migration of this legacy content

• Labour costs, capital equipment investment, storage and bandwidth calculations can all be overwhelming

• A plethora of difficult decisions must be made up front including which compression format, what bitrate, which wrapper format, what storage technology…

• And what about the metadata?

• My tapes have been around for 50 years already...maybe I can just wait another couple of years?
Large Scale Migration

• An hour of linear media migration takes quite a bit MORE than an hour using traditional approaches:
  – Tape inspection, cleaning and preparation
  – Metadata entry
  – Level checks
  – Quality assurance
  – Post migration checks
  – Storage on nearline/tape media

• Skilled video professionals managing each step of the migration process can be very cost prohibitive

• Video tape accumulation and decay continues as people seek the holy grail...
The Goal

- Define a straightforward media preparation process to alleviate bottlenecks and minimize the required operator skill set
- Allow machines to manage the migration process efficiently and in an unattended fashion
- Leverage high speed video tape robotic systems to perform repetitive tasks efficiently
- Rely on software based tools to identify problematic content
- Automatically collect valuable metadata during each step of the process
The Goal

- Generate multiple formats of digital content in a single pass during the migration process (including a preservation master)
- Automatically manage the long term storage of this content on disk and data tape storage systems
- Allow rules and policies to define content replication practices
- Empower users by allowing desktop browsing of the migrated content minutes after each migration operation is completed
- Provide software based tools for easy content access, searches, timecode based partial restore, etc.
- Now let us look at the process...
Media Preparation

• Allow for moderately skilled employees to follow a clearly defined workflow in efficiently preparing media for migration

• Leverage software based tools to manage the process

• Media inspection, bar coding and database checks reduce the potential for downstream bottlenecks or errors

• Importing legacy video tape database/catalog information enables metadata flow and enrichment at each step

• The media preparation process should take only a minute or so per media item
Tape Cleaning

• Even for properly stored media, particulate and other factors can affect migration quality

• As part of the tape preparation step, running the media through an archival quality tape cleaner is recommended

• Detection and removal of dirt, residue, etc. as well as tape re-tensioning will reduce head clogs and other downstream bottlenecks

• Metadata is collected during the cleaning operation to indicate cleanliness of the media, tape damage, tension issues, stretching, etc.

• Tape cleaning can occur at approximately 20x faster than realtime adding no more than a couple of minutes to the preparation process
Archival Tape Cleaner

The past, present and future of media
Migration Engine

• The migration engine should capture any number of encoded versions of the content in a single pass (you may only have one shot!)

• Downstream transcoding is an option but with large volume migration this can be costly and present a significant bottleneck

• A preservation quality master and an intermediary high resolution version should be captured along with a number of frame accurate proxy versions

• Metadata should be collected on a frame by frame basis identifying video quality, audio levels, drop outs, etc.
Migration Engine

• Supported encoder formats include:
  – JPEG2000
  – MPEG2 30/40/50
  – DV25/50
  – Windows Media
  – Flash Video
  – H264
  – MPEG1
  – Etc

• Supported wrapper formats include:
  – MXF
  – QT
  – DV
  – Etc
Migration Engine
Video Tape Robotic Library

- As media progresses through the preparation and cleaning process, it can be loaded into the video tape robot for unattended migration.

- With support for up to 60 tapes and up to seven BETA, UMATIC, VHS or DV decks nearly 1000 hours of content can be migrated each week (~80% efficiency).

- Each deck is connected to its own Migration Engine providing high volume single pass, multi-format encoding.

- A simple user interface allows efficient monitoring of the overall process.
Content Storage Management

• Obviously a significant amount of data is generated during the migration process.

• Consider the generation of the following formats for each migrated item:
  – JPEG2000@80Mbps
  – MPEG2@50Mbps
  – H264@3Mbps
  – FLV@1Mbps
  – WM9@500kbps

• This equates to approximately 500GB/hr for a seven stream robotic system or up to 60TB per robotic system per week!

• Obviously an efficient, scalable and robust storage infrastructure is required to deal with this digital content...
Content Storage Management

- Content Storage Management (CSM) is a content aware, intelligent storage middleware solution and is **not** to be confused with HSM.

- CSM systems interface to any type of device which produces encoded content (images, audio, video, etc.) and bridges the gap to any type/amount of nearline, archive and offline storage.

- CSM solutions differ as they provide content aware features such as transcoding, timecode based partial restore, file based quality assurance, usage analytics, storage metrics gathering and reporting, etc. HSM systems do not...

- These solutions allow direct connection to MAM systems (or other metadata management products) to fundamentally link the metadata with the digital content.
Content Storage Management

• Generating and storing preservation copies of valuable assets eliminates the need for immediate corrective actions (or archaeology) as we can always fall back to the master.

• Intermediary formats allow easy interaction and repurposing of the content today while integrated in-path transcoding provides comfort for future format/wrapper generation.

• CSM based quality assurance tools are our “eyes and ears” and help us focus in on the small percentage of problematic material in a vast content repository.

• Timecode based partial restore can allow user to browse frame accurate proxies at their desktop and quickly identify and restore only required portions of high resolution digital assets.
CSM Storage Tiers

- CSM systems support multiple “tiers” of storage within the digital storage infrastructure each with its own pros and cons

- The desired workflow will help define which tiers are required, capacities of each, bandwidth, etc.

- Each storage tier can be scaled incrementally at any time to fit evolving requirements and workflows

- Several physical storage technologies are available and should be considered
CSM Storage Pyramid

- Fast Local Disk Storage (ONLINE)
- Fast Network Disk Storage (NEARLINE)
- Robotic Library Storage System (ARCHIVE)
- Digital Media Shelf Based Storage (OFFLINE)

Faster Access and Higher Storage Costs

Storage Capacity
File Based Storage Benefits

- Storage Density!!! More than 20 hours of preservation quality content can be stored on a single LTO4 data tape

- Low media costs (~$60) allow automatic creation of duplicate copies of valuable content for distribution and disaster recovery

- No QC or human intervention is needed after migration as files are always guaranteed to be exact copies

- Simple migration to new digital storage formats and encode formats via integrated transcoding solutions protects valuable content
## Significant Real Estate Savings

<table>
<thead>
<tr>
<th>Tape Format</th>
<th>Per Unit Volume (m³)</th>
<th>Per Media Capacity (Hours)</th>
<th>Media Units Per m³</th>
<th>Hours of Content Per m³</th>
<th>Density vs Betacam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betacam</td>
<td>0.00087</td>
<td>1.00</td>
<td>1,152</td>
<td>1,152</td>
<td>1.0</td>
</tr>
<tr>
<td>VHS</td>
<td>0.00049</td>
<td>1.00</td>
<td>2,037</td>
<td>2,037</td>
<td>1.8</td>
</tr>
<tr>
<td>DV</td>
<td>0.00014</td>
<td>1.00</td>
<td>7,267</td>
<td>7,267</td>
<td>6.3</td>
</tr>
<tr>
<td>LTO1</td>
<td>0.00023</td>
<td>4.44</td>
<td>4,326</td>
<td>19,228</td>
<td>16.7</td>
</tr>
<tr>
<td>LTO2</td>
<td>0.00023</td>
<td>8.89</td>
<td>4,326</td>
<td>38,456</td>
<td>33.4</td>
</tr>
<tr>
<td>LTO3</td>
<td>0.00023</td>
<td>17.78</td>
<td>4,326</td>
<td>76,913</td>
<td>66.8</td>
</tr>
<tr>
<td>LTO4</td>
<td>0.00023</td>
<td>35.56</td>
<td>4,326</td>
<td>153,825</td>
<td>133.6</td>
</tr>
</tbody>
</table>
## Many Physical Storage Format Choices

<table>
<thead>
<tr>
<th></th>
<th>Hard Disk Storage</th>
<th>Flash Storage</th>
<th>Optical Storage</th>
<th>Data Tape Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fast transfer speeds</td>
<td>• Random access</td>
<td>• Random access</td>
<td>• Fast transfer speeds</td>
</tr>
<tr>
<td></td>
<td>• Good storage density</td>
<td>• Persistent storage</td>
<td>• No contact read/write</td>
<td>• Storage density</td>
</tr>
<tr>
<td></td>
<td>• Random access media</td>
<td>• Low energy costs</td>
<td>• Acquisition format</td>
<td>• Rugged media</td>
</tr>
<tr>
<td></td>
<td>• Fast mechanical times</td>
<td>• Acquisition format</td>
<td>• Emerging holographic</td>
<td>• Cost per TB</td>
</tr>
<tr>
<td></td>
<td>• Multiple R/W streams</td>
<td>• Fast access</td>
<td>• Low replication costs</td>
<td>• Extremely portable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Rugged</td>
<td></td>
<td>• Expansion costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Low replication costs</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tied to host chassis</td>
<td>• Limited read/write</td>
<td>• Low storage density</td>
<td>• Sequential data access</td>
</tr>
<tr>
<td></td>
<td>• Not portable</td>
<td>• Cost per TB</td>
<td>• Low transfer speeds</td>
<td>• Head and tape wear</td>
</tr>
<tr>
<td></td>
<td>• Cost of ownership</td>
<td>• Proprietary formats</td>
<td>• Questionable shelf life</td>
<td>• Slower access</td>
</tr>
<tr>
<td></td>
<td>• Highly mechanical</td>
<td>• Low transfer speeds</td>
<td>• Portability</td>
<td>• Single R/W stream</td>
</tr>
<tr>
<td></td>
<td>• Expansion costs</td>
<td>• Low storage density</td>
<td>• Single R/W stream</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• High replication costs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# A Question of Scale

<table>
<thead>
<tr>
<th>Format</th>
<th>File Size (GB/h)</th>
<th>Transfer Time (GBE Network)</th>
<th>Cost on Disk Storage</th>
<th>Cost on Data Tape Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM9@500kbps</td>
<td>0.23</td>
<td>4s</td>
<td>$0.46</td>
<td>$0.03</td>
</tr>
<tr>
<td>MPEG2@50Mbps</td>
<td>23</td>
<td>6m</td>
<td>$46</td>
<td>$2.75</td>
</tr>
<tr>
<td>JPEG2K@80Mbps</td>
<td>36</td>
<td>10m</td>
<td>$72</td>
<td>$4.30</td>
</tr>
<tr>
<td>ProResHD</td>
<td>65</td>
<td>17m</td>
<td>$130</td>
<td>$7.80</td>
</tr>
<tr>
<td>DPX Film Scan</td>
<td>3,400</td>
<td>15h 16m</td>
<td>$6,800</td>
<td>$400</td>
</tr>
</tbody>
</table>
Storage Expansion and TCO

- Disk and tape based storage systems both represent a fairly significant initial capital investment.

- Disk storage need only represent a nominal percentage (<10%) of the overall storage infrastructure in most cases.

- Expansion of disk arrays involve chassis, RAID controllers, disk drives, fibre channel ports, etc. while expansion of the data tape library is merely the cost of the tape media itself.

- By remaining current with the latest data tape technologies, physical library capacities naturally expand over generations.

- A data tape sitting idle in the robot or on a shelf consumes no energy and requires no cooling.
Media Longevity

- Media life no longer need be a concern as migration to new storage media does not require human intervention.

- Migration can be done many times faster than realtime as a background task by the CSM solution.

- Migrated content is a byte-for-byte duplicate of the original with NO degradation.

- Additional software based “checks and balances” can be added to the migration process for additional comfort.

- Recent trends show a 50% increase in storage performance and 100% increase in storage density per generation.
Content Protection and DR

- Most video tapes archives in the world consist of one single copy of each asset

- The Universal Studios fire in June 2008 destroyed nearly 50,000 valuable archive assets, some irreplaceable

- Once in the file-based domain, identical duplicate copies of all assets can be easily made and moved offsite for additional protection (for the cost of the media itself)

- Distributed and interconnected digital repositories can provide for true resiliency and extremely high SLAs
Metadata

- We touched briefly on metadata at several points and hopefully its clear CSM and MAM are complimentary technologies.

- Metadata management is as important of a component as the essence managed by the CSM solution.

- MAM systems use unique identifiers to link metadata records with the actual digital essence contained and managed by the CSM system.

- Metadata and essence paths should be naturally independent and rely on best of breed, application specific solutions.
Standards

• Front Porch Digital is leading the SMPTE 2034 (AXF) initiative for industry standardization of CSM storage formats

• This committee involves other manufacturers as well as user groups in the media and entertainment space such as ESPN, Discovery and the BBC

• Our goal is to standardize CSM optical/tape media storage formats for long term protection and accessibility

• This will allow portability between CSM systems as well as recovery of stored data leveraging open source utilities
SMPTE 2034 (AXF) Initiative

- Although important to ensure long term accessibility to stored assets, this should not act as a barrier to adoption as data migration can later occur.

- Existing IT standards (TAR, etc) are not acceptable for these specialized CSM applications because they prevent some more advanced functionality such as partial restore.

- AXF is fundamentally a combination of XML based descriptive metadata and block based highly accessible essence storage.

- Committee work updates are readily available via the SMPTE website.
Clear Benefits

• Migration of legacy video tape archives is a pressing although daunting issue

• Proven solutions are key to successful migration and the long term preservation of your valuable media assets

• Partnering with an expert in the industry who currently manages some of the largest digital media archives in the world reduces risk

• Preservation and storage is the only focus of our business

• Customers such as The Library of Congress, Shoah, LDS, The Holocaust Museum, IOC, and 300+ others entrust their valuable assets to Front Porch Digital
Rescue and Preservation of Videotape Collections

Thank You!

For more information, please contact me at:
brian.campanotti@fpdigital.com