

3|2011

€ 15,-
65. Jahrgang

www.FKT.schiele-schoen.de

FKT

RTM
RUNDFUNKTECHNISCHE
MITTEILUNGEN DES IRT

DIE FACHZEITSCHRIFT FÜR FERNSEHEN, FILM UND ELEKTRONISCHE MEDIEN

**Bedeutung der Archive
für den Rundfunk**

**Langzeitarchivierung
digitaler Daten**

**Archive für
Videoproduktion und
Broadcast**

**Methoden zur
Umgehung von
IPS-Technologien**

**Bildqualitätsmessung
komprimierter
Videsequenzen**

**IRT-Symposium: Vom
Suchen und Finden**

**Standards für
digitale Archive**



Archive for Video Production and Broadcast

An archive is an investment in systematic collection, securing and administration of finished productions and assets. LTO tape represents the state-of-the-art for long-term and off-site storage. Combined usage of an LTO library for archive and backup increases efficiency and facilities early payoff. Choosing the right timing for archiving results in avoiding unnecessary restore and rearchiving processes. This article points at adequate solutions.

Archive – the unknown being

The term "archive" (latin: Archivum) originated from the Roman official building that was used for storing scriptures that were no longer needed for current affairs. The actual storage room, together with the stored documents, later became the nouns that we use today. The key phrase here is "no longer needed for current affairs", as will be clarified later.

Today, data archiving is generating a whole new list of questions. At first glance, technology has made it easier than ever to transport and duplicate data, but this advantage can quickly disappear when taking a closer look. The topic gains scant public awareness, although the passing on of culture and science to future generations depends on it. In the US there are efforts to increase the visibility and reputation of archivists.¹ Some far-reaching projects with detailed analyses, experiences and practical advice include, among others, the Digital Preservation Project, Preserving Public Television, and the American Archive Content Inventory Project.²

A Backup is not an Archive

To address a frequent misconception right from the beginning, the fundamental difference between Archive and Backup needs to be clarified. A Backup is always a duplication of data, and data that is needed for daily operation requires an additional copy for security reasons. If something gets lost, it can be restored from the Backup at any time. Old Backup runs are overwritten in a cyclic process based on a specified schedule. The size of the Backup therefore changes only minimally.

An Archive always consists of data from completed projects, transferred from the expensive online storage to more cost-effective nearline or offline storage. After a copy is transferred from its place of origin, the data is deleted. Thus, an archive is always growing.

Archive from small to large

A proven procedure in the video field was the archiving of video tapes. Often was the case that

every production was simply put onto a shelf and organized more or less through systematic labeling. In case of emergency, or so the reasoning went, we can always go back to the tape. Using file-based workflows, this is no longer possible. Occasionally this is unfortunately recognized too late, as soon as a file is deleted from a card or disk it is lost forever in most cases. But the familiar relationship of "one tape=one production" can be continued even with a digital archive.

The minimum requirement in this case is a single LTO tape drive. Using the setting to write each archive job on one separate tape (even if not filled up completely) fulfills the task as outlined above. Several advantages result from this procedure. Standard procedures can remain unchanged and shelves are available to be re-used. At the same time, there is the added benefit from the security features LTO tapes have to offer, which I'll cover later. On top of that, Archiware's P4 Archive³ allows browsing and searching the archive's index even without the actual tape being in the drive. These previews make finding the right clip even easier. Once the desired media has been found and you wish to restore, the software calls for the respective tape to be loaded. For small to medium sized environments this can be an elegant migration to a digital archive. Larger environments, where multiple users might want to restore at the same time, need hardware that changes tapes automatically – a tape library, tape robot or jukebox.

The advantages of libraries extend beyond this simplification of media management. Having multiple tape drives allows the user to write and read at the same time. Data throughput can be further optimized by parallelization where two or more drives write one data stream. At data rates of 120 MB/s (LTO4) and 140 MB/s (LTO5) for one drive, drive multiplication puts the power at your fingertips.

Systematic archive

The decision to build an archive often involves considerable investment. A long-term return on that investment is optimal, meaning a systematic

approach is favorable. Along the way many questions are to be addressed, some of them decidedly organizational in nature.

Tapeless with Tape

To suggest tape as a backup medium might sound anachronistic after spending years in building tapeless video workflows. But there are good reasons for it. To begin with, LTO tape (Linear Tape Open) has almost nothing in common with videotape. The origin of LTO specification is itself remarkable: the LTO consortium took proven features from then existing tape formats and combined them to a "best of all worlds" solution. From DLT came the cartridge technology, from AIT the memory chip in the cartridge, from DDS the error correction, and so on. The result is the most reliable and most cost effective (per TB) long-term storage medium: LTO. It even combines multiple security levels including 'write verify', checksums for automatic error correction, adaptive write speeds and a magnetic layer that is certified for 30 years.⁴

Safety First

A glance at the security features shows the dramatic difference between LTO and videotape. Servo tracks that are pre-recorded during manufacturing provide drive independent precision positioning of the heads. Thus any tape can be read in any drive reliably.

Read and write heads that are positioned in sequence provide immediate control of the written data much like read-after-write controls. In case of a defective sector, data is re-written and re-checked again automatically. Additionally, checksums are calculated for each sector. They allow for automatic correction if a dropout occurs on the tape.

When stored correctly, LTO tapes are certified readable for 30 years of storage, and are currently the only storage medium that can prove this claim through practical experiences. Security for the investment in this technology results from the combination of the consortium members HP, IBM and Quantum who manufacture drives and the existing roadmap up to generation 8 (right now LTO5).

Marc M. Batschkus
MD PhD is medical
informatics specialist.
At Archiware he is in
charge of business
development and
marketing.



Tandberg Data is also a member of the consortium, developing small to mid sized libraries alongside a number of licensees. LTO drives can always read tapes of the two previous generations (e.g. LTO5 drive to LTO3 tape), and they can write on tapes of one previous generation.

Archive: Questions for planning

1. Define users: who configures, who uses and who archives?
2. Decide and select: what goes into the archive and when?
3. Which metadata describe the data (what should the search parameters be?)
4. Actual security: on what operating system, hardware, software, infrastructure, interfaces, network and medium will you be archiving?
5. What security requirements should be fulfilled: off-site storage, duplication, storage duration, test cycles of media, generation migration, etc.
6. Retrieval:
 - Who searches?
 - With what criteria?
 - Who is allowed to restore?
 - On what storage?
 - For what use?

Optimizing Archive Capacity

Deciding at what point in time productions are archived is crucial for optimum dimensioning of the archive. If archiving happens early in the production, additional changes and edits are likely and the production has to be archived again. As a result, most of the files of the production that are unchanged will also be archived again. By archiving incomplete projects as well as their complete versions, the archive grows faster than necessary. No files can be deleted from an LTO archive (only whole tapes after an optionally configurable retention time is reached). Attention needs to be called to the aforementioned definition of the archive as "no longer needed for current affairs". The optimum point in time for archiving can be determined by analyzing the use of completed projects. If a project is unlikely to be used in the near term (minimum of a few months), then it is a prime candidate for archiving. There are two distinct advantages of this procedure: productions remain readily available when editing is likely, and the size of the archive is optimized the number of duplicate files are reduced.

Securing the Workflow

The primary storage for active productions should always be secured by a backup, allowing

for a quick restore in case it is lost or deleted. The Backup renews itself by overwriting previous backup runs after their retention time is reached. Actually, multiple workflow stages can be independently secured. A special case here is data availability; time sensitive data with a stoppage tolerance of minutes can be secured by a mirror. In case of emergency, this mirror storage can take over production immediately without a restore process.

Since an archive is a migration of data from the primary storage to an off-line storage, meaning that data will be deleted from its source, securing the archive is necessary. For this the archive job can write on two identical tapes at the same time. One of them should be kept offsite. This off-siting is important and a requirement of film set insurance policies, expressly on LTO tapes (!).

Planning the scale of archive hardware necessary can be a challenge. Tandberg Data offers a way to combine security of investment, flexibility and reasonable cost. The tape library T40+ can be extended in multiple steps from 24 tape slots to a maximum of 151 tape slots (=226TB capacity uncompressed LTO5). The drives are mounted within the base unit, and additional expansion units can be mounted on top of it and connected by a transport slot. If desired, up to four (later five) T40 units can be mounted on top of each other. They behave as one library. Thus just-in-time expansion is possible any time.⁵

Cloning as Security

Data that was archived is deleted on the primary storage, but an additional security step is necessary. During the archive process, a duplicate tape set can be created via cloning. One copy can be taken out of the library and stored off-site, so the archive is protected against the risk of flood, fire, theft, etc. A particular case of a production house shows that this can be necessary even without external catastrophe: one power supply at the base of a server rack overheated and caught fire, effectively destroying the whole rack. Only the data stored offsite survived the incident.

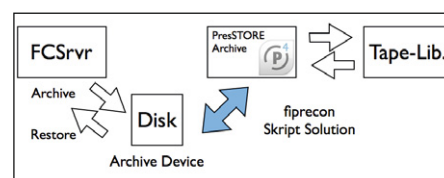
A MAM is (in most cases) no Archive

Most of the larger work environments today use a media asset management (MAM) system to administer and search available assets. Sophisticated search queries can be used and media checked out for editing to prevent double effort. Since this implies investment usually in the six-figure range, smaller production companies tend to shy away from it. Thorough examination of the whole field can be found in David Austerberry's

work.⁶ Asset management is useful in smaller environments as well and some solutions like Apples Final Cut Server, CatDV by Squarebox can also be implemented with a modest effort.

In most cases asset management is no archive. Actually Final Cut Server as well as CatDV can save to disk only. Therefore the question needs to be raised if both can be combined and in what scope. The archive software P4 Archive already offers asset management with previews and search options via the browser. Metadata can also be added to the assets. If requirements remain limited this software can fulfill both requirements. Additionally far-reaching options to save data are available. Manual archiving as well as automatic archiving is possible. Configurable login areas allow users to browse/search or restore according to their privileges. As part of a complete data management suite, P4 Archive seamlessly combines with P4 Backup and P4 Synchronize for a complete workflow solution.

Beyond that a combination of FCSrvr+P4 Archive or CatDV+P4 Archive can fulfill more complex needs concerning management and security. This combination seems natural since both MAM solutions are unable to archive to tape. The MAM system puts files to be archived in a specific directory on disk where a script "picks them up", writes them to tape, then deletes them after completion. This way storage space gets freed up on the expensive online storage and gets moved to the less expensive offline storage.



Graphic 1: Final Cut Server integration with P4 Archive, fiprecon and tape library

Tape is the most economic storage medium per TB as well as the safest, since tapes that are not in the drive mechanism are completely secure from viruses, malware, deletion and other accidents. The German broadcast specialist André Aulich developed a script solution called fiprecon and has made it available for FCSrvr and P4 Archive.⁷

Archive at Work

Meta Media Creative Technologies (<http://metamediotech.com>) offers archive installation and services. They count the Major League Baseball Network MLBN (<http://mlb.mlb.com/network/>) among their customers. All media from their site is held in a Final Cut Server. Since FCSrvr has no tape archive option, P4 Archive by Archiware is

used to extend the functionality with an LTO tape archive. Every night, search procedures run to find the oldest media. Assets that have not been used for a specified number of days are collected and moved to a specific disk where a script hands them over to P4 Archive. In one single archive job they are written on a Spectra Logic T50e library. After 30 days the copy on the disk is deleted, so that from that moment on, the file only exists on tape. If an editor needs the file, they select it in FCSrvr to be restored. The fibrecon script triggers the restore process of P4 Archive. As soon as the file is read from tape FCSrvr, the editor has it as their disposal. During this process, the interface of P4 Archive is not shown to the user. He works only within the FCSrvr interface.

Spoiled by Choice – the right archive format

What is the right format to store productions? This question implies another one: how long should the archive save the data? Looking at either end of the extremes might help. There are production houses, postpro specialists, 3D and SFX companies that rarely want to archive for more than 5–10 years. On the other end of the spectrum there are public TV stations with the expressed mission to

safeguard cultural assets. In this case, they want to store for at least a couple decades. In the first example the production or delivery format is a good choice. If need be a file can quickly be integrated into production again. Since .AVI as well as .MOV, Final Cut and Avid are directly connected to the respective manufacturer, there is also a direct dependency on their plans and operations. This was the reason for SMPTE to develop an independent standard that was also supposed to ease the communication between systems. The result, MXF (Material Exchange Format) has been available for number of years. Sony and Panasonic among others have adopted it for their products. No available operating system can read or write MXF out of the box. But there are easy solutions available like MXF4mac by Hamburg Pro Media, which make the operating system and applications capable of using MXF.⁸

Metadata precisely

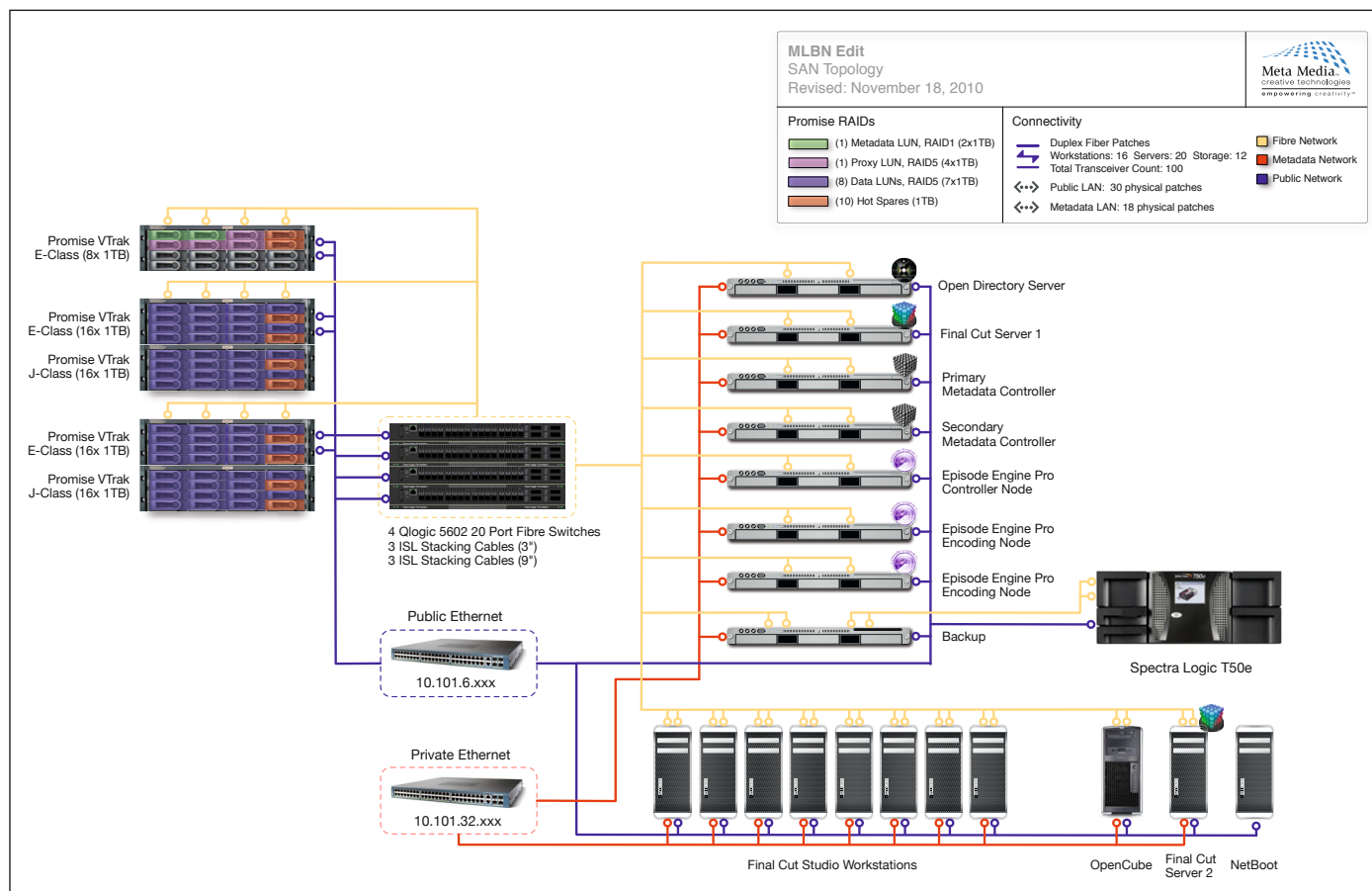
In order to retrieve a video clip long after its creation, it is necessary to be able to search for descriptive criteria. These metadata can either be technical (file name, time, format, codec) or descriptive (actors, elements and scene descriptions). The term metadata is new but librarians

have been working on descriptions of books for centuries, yet their experience and systems are relevant and helpful for the description of video as well.

Proven through extensive use has been the Dublin Core standard, which tries to facilitate findability through defined fields. It is in continual development.⁹

Library archives are targeting standardization named EAC – Encoded Archival Context – and is an XML format.¹⁰ Specifically addressing the needs of broadcasters is the Public Broadcasting Metadata Dictionary Project.¹¹ Metadata can also facilitate the exchange among organizations like public broadcasters and educational institutions.

One warning should be issued at this point: metadata is NOT a pure technical topic. Quite the contrary; it is mainly an organizational topic. Relevant questions are e.g. who decides which form of description is used in general and on a detailed level? Who formulates the metadata? Who oversees its consistent application? An organization has to face these questions and more in order to be able to implement useful metadata oriented for future use. The University of Utah alone has been working for years on these issues in all their complexity.¹²



Graphic 2: Large scale integration: Setup of MLBN using Final Cut Server, P4 Archive and fibrecon.

Archiving pays off

When comparing costs of saving media to disk or to tape a considerable difference is seen. Saving to disk cost more per TB than saving to tape. Costs for additional RAID's, their corresponding power consumption, cooling requirements, and infrastructure to integrate the extended capacity into the setup are to be considered. Tape uses only about 10% of the electric power of disks and needs no additional investment in cooling. One library can write an almost infinite number of tapes if older tapes are taken out of the library. The potential for savings is considerable. All in all, the investment in an LTO-based archive pays for itself within one to three years depending on the specific situation.

Archiving and the law

When planning an Archive, the compulsory archiving of business data can be easily integrated. Stringent archiving requirements exist on international, national and also local levels.¹³

The easiest technical solution for archiving unalterable business data is using WORM tapes. These tapes are specified so that they can only be written once (though in multiple parts). Data can never be erased from them. Saving relevant business and financial data on it provides a reliable and extremely cost effective solution. Written WORM tapes are simple to store in a safe location.

Summary

An archive is a fundamental investment covering the systematic collection, safeguarding, administration and re-use of completed productions, media and data. Choosing the appropriate point in time for archiving avoids unnecessary restore processes and double archiving of unchanged assets. Collecting requirements and analyzing the workflow are prerequisites to construct an optimal solution. Evaluating the need for promoted expensive and often complex features needs to be carefully executed. A straightforward and transparent solution that can be maintained in-house has its advantages. Saving and off-site storing of LTO tape is state-of-the-art at present without alternative for long-term storage. Combined use of an LTO library for Archive and Backup increases usefulness and ensures faster amortization.

- [1] Society of American Archivists SAA
<http://www2.archivists.org/>
- [2] Digital Preservation Project
<http://www.digitalpreservation.gov/>
Preserving Digital Public Television Final Report 2010
<http://www.thirteen.org/ptvdigitalarchive/files/2010/07/FinalReport.pdf>
American Archive Content Inventory Project
<http://americanarchiveinventory.org/project/>
- [3] PresSTORE P4 by Archiware
<http://www.archiware.com>
- [4] Technical data, specifications, roadmap and tests can be found with the LTO consortium
<http://lto.org/>
- [5] Tandberg Data scalable LTO library T40-T160
<http://www.tandbergdata.com/us/index.cfm/products/tape-automation/storagelibrary/storagelibrary-t40/>
- [6] David Austerberry: Digital Asset Management
<http://www.davidausterberry.com/>
- [7] firecon script solution by André Aulich to integrate P4 Archive with Final Cut Server
<http://www.andre-aulich.de/en/perm/connecting-final-cut-server-to-archiware-presstore>
Video shows solution at work:
<http://www.archiware.com/pics/FCSrvrArchiveIntegration.mov>
- [8] MXF4mac by Hamburg Pro Media
<http://mxf4mac.com>
- [9] Hillman, D. 2001. Using Dublin Core, DCM Recommendation
<http://dublincore.org/documents/usageguide>
- [10] EAC – Encoded Archival Context
<http://eac.staatsbibliothek-berlin.de/>
- [11] Public Broadcasting Metadata Dictionary Project:
<http://pbcore.org/2.0/>
http://pbcore.org/PBCore/PBCore_Tools.html
- [12] University of Utah Media on Demand Wiki
http://library.med.utah.edu/wiki/MODwiki/index.php/Main_Page
Media Exchange Project
http://www.pbcore.org/PBCore/PBCore_CaseDocs/Case_MediaExchange.html-Metadata: Describing Media Assets to Improve „Findability“
http://library.med.utah.edu/wiki/MODwiki/index.php/Metadata:_Describing_Media_Assets_to_Improve_%22Findability%22
- [13] Federal Rules of Civil Procedure (FRCP) <http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2010%20Rules/Civil%20Procedure.pdf> 'Records management and ISO 15489:
http://en.wikipedia.org/wiki/Records_management
ISO 15489-1:
http://www.archives.gov.ua/Eng/Law-base/ISO_15489.pdf

Originally published in German.
FKT is the magazine of German "Fernseh- und Kinotechnische Gesellschaft" (TV and cinema technology society)
All rights reserved – © Copyright 2011 by Fachverlag Schiele & Schön GmbH, Berlin
Translation of article originally published in German.
Translation: Marc M. Batschkus