FORWARD INTO THE PAST—PROTECTING OUR MUSICAL HERITAGE*

An Architecture Guide for the Music Community

James Fleming
adeo...Solutions for the New Entertainment Economy
Washington, DC 20037, USA

A broad overview is given of the key features and issues that need to be considered when developing and building a successful multimedia archiving system. Technology changes are taking place at an ever increasing rate, and the critical system decisions necessary to embrace this acceleration of change are explained. Complex storage systems of this type can become difficult to maintain at optimum performance, and it is shown how effective management can resolve this potential problem. In addition, various approaches are discussed that can help ensure long-term viability of a deep corporate archive and asset management system.

INTRODUCTION

In 1985 "Back to the Future," starring Michael J. Fox, was a very popular movie. The basic plot involved going back in time to the 1950s to ensure one's very survival in the present day. Our hero is successful in his quest, and returns triumphant to the 1980s. An entertaining story, with a happy ending. Conversely, archiving of rich content requires moving forward into the future to protect the past, the rich past of our musical heritage. What plot line is required here to achieve a successful outcome? Many companies have invested millions of dollars acquiring archives, containing a wide array of content, from music, the spoken word, to photos and movies. They want a return on that investment, in a timely manner. They want to know that "moving forward into the past" has a successful conclusion. Unless the raw assets are stored and protected in a manner that permits their subsequent efficient repurposing the investment will be useless and there will be no heroes.

With the advent of high-performance computers and cheap mass storage we can now store and maintain vast quantities of digital material but how do we know what to store? Digital music assets are being generated at an enormous rate in the early part of the twenty-first century. For example, more artists are entering the charts, videos are becoming a key part of success, and Internet content is being created and delivered as part of the value chain. So what issues are important in integrating a philosophy of long-term asset protection and effective management with a marketplace that is becoming more and more fragmented?

KEY POINTS

Archiving requires a management mind-set that focuses on planning for 50 years and longer, not just the next couple of quarters. Shareholder benefit is created over the long term. Allocate the appropriate human resources to the project.

• Store all music at the highest sampling rate and bit depth possible. This ensures that you store the best quality of the music for the benefit of future generations.

• Know your existing company process prior to introducing any complex asset management tool. Technology cannot streamline or fix a system that is not well understood.

• Never throw away any hardware or software unless you are sure all assets can be readily accessible. This approach covers everything from client computers to servers, to analog tape playback machines, and storage devices.

• Always have an effective disaster recovery plan that will ensure the assets are protected, regardless of system breakdown.

• Develop a core group of individuals within your company to develop and manage the enterprise wide archiving and asset management system. Give them the resources they need to produce an efficient and appropriate solution. This will help in the effective repurposing of all assets required for music and graphics production.

BROAD GENERIC ARCHITECTURE

A typical archiving system can be split up into certain key partitions, with standard lines of demarcation between them. Thus any component part can be replaced without major disruption to the system in general. This is one major benefit to an open architecture approach compared to a closed vendor-specific solution. Each section can then be fully analyzed and the most appropriate solution found.

Acquisition and Conversion

Always keep conversion devices, such as analog to digital and digital to analog external from any computing or storage device. Low-level system noise and clock jitter have an adverse effect on digital music. Also this gives the studio personnel a wide range of equipment to choose from, especially with the advent of surround sound and high-quality two-channel audio. During the process of conversion from the physical asset to the digital one, crucial track timing, track name, and PQ information is likely to be readily available. It is imperative that such metadata be entered at this stage of the process. ➥

*Manuscript received 2001 June 24; received 2001 June 29.

Waiting until later in the archiving cycle can become time-consuming and costly.

Local Storage

Many client computers are self-contained and not part of a network. Subsequently, removable storage has become very popular in the last few years and is a very efficient method of data transfer. Having an 18-Gbyte disk allows the archiving of about 30 albums. Typically, the transfer engineer would take at least double the time of the selection. Thus this single inexpensive removable disk would contain a minimum of 60 hours of engineer activity. Music and library archivists often find that transfers tasks require an average of 3 to 4 hours of engineer time per hour of source. Thus we can quickly see how expensive a disk crash would be. Going to larger disks therefore increases the exposure to failure.

Online Server-Based System Storage

Although a music archive will keep all of its assets offline, these assets need to be readily available online for easy and quick access. Therefore adequate disk space must be allocated. The increased disk size and reduced cost of RAID (random-access inexpensive disk) now makes it possible to consider purchasing anywhere from 500 Gbytes to 1 Tbyte of disk for online storage. There are a number of RAID configurations available and tradeoffs may be made between required reliability and cost. Depending on the number of clients, and the total transactions taking place per hour, this available space may fill up quickly. For example, a typical archive, supporting music and graphics, could have requests for 20 albums, be archiving 50 albums of music with the associated graphics, and have numerous requests for other graphical assets. Thus there could be 100+ Gbytes of data being stored on disk during the transfer process. If we include all the asset-tracking databases necessary, we see that server disk allocation needs to be carefully thought through. After a system has been in place for a number of years it will naturally contain many Tbytes of data in various forms, and to have to modify and reallocate mass storage will be costly and could be disruptive to the daily workflow and needs of a company.

Distributed Storage

With the advent of distributed client–server systems designers found that massive amounts of data were often divorced from where they were needed. Thus networks became clogged as these data were being transferred. This limitation in system performance has been addressed by storage suppliers, and now Network Attached Storage (NAS) and Storage Attached Networking (SAN) solutions can be possible effective solutions for system developers. In general we always need to limit the number of read/write processes taking place.

Long-Term Archival storage

Any thorough analysis of storage technologies will indicate that tape is the most cost-effective mass storage media available today, and indeed there are a number of companies providing automated robotic tape based mass storage systems. These provide relatively infinite amounts of storage in a small footprint. At present tape sizes are nominally 50 Gbytes per cartridge, and the tape manufacturers are predicting migration paths up to 200 Gbytes. Critical issues to be considered here are the input–output activity, which will define the number of drives required and the drive data transfer rate.

LINES OF DEMARCATION

By having clearly defined lines of demarcation, which follow known standards, the designer can easily upgrade any portion of the system for improved performance when desired. We can define various lines for both software and hardware.

Hardware Demarcation

Asset acquisition: Use standard AES/EBU interfaces or Firewire between external interfaces and the chosen digital audio workstation.

Communication: Many standards exist, such as FDDI, Fiber Channel and Ethernet, 10 baseT, 100 baseT, and Gigabit. These are all serial implementations, whereas a SCSI standard used for disk transfers is parallel.

Storage Systems: External RAID solutions provide self-contained mass storage which, separate from the server computing platform, allows upgrades and additions to be done while not affecting the base server installation.

Software Demarcation

Disparate Operating Systems: They must communicate with each other in distributed client–server systems. For example, a Windows or Macintosh platform linked to a UNIX-based server is a common solution today. Therefore file transfer testing must be performed between the machines to ensure that there are no compatibility issues.

Application Software Updates: In complex distributed systems it is impossible to test every possible permutation of performance. Thus when new versions of application software are introduced, it is imperative to run a suite of test routines with known data to fully check out the new application. It could be troublesome not to check a system update fully and find subsequent problems with the archived data.

DATA AND ASSET ACQUISITION

The cost of storage has come down dramatically over the last 10 years. Therefore it is imperative, as mentioned, to acquire assets at the highest quality possible. Existing physical assets will typically include:

- Analog tape
- Digital tape (1630s and DAT)
- Metal disk masters
- Disk lacquers
- CD-ROM
- DVD audio and video
- SACD

With original digital assets we have no choice but to be limited to the initial sample rate and sample size. However, resampling and interpolation techniques may prove to have some benefit. Typically the red book standard gives us a sample rate of 44.1 kHz and 16-bit sample depth. This has rightfully

been criticized as a format, and is not appropriate for archiving.

Sampling with standard linear PCM (pulse-code modulation) at 96 kHz and with a sample size of 24 bits results in improved quality. Unfortunately many mastering engineers and equivalent critical listeners can easily distinguish between a good analog recording and this resultant digital file.[1] Therefore, where possible, it is suggested that studios and companies investigate the viability of archiving either at 176.4 kHz (4 times red book) or at 192 kHz.

1630 tapes have been a popular choice of media for final two-track production masters. They produce an excellent sound which some individuals prefer over a CD-ROM. However, they need to be evaluated carefully and the data archived for protection because the tape and the coding method used were never intended as an archiving standard. We found them to be unreliable over time.

Much has been written about the precautions necessary for storing physical assets. For example, analog tape can be susceptible to very humid conditions. Suffice it to say that any physical archive must be checked thoroughly for absolute physical conditions with a cataloging database developed to ensure a complete record of the assets.

Disaster Recovery

Any large-scale system must be designed against possible disaster and major data loss. Thus all near-line tape cartridges need to be backed up, with the back-up tapes stored off site in a high-security facility. The power systems should be backed up to protect from both blackouts and brownouts, a brownout being defined as a transient loss of power. This is an important consideration within global affiliates where the power may not be as reliable as in the United States. Any archiving installation must be in an environmentally controlled location to ensure that all tape storage cartridges are stored under optimum conditions.

[1] The Harvard Loeb Music Library has archived a good deal of band-limited material to 96 kHz, 24 bit and they report that listeners are unable to distinguish the digital playback from the analog source. The widely varying empirical reports call out for formal psychoacoustic testing.

PROJECT LIFETIME ISSUES

It appears paradoxical that companies are building tape-based silos to replace physical assets which contain tape in the knowledge that the tape is not a viable long-term storage media. So how do we resolve this problem and use tape as a building block for a long-term, deep archive? The key to this decision is in current digital signal processing and digital storage formats. All tape drives incur raw errors during the read and write processes. These are minimal and may be one error in 10 000 bits of data. Therefore coding designers developed a method of minimizing this problem: all data stored on tape have additional data which can both detect and correct the actual raw digital information. Therefore the apparent error rate drops to approximately one in 10^{12}. What we need is to keep track of this error rate and to make a copy before the raw errors become uncorrectable. Tape-based silo vendors must make this feature available to ensure the long-term protection of any tape-based mass storage system. The system would automatically check each tape and keep track of the raw errors for all tape cartridges. Long before this value reaches a critical level, the data can be cross-copied to a new tape cartridge.

Data Format, Naming Conventions, and Metadata

File formats are crucial for any long-term archive, and standardizing on common ones is important. Popular formats are WAV, AIFF, and DDP. WAV files are common with PC-based audio work stations, and the AIFF format for Mac-based machines. DDP (disk descriptor protocol) has been developed as a standard for supplying masters to CD manufacturing plants. This is becoming popular as a common format to be used both for physical disk manufacturing and for electronic music delivery projects. Also AES31 is gaining industry support. One thing to mention is the importance of EDL compatibility. System integrators should consider making sure that EDLs containing vital PQ data, fades, and track listings can be transferred between audio work stations.

Metadata, or data about data, often called associated data, needs to be an integral part of the data entry process in any successful archive, and the earlier the better. Record companies have all had to pay up when assets used were found to have no rights for use. Any data structures used should be extensible and an approach of graceful degradation should be followed, that is, file names should allow for some understanding of the content if a database becomes inoperative.

MANAGEMENT ISSUES

Complex enterprise-wide systems, such as an archive system, protecting a company's core assets, embrace hardware and software probably supplied from a myriad group of vendors. The main focus of the system integrator is to take these diverse solutions and provide a seamless, efficient mass storage system satisfying the daily short-term and more long-term needs of the project. Management plays a crucial role here and must allocate the intellectual and psychological resources necessary to support this asset protection initiative. Often the elegant solution is arrived at through empathetic dialog and synergy between departments which have formerly had very little interaction. That is, an integrated, balanced approach must be encouraged, which will embrace not just the hard science of logical thinking but also a more intuitive style. Many companies have vast treasure troves of musical heritage which are in immediate need of attention. The management techniques which created this situation therefore cannot be used to provide solutions to this on-going problem.

Traditionally companies have been used to operating alone in very competitive markets, without the need to develop and nurture close relationships with other businesses. The intellectual property and centers of excellence built up by vendors and suppliers need to be utilized by system integrators much more than in the past. Often major changes made in application software or hardware can introduce unforeseen problems, which require immediate attention. Working closely with technical support groups and service centers can help reduce the time spent in problem solving. Over the long life of an archiving project this can save both valuable time and financial resources.

The people who run asset management and archival systems are providing a valuable service to the creative staff for many kinds of multimedia projects. They must remain attentive to the needs of studios and art departments, providing them with assets in a timely manner, hence helping to make the creative process more efficient.

CONCLUSIONS

This report covered many of the crucial issues involved when building a long-term archive. By considering and building on these points, companies can embark with confidence on their journey "forward into the past." However, each individual company must review their own specific needs carefully prior to developing and building a long-term archive.

THE AUTHOR

James Fleming, in his previous position as director of technology for the Recording Industry Association of America, was responsible for working with business and technology leaders of each major record label to assess, interpret, and evaluate various technological developments that may effect the worldwide recording industry. Additionally, as a member of the Executive Committee of the Secure Digital Music Initiative (SDMI), Fleming had extensive interaction with executive-level management of numerous major companies in the information technology, consumer electronics, telecommunications, and entertainment industries. He was a voting member of the U.S. National Body to the Motion Pictures Expert Group (MPEG) Standards Body. Before joining the RIAA, he specialized in large-scale embedded system development and project management.

Email: James@adeo-thinks.com