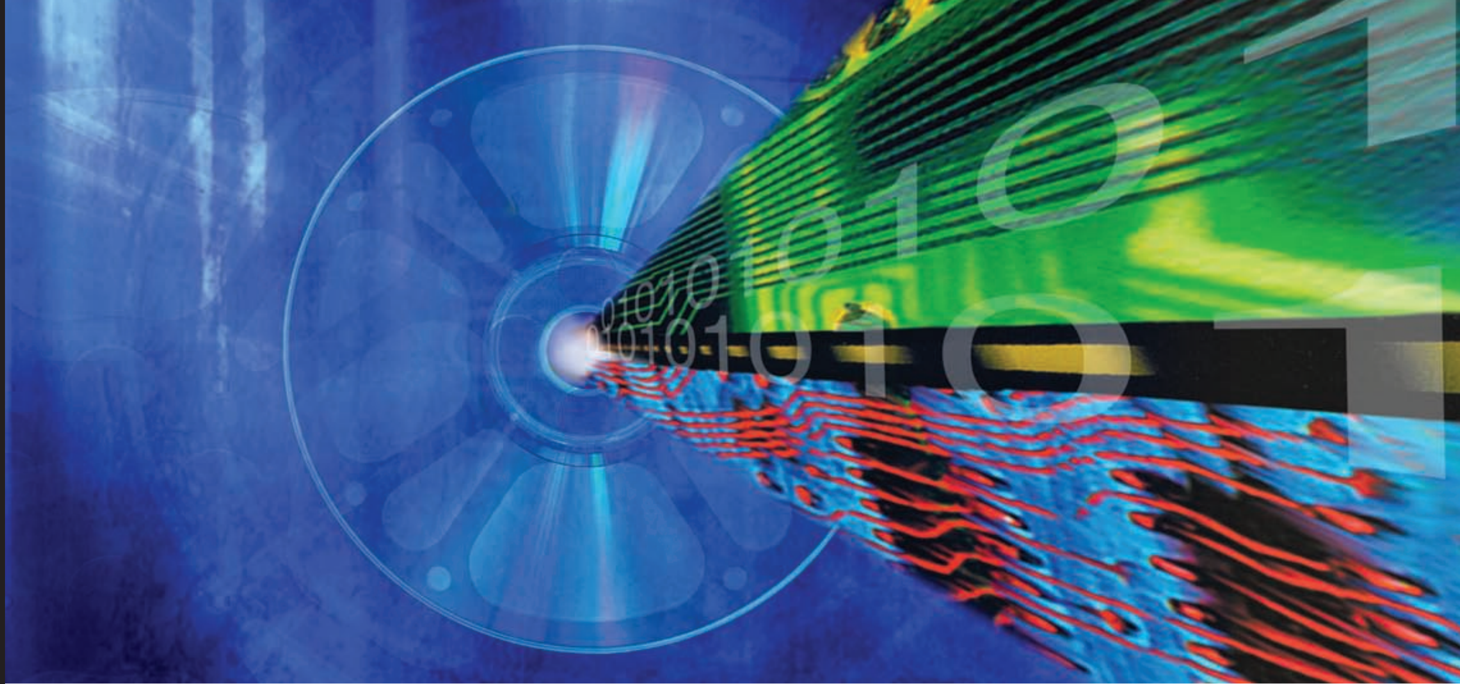


Science and Technology Council

THE DIGITAL DILEMMA

STRATEGIC ISSUES IN ARCHIVING AND ACCESSING DIGITAL MOTION PICTURE MATERIALS

THE SCIENCE AND TECHNOLOGY COUNCIL
OF THE
ACADEMY OF MOTION PICTURE ARTS AND SCIENCES

Science and Technology Council

THE DIGITAL DILEMMA

STRATEGIC ISSUES IN ARCHIVING AND ACCESSING DIGITAL MOTION PICTURE MATERIALS

TABLE OF CONTENTS

Table of Contents

Table of Contents

Foreword

DIGITAL TECHNOLOGY IS ALREADY PRODUCING SIGNIFICANT benefits for the motion picture industry. As evidenced in image capture, visual effects, mastering and final color grading; in sound capture, sound effects, and sound editing and mixing; and in the continually increasing digital distribution to theaters and other platforms, the digital era is not approaching – it's here.

However, the changes have tended to arrive piecemeal, and so rapidly that the industry has not yet had a chance to step back and consider the digital revolution and its long-term implications as a whole. Even some of the artists who have been the most evangelical about the new world of digital motion pictures sometimes seem not to have thoroughly explored the question of what happens to a digital production once it leaves the theaters and begins its life as a long-term (if all goes well) studio asset.

To date there have been no definitive studies comparing current costs of digital or hybrid systems to those of the analog photochemical systems that have long been the standard in Hollywood. The long-term preservation of, and convenient access to, a company's cinematic assets is clearly going to be an ongoing concern, and yet a danger exists that in an effort to stay on the curl of the digital wave – an effort not surprisingly encouraged by the vendors of digital technologies – the industry may make decisions that produce unfortunate financial and cultural consequences. Herein lies the Digital Dilemma.

This project originated two years ago when Phil Feiner, chair of the Digital Archival Committee of the Academy's Science and Technology Council, proposed convening a "summit" that would for the first time bring together archivists and senior technologists from the Hollywood studios and those charged with the preservation of moving images and recorded sound by universities, the U.S. government and other organizations. That summit led to the realization that the marked acceleration in the use of digital systems was not being accompanied by appropriate planning, or even in some cases by a full understanding of the potential impact of the digital revolution.

The Science and Technology Council subsequently surveyed experts in the field – from studio executives and technology department heads to those charged with preserving medical, military and geographical data – and collected detailed information on the issues. This report, defining the issues that the motion picture industry faces with respect to long-term storage of and access to digital motion pictures and other digital assets, is the first in a series of Academy studies.

As an organization historically concerned with the art rather than the business of motion pictures, the Academy is appropriately concerned primarily with the cultural consequences mentioned above. But because the business decisions that companies make about how to preserve their cinematic holdings, and about how much of them to preserve, have clear consequences for the art of motion pictures, this study falls very much within the Academy's mission.

Foreword

While there have been several well-researched and informative papers on the problems associated with digitizing existing media archives and on digital data preservation in general – and we liberally reference some of those works here – none have examined the topic from the unique perspective of the Hollywood studios, a perspective that developed over a 100-year period.

From this perspective, it's clear that a totally committed, binding switch to digital has one major drawback: the absence of guaranteed, long-term access to created moving image and sound content.

"The Digital Dilemma" is designed to bring industry executives up to date on major technological changes that are affecting and will continue to affect how content owners create and manage their digital motion picture materials. The replacement of analog film systems with digital technology has a significant impact on costs, operations, staffing and long-term access. But the motion picture industry is by no means the only one wrestling with these issues. As this report demonstrates, the federal government, the medical profession, astronomers and other scientists, the military and other entities are all struggling with remarkably similar issues. Through our research, we endeavored to learn what is happening now, what problems they have encountered, what they foresee, and what plans, if any, they are making to accommodate the changes that come with digital storage technologies, as well as the unintended consequences of those changes.

It is a study that offers more questions than answers. But the questions are enormous ones and they need to be addressed very soon by the motion picture industry as a whole, starting with those in the key corporate decision-making positions. We offer this report as a call to action to generate fruitful collaborations and workable long-term solutions.

MILT SHEFTER, ***Lead,*** *Digital Motion Picture Archival Project*
ANDY MALTZ, ***Director,*** *Academy Science and Technology Council*

A NOTE ABOUT SOURCES
Many senior and staff-level employees of the major Hollywood studios, laboratories and archive facilities spoke openly and candidly about what is going on in their organizations and what they see happening around the industry. They also provided us with their personal views of the issues of preservation of and access to digital motion pictures. We chose to encourage the beginning of a productive industry-wide conversation by providing a safe environment to express the unfettered views and facts as seen by the "boots on the ground," and in support of that openness we chose to leave this information unattributed.
-Ed.

1 Executive Summary

IN THE MOTION PICTURE INDUSTRY, THERE IS A MAJOR DIFFERENCE between an archive and a library. The archive holds master-level content in preservation conditions with long-term access capability. A library is a temporary storage site, circulating its duplicated holdings on demand. An archive that stores digital materials has long-term objectives. By current practice and definition, digital data storage is short-term.

For Hollywood studios, the "library," or their collection of titles, is arguably one of their largest and most valuable assets. For most of the last 40 years, and in many cases longer than that, they and other content owners stored all motion picture film records – original camera negative through final release prints – not throwing anything away. The "save everything" strategy was possible because of the low cost of storage and long-term life of film and its supporting photochemical technology. Film assets also served content re-purposing, even for distribution channels and markets unknown at the time the film materials were created and saved.

In contrast, digital data practices generate much greater amounts of material, and currently very little of it is preserved. The digital master, created during the Digital Intermediate process, is recorded to very stable yellow-cyan-magenta (YCM) separations on black-and-white film with an expected 100-year or longer life. However, this preserves only that singular version of the created content. The digital equivalents of "B neg," trims and outs, and other ancillary materials available and commonly used for non-theatrical distribution, are not saved as film but as digital data that needs to be actively managed or "migrated" to new digital media formats every few years.

The exploding use of digital technologies in acquisition, postproduction and distribution raises new issues related to production workflows, organizational responsibilities and business models. Data explosion also comes with the threat of data extinction and, therefore, the loss of valuable content. With a single digital motion picture generating upwards of two petabytes of data – the equivalent of almost half a million DVDs – the decisions as to what materials to hold, what to preserve and what risk management decisions are needed before the migration decision, all place new pressures on management.

Current practices in other sectors such as medical, earth science, government, corporate businesses and supercomputing have spotlighted two major findings of interest to the motion picture industry:

1. ***Every enterprise has similar problems and issues with digital data preservation.***
2. ***No enterprise yet has a long-term strategy or solution that does not require significant and ongoing capital investment and operational expense.***

Experience in the above sectors underscores the fact that ongoing labor and energy costs add significantly to the total cost of ownership of digital materials. Economic models comparing long-term storage costs of film versus digital materials show that the annual cost of preserving film archival master material is $1,059 per title,[1]

[1] *Based on a monthly cost of 40 cents per 1,000 foot film reel in preservation conditions plus the amortized cost of film archive element manufacture.*

and the annual cost of preserving a 4K digital master is $12,514,[2] an 11-fold difference. The annual preservation costs for a complete set of digital motion picture source materials also are substantially higher than those for film, and all digital asset storage requires significant and perpetual spending to maintain accessibility.

Advice from the above sectors includes not allowing the equipment manufacturers and system designers to continue to foster technology obsolescence as they did in the television industry and are now doing in the information technology realm. Instead, the stakeholders must be the driving force.

There is an urgent, historically justified opportunity for content owners and archivists to manage the transition from current to future practices. This is best accomplished while film preservation can still be done in parallel, and essential digital assets that are not suitable for film preservation are small in number and relatively young. Furthermore, the task of preserving digital assets is too large for isolated or piecemeal efforts.

The primary challenge for proponents of digital systems is to meet or exceed the benefits of the current film system. These benefits include worldwide standards; guaranteed long-term access (100-year minimum) with no loss in quality; the ability to create duplicate masters to fulfill future (and unknown) distribution needs and opportunities; picture and sound quality that meets or exceeds that of original camera negative and production sound recording; independence from shifting technological platforms; interoperability; and immunity from escalating financial investment.

The risk management decisions about what digital materials to keep, migrate, or otherwise manage must consider the broad set of issues inherent with digital storage technologies. The passage of time will inevitably determine the cultural value of assets, but economics will force an ongoing assessment of the future financial value of assets each time a major data migration is considered. The risk management decisions cannot be postponed until the data migration deadline arrives. Digital archiving is an enterprise-wide consideration that requires support at the highest level to be successful.

This report is a call to action for the motion picture industry to understand the issues, clearly define the problem, and create discussion among all the major stakeholders to produce standards and technological alternatives that will guarantee long-term access of digitally created motion picture content. To this end, the Academy has initiated several collaborative projects, which include:

- *research on related digital preservation issues and potential solutions*
- *development of digital file formats for acquisition, mastering and archival applications*
- *development of a digital preservation case study system*
- *facilitating productive dialogue among the stakeholders*

The digital dilemma arrived with the digital era. It demands concerted, committed industry action.

[2] *Based on an annual cost of $500 per terabyte of fully managed storage of 3 copies of an 8.3 terabyte 4K digital master.*

2 Archiving

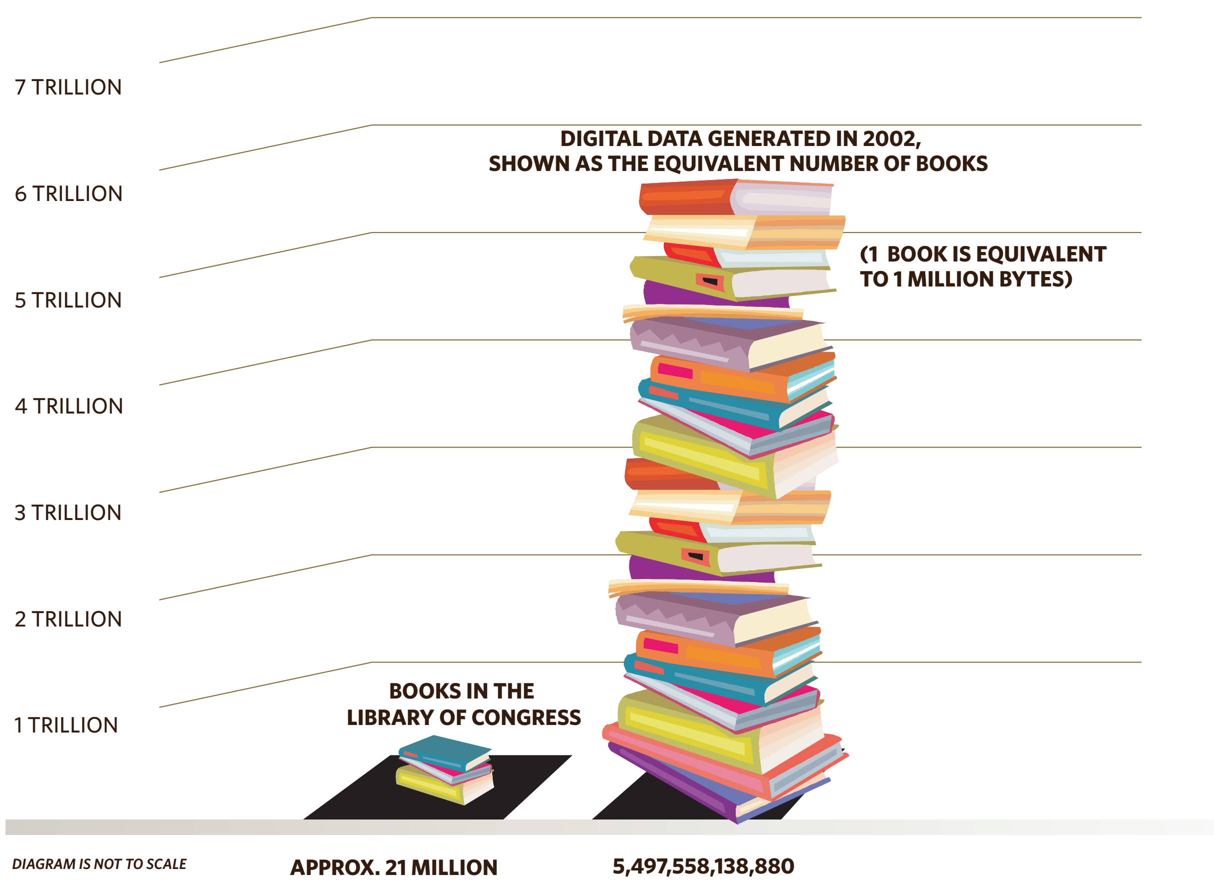
The goal is preservation without errors, access without end.

ARCHIVING HAS A LONG HISTORY IN HUMAN SOCIETY. KEY INSTITU-tions in every society in every era have invested in the long-term preservation of records and other objects deemed important by that society at that time. From pre-history until the present age, all archives consisted of "things" that exist in the physical world, preserved in the physical media of each era – papyrus, parchment, paper, leather, canvas, wood, stone, ceramic, metal, silk, photographic plates, sheets and rolls of film of various gauges and specifications.

An archive is not just a collection of old content. An effective archive integrates its holdings with up-to-date catalogs, indexes and other tools needed to search and retrieve assets stored in it. Archiving purposes vary from domain to domain, and from community to community, but, in general, archiving is meant to systematically collect and protect assets considered valuable enough to keep "for the future." Ideally, the contents of the archive must be reliably authentic, accurate and complete. The goal is preservation without errors, access without end.

Enter digital information: according to a 2003 study by researchers from the School of Information Management and Systems at the University of California, Berkeley [Lyman and Varian 1], the world generated 5 exabytes – the equivalent of 5.5 *trillion* books – of new data in 2002, stored in four physical media: paper, film, magnetic and optical – that are seen or heard primarily through four electronic channels: telephone, radio, television and the Internet. The UCB report estimated that the 5 exabytes of information recorded on storage media comprised less than one-third of the total volume flowing through telephone, radio, TV and the Internet in 2002.

How Much Data, 2002 – Book Equivalent



Digital media are adding to the explosion of data in the world. This data explosion also comes with threat of data extinction.

> *"In the excitement about the solutions digitization offers, the right questions about costs are often not asked, especially about long-term costs for keeping the digital files alive. This enthusiastic attitude is risky, for the conversion process to create the digital files may well be quite expensive to start with, and these investments may turn out to be wasted if planning for the future is ignored and no structural funding for maintenance is secured.*
>
> *Without such long-term planning, digitization projects can come to behave like black holes in the sky. Scanned information, which in the analog world could be accessed simply by the use of our eyes, is suddenly stored in an environment where it is only retrievable through the use of technology, which constitutes a constant cost factor. The more information is converted, the more the costs for accessing it go up. The digital black hole has got its firm grip on the project. It will go on swallowing either money or information: the funding must be continued or the input will have been wasted. If funding starts to fade, the information may still be retrieved but after a while it will no longer be accessible due to corrupted files, or obsolete file formats or technology. Then the digital information is lost forever in the black hole."*[3]

The problems of "data extinction" are only growing as more and more aspects of human activity move to the digital domain. Consumers have to move (migrate) their downloaded digital music to new media players when their old players get too full, sometimes requiring re-registration of their Digital Rights Management (DRM) authorizations to insure they do not lose access to any favorite songs. Authors must find current applications that are interoperable with their old word processing software in order to read manuscripts originally written with software that has since become obsolete. Digital photographs recorded on old floppy disks cannot be accessed on modern computers, which no longer have floppy disk drives. The only way to play old video games is to keep the old game system hardware running, which often requires scouring flea markets for old circuit boards that can be cannibalized for obsolete parts. Modern digital data – the media on which it is stored, the hardware needed to play it and the applications that use it – are all changing at a rapid pace. In the face of these challenges, preserving digital data and assuring its accessibility over the long term requires a systematic process that is generally described as "digital archiving."

[3] *From "The Digital Black Hole" by Jonas Palm, Director, Head of Department of Preservation, Riksarkivet/National Archives, Stockholm, Sweden.*

ARCHIVING

2.1 History and Characteristics of Hollywood Film Archiving

The "library" is one of the most valuable assets possessed by a studio.

MANY CINEMA ARCHIVES AROUND THE WORLD CONTINUE TO operate as "public" archives, such as those at the Academy of Motion Picture Arts and Sciences, UCLA, the Library of Congress, the Museum of Modern Art, Eastman House and others. The creation of "private" archives owned by corporations for the purpose of making money is a relatively new phenomenon in archiving history.

But in Hollywood, private film archives have emerged as valuable corporate assets that can appreciate in value over time and can be bought and sold for large sums. The "library" is one of the most valuable assets possessed by a studio. Assets are preserved so they can be exploited to create new media products for future markets. Making new revenue from old assets is a very profitable approach when it can be done without incurring undue new costs in adapting the old media format to the new market demand.

The explicit commercial motivations of the Hollywood studio archives are among the factors that set Hollywood cinema archives apart from many public archives. Hollywood cinema archives are maintained by and for the content owners themselves, not by stewards holding community assets "in trust." Another distinguishing feature of Hollywood cinema archives is the sheer size of the body of assets to be preserved, including the number of new productions that must be added to the archive every year to keep the collection complete. Just counting MPAA-rated films, the total number of films released in 2006 was 607, an 11% increase over 2005's 549 films [Motion Picture Association of America 3].

While it has not always been the case, current Hollywood studio archiving policy is to "save everything," starting with the various versions of the finished movie, but also including all the original camera negative (OCN) film, all the original audio recordings, all the still photographs taken on-set, all the notated scripts and more. Everything is saved from the biggest hit movies and everything is saved from the worst commercial flops.

The modern motion picture business does a very comprehensive and reliable job of archiving feature-length motion pictures using film archives. But looking back over the past 100 years, Hollywood's history of archiving has been uneven. Many of the earliest movies were lost because long-term preservation of motion pictures was not considered important – either commercially or culturally. Many titles in early film libraries on flammable nitrate stock were destroyed by fire or merely thrown in the trash; other generations saw their film masters turn to "vinegar" in hot, humid warehouses until current climate control requirements for long-term film preservation were well understood. As a result, fewer than half of the feature films made before 1950 have survived, and less than 20% survive from the 1920s [US, LC, NFPB, Natl. Film Preservation Plan].

With the arrival of black-and-white TV in the 1950s, movie studios happily discovered they could generate profitable new revenues by converting old movies to video for broadcasters to beam to consumers in their homes. The introduction of color TV in the 1960s made the color movies in the archives even more attractive for broadcasters, and more profitable for studios.

The film archive business model became even more profitable with the widespread adoption of home entertainment packaged media, first as VHS tapes in the 1980s and then as DVDs in the 1990s. To differentiate DVDs from VHS tapes and to justify keeping a high consumer pricing model (despite lower per-unit manufacturing costs), the studios learned to pack DVDs with "extra value" by adding scenes, bloopers, outtakes, etc. Since it is nearly impossible to tell in advance which shots will be selected for inclusion in a future DVD, it has become industry practice to save all processed OCN from the original production, in addition to preserving various combinations of film separations and the interpositive (IP) of the final movie, at different physical locations in order to protect assets by geographic separation.

2.2 Current Hollywood Film Archiving

The terms "preservation masters" and "archival masters" describe the 35mm original camera negative (OCN), interpositive (IP) and yellow-cyan-magenta (YCM) separations on black-and-white film stock stored in environmentally secure film vaults.

SINCE THE BEGINNING OF THE "HOME video era" around 1980, most studios have come to recognize the potential long-term value of their film libraries and some have embarked on ambitious "asset protection" programs. Paramount Pictures is a case in point. From 1987 to 1993, Paramount reportedly spent over $35 million inspecting its negatives, audio tracks and black-and-white separations; doing film repairs; and printing new preservation materials. In 1990 it opened a new $11 million archives building, with environmentally controlled vaults for preprint and color materials. Paramount stores some of the master elements in an underground facility in Pennsylvania and tracks millions of items worldwide through a custom-designed computerized inventory and tracking system. By investing in the physical care of its collections, the studio expects to extend the shelf life and revenue potential of film elements as well as to expedite retrieval. A similar archive construction and asset-protection project was undertaken by Warner Bros.

Industry storage practices vary, of course. Other studios have their own film vaults on their premises, and store other film material at commercial vaults. In addition, most large studios routinely keep preservation masters of films they produce as well as additional materials – such as foreign-language soundtracks or edited television and airline versions – as required for ancillary markets. For each title, a studio may keep many different preprint and sound elements. The depth of preservation protection depends on the scope and duration of the studio's commercial rights and the film's expected value over time.

For most people in Hollywood today, the terms "preservation masters" and "archival masters" describe the 35mm OCN, IP used in the print manufacturing process, and YCM separations on black-and-white film stock stored in environmentally secure film vaults. The OCN is the most fragile and is only accessed to make new IPs or restoration elements when needed. The IP, usually extensively duplicated, is typically used to make new printing masters, and the separations are used when all else fails.

Each studio has its own list of what specifically should be archived. But generally speaking, the issues of analog archiving are well understood by many people. After more than one hundred years of technical innovation and market forces, the photochemical media formats have settled down to just a few remaining choices, manufactured by only a few remaining vendors in widely accepted standard formats. Today, there is broad consensus about how to manufacture and preserve 35mm film archival masters.

One of the largest film vaults in Hollywood, as an example, holds about 425,000 film elements of various kinds. This particular vault holds film elements from motion pictures produced in 1912. Like all the Hollywood film vaults built (or renovated) in the last fifteen years, this facility is designed for preservation storage with cold temperatures, low humidity and fire suppression systems.

Archive services are basically the same for whatever film element is placed in the vault, whether it is a television program, theatrical release, or documentary, no matter if it is cut or uncut. Every piece of film coming into the vault is inspected before storage. Inspectors manually inspect the film, look at the information on the leader and compare it to the container labeling. In addition to inspecting the film element for physical and photographic integrity, a staff restoration management director may have a laboratory make a viewing print to ensure that there is nothing wrong with the master negative, and that it is intact from first frame to last before committing it to archival preservation.

As part of the intake process for every film element, the archive staff manually logs basic asset management information such as element title, reel information, element type (OCN, IP, etc.), version description (editor's cut, etc.), program type (theatrical, TV, cartoon, documentary) and aspect ratio. Also, one or more unique barcode identifiers are assigned for inventory management. For a new release today, by the time the original lab materials reach the vault, a High Definition (HD) master has already been made, and this is used to supply various downstream electronic distribution platforms.

ARCHIVING

2.3 Hollywood Archives vs. Libraries

Digital archives are only truly protected by redundant replicas of the structured digital assets themselves.

FILM ARCHIVES IN DISTANT, COLD-TEMPERature underground vaults today are accessed only when necessary – for example, if no other good-quality film print master elements can be found locally. Sometimes this means an entire movie must be retrieved; sometimes just some short elements are needed to repair or replace a particular scene. They function as a kind of insurance policy to protect valuable assets produced at great expense. These are carried on the financial books of global media companies who have, over the years, bought and sold their film collections for millions or even billions of dollars.

In parallel to film archives intended for long-term preservation, studios operate short-term film distribution libraries containing release prints, interpositive and/or internegative film copies that can be used to manufacture new release prints and other finished elements (including sound tracks) needed to meet commercial distribution requirements. The assets stored in the distribution libraries are accessed frequently and are very actively managed to satisfy customer demands and maximize revenue potential during the primary commercial window for each title produced, typically a period of three to five years.

While the major studios' archives have stayed almost 100% film-based so far, the commercial distribution libraries operated by these same studios have expanded in recent years to include not only film prints, but also Digital Cinema Packages (DCP) as well as derivative versions of programs in digital formats for non-theatrical release, such as Standard Definition and High Definition video for sale to television broadcasters and cable and satellite system operators. Several people interviewed for this report believe that as higher percentages of a studio's revenue potential for a given title come through non-theatrical digital markets in new formats, there will be growing pressure to move the distribution libraries to digital platforms to stay competitive. At least two major studios, Sony Pictures and Warner Bros., have digital distribution library projects underway: ATLAS (through Ascent Media Group) and DETE (Digital End-To-End), respectively.

The traditional analog system that separates archives for asset preservation from libraries for distribution is being carried over to the digital domain. The digital media archival assets are most likely full pixel count, full bit-depth, uncompressed and unencrypted, as compared to the digital media distribution library content, which is most likely formatted at lower pixel count, lower bit-depth, and compressed.

Titles in the distribution library might be pre-encrypted, ready to go on demand. Or they might be stored un-encrypted inside the library, but always encrypted as part of the distribution process. Titles in the library might also contain embedded watermarking and other DRM metadata.

Digital archives are only truly protected by redundant replicas of the structured digital assets themselves. New titles move into the distribution library faster than they are added to the archive because the distribution library is used to generate revenue while the archive is intended to act as insurance against any loss of corporate assets. But if digital motion pictures can become "born archival," they can be ingested into the archive quickly and easily as part of a largely automated file-transfer process.

The storage and administrative systems for the digital preservation archives and for the digital distribution library might well merge into one unified repository, perhaps employing different user interfaces – one for library services, the other for archive services. Archival assets would typically require reformatting when they are retrieved as a library service, but not when they are accessed through the archival interface.

The conversion of archival formats into distribution formats has historically required slow and/or expensive processing, often using purpose-built hardware for speed. But continuing increases in digital processing power, digital storage capacity and digital networking bandwidth mean that it may become more efficient to co-locate the archive, library and distribution infrastructure. This could reduce the number of data transfers from archiving to processing facilities and consolidate many redundant functions shared by libraries and archives. It also would bring together the people responsible for preservation and those working on distribution and media processing R&D.

3 The Transition to Digital

The recent introduction of digital technologies into the final links in the production and distribution chain is, in fact, a "tipping point" that fundamentally changes the industry's economics and practices.

IT IS IMPORTANT TO UNDERSTAND THAT THE MOTION PICTURE industry has been adopting digital technologies in a piecemeal fashion over the last 25 years. The following sections present a brief history of the digital conversion. The recent introduction of digital technologies into the final links in the production and distribution chain is, in fact, a "tipping point" that fundamentally changes the industry's economics and practices.

The digital transition affected different aspects of the production process at different times, although the fully digital production still results in a "film-out," or creation of a film negative from the final digital master. This, along with the YCM black-and-white separations made from the final digital master, is the only finished film asset that is currently being saved using a well-recognized technology with understood and accepted long-term preservation and access characteristics.

3.1 Audio Converts First

MODERN AUDIO RECORDING, POSTPRODUCTION AND DISTRIBUTION all use fully digital workflows yielding digital audio files best saved on digital storage media. In fact, analog audio tape is rapidly disappearing. There are few remaining manufacturers of analog audio tape or the professional recorder/player devices that can handle such tape. This is compelling the sound departments of the major Hollywood studios and elsewhere to transition to digital archiving for lack of a better alternative.

Digital Audio in Acquisition and Postproduction

The introduction of digital audio recorders and processing equipment in the early 1980s was the start of the motion picture industry's conversion from analog electronics and film technology. The Nagra series of analog audio tape recorders manufactured by the Swiss company Kudelski, S.A., long the de facto standard for motion picture production sound recording, began to be replaced by the Digital Audio Tape (DAT) format, which was subsequently replaced by field recorders that use hard drives and recordable optical storage devices. By the late 1980s, the supporting analog mixing consoles and tape recorders used downstream for sound editing, effects, and mixing began to be replaced by Digital Audio Workstations (DAW), although the final soundtrack continues to be output in analog form to film stock coated with a magnetic layer ("fullcoat mag"), and then ultimately as an analog optical track on film prints.

Digital Audio in Exhibition

Although it was announced in late 1990, it wasn't until 1992 that Dolby Laboratories first introduced the SR/D format, known today as Dolby Digital, with the release of BATMAN RETURNS. The development that made this format possible was the AC3 audio data compression algorithm for 5.1 audio channels, with the ".1" signifying a limited-frequency bandwidth subwoofer channel. Real estate on film is precious, and so Dolby opted to record the "bit map," or images representing the actual digital bits, between the sprocket holes. It should be noted that the optical analog sound track was retained as a backup measure, which still remains on film prints today.[4]

More digital formats then appeared in the cinema marketplace. Digital Theater Systems (DTS) introduced the DTS digital 5.1 format in 1993 with the release of

[4] Optical Radiation Corporation was the first producer of a commercial theatrical digital audio reproduction system, used first on DICK TRACY in 1990, but the lack of an optical backup track, coupled with the system's complexity, prevented the system from being adopted by the major studios.

3.1 Audio Converts First *continued*

Digital audio tape degradation manifests in the inability to recover any of the sound at all.

Jurassic Park. DTS places the digital audio bits on CD-ROMs in a proprietary format and records only an analog synchronization track on the film, also preserving the analog optical track as a backup.

In 1993 Sony introduced the SDDS digital audio format with the dual releases of In The Line of Fire and Last Action Hero. Unlike Dolby Digital and DTS, SDDS is a 7.1 format, resurrecting the additional full-range effects channels of the Todd-AO 70mm magnetic format, although not all feature films are released using this capability. As with Dolby Digital, the SDDS data is recorded directly on the film, and as with both of the other digital formats, SDDS relies on the stereo optical track for backup [Karagosian].

It is important to note that each of the existing digital formats occupies an exclusive physical area of the film. In practice, it is more and more common to release a film print with the printed digital audio data or time code for more than one format. Film producers enjoy the choice and innovation that come along with multiple competitors in the marketplace. There are limitations and advantages to each of the formats, in terms of sonic capabilities, distribution capability, and the economics of the film print itself. For the foreseeable future, there will continue to be a variety of formats for multi-channel sound for cinema.

Archiving Digital Audio

Studio sound and preservation departments have long known that digital audio tape formats do not have adequate long-term survival characteristics, primarily due to their "brick wall" failure mode. That is, while analog audio tape degradation manifests as increased audio "noise" which can generally be filtered out, digital audio tape degradation manifests in the inability to recover any of the sound at all. It is for this reason that some studios have backed up their digital audio data to recordable CDs with scheduled migration to recordable DVDs. However, according to the National Institute of Standards and Technology, DVD technology has degradation characteristics such that approximately half of a collection of disks can be expected to last more than 15 years, and therefore half will not [The X Lab].

Digital audio preservation methods are getting more sophisticated. In a presentation at the Association of Moving Image Archivists' May 2007 Digital Asset Symposium, NBC/Universal Studios discussed the development of its digital delivery and preservation system that uses a combination of online hard drives, LTO3 data tape, and DVD-R optical disks to access and preserve its motion picture sound elements [Taylor and Regal].

Digital Asset Management systems require ongoing investment in infrastructure, hardware, software, and highly trained personnel.

JURASSIC PARK WAS NOT ONLY A WATERSHED EVENT FOR MOTION picture sound; it is also widely regarded as the first major motion picture to use photorealistic digitally created characters in a central role.[5] The movie's dinosaurs were originally planned to be shot with traditional stop-frame animation techniques using miniature models, but the initial tests of the digital dinosaurs were so promising that a commitment was made to go completely digital. The final product was impressive, and the popular folklore is that audiences "could not tell the digital dinosaurs from the real ones."

1995's TOY STORY was the first feature film with completely computer-rendered 3D characters, and in the years since, 2D animation and visual effects have been almost completely created using digital tools.

The adoption of purely digital tools for visual effects and animation created a need for effective digital data management tools for production activities, also known as Digital Asset Management systems (DAMs). DAMs, in most cases, effectively enable the backup and production-related access of the digital character models. This is not without its costs, as they require significant investment in Information Technology (IT) infrastructure, ongoing hardware and software upgrades, and highly trained personnel. But the combination of digital visual effects and DAM has proven effective in making some of the most commercially successful movies of the past few years.

3.3 Postproduction

Picture Editing

Interchange of images between facilities, a requirement in today's world of multi-facility collaboration, is problematic, given the lack of digital file format standards.

The transition from cut-and-splice film editing to electronic nonlinear editing began in the mid-1980s with the introduction of computerized videotape- and videodisc-based editing systems. Film-originated television programs were the first to adopt these systems because they did not require the conforming of the film negative to produce the final edited master. Television program masters were assembled from master videotapes using electronic "instructions" generated by the nonlinear editing systems. The development of "negative cut lists," coupled with the instantaneous access of digital video stored on computer hard disks in the early 1990s, made electronic nonlinear editing practical for the editing of feature-length motion pictures.

Today, almost every theatrical motion picture is edited on a digital nonlinear editing system, and consumer versions of this professional tool have found their way into tens of millions of homes. For better or worse, the rise of personal video-sharing websites such as YouTube would not have happened without the development of professional digital video editing tools.

It should be noted that in the three cases discussed to this point, the transition to complete adoption of each of the digital technologies took no more than ten years from initial commercial introduction.

Mastering

The final step in motion picture production is in the midst of its conversion from film to digital technology. Generally referred to as the Digital Intermediate process (although Digital Mastering is a more appropriate term), the final color balancing and visual styling of the final master film record more often than not are done using digital tools such as interactive color correction systems rather than adjusting film stock exposure and developing controls. The Kodak Cineon system, introduced in 1992, demonstrated that analog film images could be converted to digital bits, processed and enhanced, and then re-recorded back to film with powerful results. This concept is used for both visual

[5] *Earlier motion pictures such as 1985's YOUNG SHERLOCK HOLMES and 1989's THE ABYSS integrated computer-generated characters, but in relatively small supporting roles.*

effects integration and final color balancing, and it is widely believed that more than half of all major motion pictures today are mastered using the Digital Intermediate process.

As with any newly adopted process, there are unresolved issues. For example, some mastering facilities use High Definition Television (HDTV) equipment rather than higher-quality "4K" computer-based systems[6] as a cost-saving measure. The resulting master contains less visual information in terms of fine detail and dynamic range (collectively referred to as "precision" in the diagram below), is of observably lower image quality than what has been achieved for over 100 years with film, and there is concern that the decision to archive reduced-quality masters will have adverse consequences in the future [Scherzer]. Interchange of images between facilities, a requirement in today's world of multi-facility collaboration, is problematic, given the lack of digital file format standards.[7] Furthermore, the final physical form of the digital master – data tape, optical disk, magnetic hard drive – is not defined by any standard or industry agreement, and therefore what goes into the archive is not defined.

Another unintended side effect of the Digital Mastering process is that the final digital master in many cases bears little resemblance to the original camera negative (or digital camera original data if a digital camera is used in production). The level of creative control enabled by digital mastering tools effectively shifts significant decisions regarding the motion picture's overall "look" downstream from on-set choices historically made by the cinematographer.

Visual Attributes of Image Formats

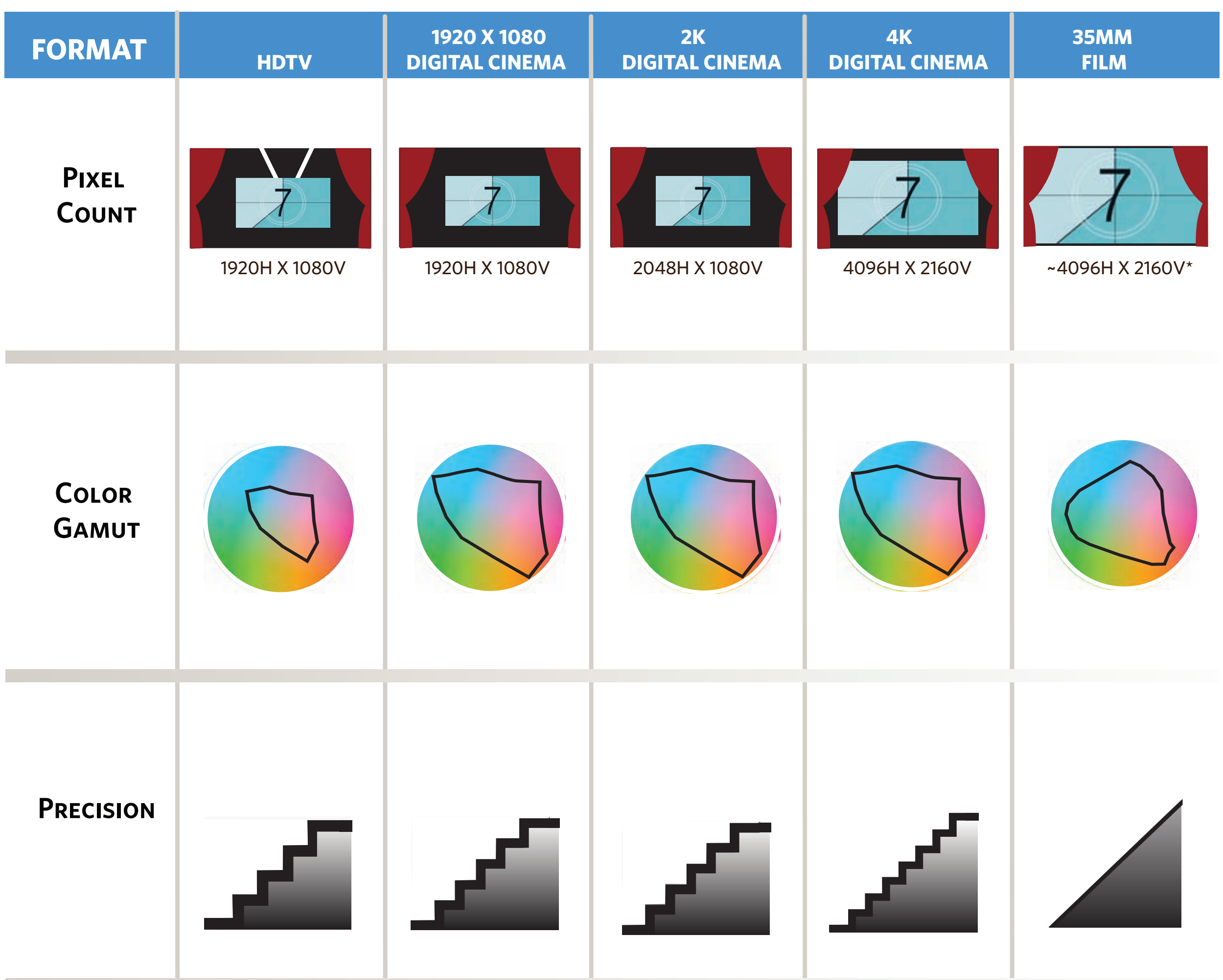


DIAGRAM IS NOT TO SCALE

** Approximate pixel count of 35mm film negative*

[6] *"4K" is shorthand for the highest pixel-count digital motion picture image format in regular use today. A 4K image has 4096 pixels in the horizontal direction and 2160 pixels in the vertical direction, which is roughly equivalent to 35mm film.*

[7] *The Academy has a project underway to address the interchange issue. More information on this project can be found at http://www.oscars.org/council/advanced.html.*

3.4 Exhibition

At what point film prints are no longer economically justifiable is unknown.

THE CONVERSION FROM FILM PROJECTION TO DIGITAL CINEMA projection technology[8] is underway, and so much has been written and is still being written about this topic that the subject will not be covered here except to point out that it is unclear at what time in the future film prints will become obsolete. Of the approximately 37,000 commercial theaters in the United States, 3,595 are Digital Cinema-enabled, and conversions are occurring at the rate of approximately 200 screens per month [DCinema Today; Overfelt]. The conversion rate is expected to accelerate when Digital Cinema Implementation Partners, a consortium representing over 14,000 U.S. screens, and Technicolor Digital Cinema begin their deployments. Assuming a doubling of the current conversion rate to 400 screens per month beginning in 2008, there would still be over 8,000 film-only screens remaining in the U.S. in 2013. The international conversion rate is expected to be slower than the domestic rate, given the unique business and governmental issues. With over 70,000 screens outside the U.S., there are likely to remain a substantial number of film-only screens for some time. At what point film prints are no longer economically justifiable is unknown.

3.5 Acquisition

There is, of course, the matter of how (or whether) to preserve the enormous amount of digital data produced when shooting digitally.

DIGITAL MOTION PICTURE CAMERAS THAT IN SOME RESPECTS MEET or exceed the perceived image quality of 35mm film negative are now in regular commercial use. The digital output of these cameras is recorded either to HDCAM SR digital videotape, a magnetic hard drive – based digital recorder or solid-state "flash" memory devices. According to the motion picture camera manufacturers interviewed for this report, approximately 20 to 30 major motion pictures per year are now shot using these cameras. Reported advantages of these cameras over film include immediate playback of recorded scenes in certain circumstances, increased color saturation in low light-level situations, and longer recording duration between media reloading. Reported disadvantages include reduced spatial resolution and exposure latitude relative to 35mm film and postproduction workflow challenges caused by the large amounts of digital data produced. Additional trade-offs must also be considered when choosing the capture medium for these cameras: HDCAM SR digital videotape or digital data recorders. For example, HDCAM SR uses mild image compression and digital data recorders do not; digital data recorders allow for deferring certain image processing choices until later in the postproduction process; and higher spatial resolutions and greater bit-depths are possible with digital data recorders.

This new capture technology has had some interesting effects on production practices. For example, the relatively low cost of digital videotape as compared to film negative has resulted in letting the camera roll for much longer periods of time than with film, enabling directors and actors to spend more time achieving a desired performance [Kirsner]. There is some concern that the greater amounts of source material generated with this production style will result in added overall costs when postproduction and archiving costs are factored in. It is also reported that some directors will do "circle take" selection on-set, deleting the digital files containing takes they know they will not use [Hurwitz]. The concern voiced about this approach is that it increases the risk of accidentally erasing a good take or eliminating potentially useful alternate takes.

There is, of course, the matter of how (or whether) to preserve the enormous amount of digital data produced when shooting digitally. This subject will be delved

[8] *"Digital Cinema" is defined as the theatrical projection standards currently being developed by the Society of Motion Picture and Television Engineers' DC28 Technology Committee.*

into in Section 6, though it is worth mentioning here that one of the motion pictures analyzed for this report produced well over 5,000 HDCAM SR videotapes from location shooting. Digital acquisition using uncompressed digital recording systems such as a magnetic hard drive recorder or a solid-state "flash" memory recorder generates 60 to more than 2,000 terabytes of data (depending on pixel count, bit-depth, number of backup copies, etc.) or the equivalent of 13,000 to 436,000 DVDs. By any measure, this is a large number of tapes or disks to consider when archiving original source material.

What is unknown at this time is whether digital cameras will ultimately supplant 35mm film as the primary capture medium for theatrical motion pictures. Cinematographers using the new digital cameras seem to agree that they are simply another tool in their creative toolbox, and that shooting with film still has its benefits. As with print film, it is unknown at this time whether the economic viability of film negative will be diminished as a result of the adoption of the new digital tools.

3.6 The Impact of Digital Technology on Motion Picture Archiving

Today, no media, hardware or software exists that can reasonably assure long-term accessibility to digital assets.

THE ADVENT OF DIGITAL CINEMATOGRAPHY, WIDESPREAD ADOPTION of Digital Mastering postproduction workflows, and the studios' push to deploy digital cinema distribution to theaters means that the cinema industry must reconsider its exclusive dependence on "film in a cold room" for long-term preservation of its motion picture assets. Studio representatives readily acknowledged in interviews that they see an emerging need to archive their digital assets, which are growing in number and variety and potential value. Traditional film archiving can no longer preserve all the forms of outputs flowing from the creative processes at the heart of the studio's business.

Generically, digital archiving is the systematic digital ingestion, storage, preservation and access, with the intention of long-term preservation, of digital "objects" comprising structured data files in a format that can be indexed and searched in some manner. When it comes to cinema, digital objects commonly include sequences of digital image frames that make up digital masters, multiple digital sound tracks, foreign-language dialog tracks, and text files containing subtitles in various languages. They may also include digital camera originals, digital audio original stem files, pre-mix/pre-dub audio files, and other digital "assets."

According to a 2005 paper from Stanford University researchers [Rosenthal, et al. 1], the goal of a digital preservation system is that the information it contains remain accessible to users over a long period of time. The key problem in the design of such systems is that the period during which such assets need to be accessible is very long – much longer than the lifetime of individual storage media, hardware and software components, and the formats in which the information is encoded. If the period were shorter, it would be simple to satisfy the requirement by storing the information on suitably long-lived media embedded in a system of similarly long-lived hardware and software. But today, no media, hardware or software exists that can reasonably assure long-term accessibility to digital assets. A dynamic approach that anticipates failures and obsolescence will be essential.

Archiving of digital assets is a new challenge for the studios. At one studio, there is a huge backlog of films waiting to be ingested into a digital archive. All digital elements are going into "temporary digital storage" where they will not be looked at again until they need to be migrated some years from now.

The discussions with studio personnel were wide-ranging and in depth on the challenges they face in this changing environment. In general, there is no clear picture of how to deal with not only the production and intermediate digital elements, but also the proliferation of different finished versions of a movie (e.g., foreign language versions, edited for cultural sensitivities in other markets, etc.). There is also much concern about the trend to create digital masters at 2K (only slightly better image quality than HDTV), which contains significantly less visual information than the film masters created today, or even those created 40 years ago. The fear is that projection and display technology will continue to improve, but the archived source material will produce nothing better than what can be seen on today's display technologies.

On the topic of storage media, LTO3 data tape is being used at one studio as an archival medium because they believe there is no better choice, and they recognize that this commits them to migrating the data to a newer format sooner or later. They also recognize that there has been no planning for that eventuality. The biggest challenge is that they are worried about making the wrong choice, given that their long-term objective is one hundred years of content life with guaranteed access. They also believe YCM separations are the best protection and insurance available today, because they provide a safety net to allow the use of a digital storage format with a much shorter lifespan such as LTO3, and to find a better solution within 7 to 10 years, assuming LTO3 lasts that long. They emphasized that digital archiving of the finished program is the number one priority. Archiving so-called "floor content" (trims and outs) that are saved today as part of the film archiving system is a secondary concern for this studio.

At another studio, digital acquisition of principal photography for "A titles" is seen as a looming challenge. This studio already saves key components, including the digital output negative and separations, but they do not have a system to save the original digital camera data. Ultimately, they want a method for long-term digital storage that works as well as film does. They are confident that it is feasible to keep digital elements protected and accessible for 5 and possibly up to 10 years, but long-term digital archiving is an unsolved problem. They are hoping for help from the storage industry in terms of new archival-quality media and other non-film archival methodologies that can be applied to digital sound and to digitally captured motion pictures. Eventually, if they are no longer able to output to film, then of course everything will need to go to digital storage. There is concern about the economics of digital archiving, but the bigger fear is that the studio will not adequately invest in future-proof archiving and access, and will thereby risk the long-term survival of expensive corporate assets that also have important cultural value.

At a third studio the most immediate problem is again how to handle digital camera origination materials stored on hard drives. Some digital cameras are using digital videotape, but many are recording straight to hard drive or flash memory storage. Without physical capture media such as tape or optical disk, there is no easy way to keep the entire digital negative. Studio archivists do not know if they have archived everything because no physical media exist – there is no equivalent of OCN in these cases. They are also encountering this problem with audio recordings, and comment that it is unlikely that every digital videotape or digital audio tape in storage today can be transferred to data tape in the future because of cost. Management of metadata – the "data about the data" that allows for efficient indexing, search and retrieval – is critical for archivists but is not a high priority for manufacturers or users.

At still another studio, senior technologists worry that there is no formal corporate archiving strategy across the whole company. The technology group can lead (intellectually) and formulate "recommended practices" that the business units may adopt or not, as they wish. Strategy decisions for archiving are largely made at a business unit level. The business unit that is tasked with storing assets has to pay the storage costs. Archiving issues are currently handled in a decentralized manner, but people are starting to realize that if the different business units were to compare notes and start to share resources even a little bit, the enterprise as a whole could be more efficient about how to tackle the digital archiving challenges. Given the complexity of internal accounting, operational and business responsibilities, the best approach, they think, would be to centralize knowledge about digital archiving but decentralize control and budgeting for specific archiving facilities and the assets they hold. Stand-alone "silos" of digital archives remain, and basic problems related to internal archiving are unresolved. Therefore, compatibility with external archives is still a low priority.

One executive argues that digital archiving is strategic to the future global media business of the studio, and that this work needs to be done fairly close to home because the production processes and the archiving processes are getting interwoven. There is an assumption that everything produced in the future is going to get re-purposed, sliced and diced in many different ways for different markets over many years. "It's not like the film vault of old where you could ship stuff off to

3.6 The Impact of Digital Technology on Motion Picture Archiving *continued*

underground mines used for storage, and then call a few times a year and tell them to ship stuff back to you." The assets in the archive all need to be treated dynamically now.

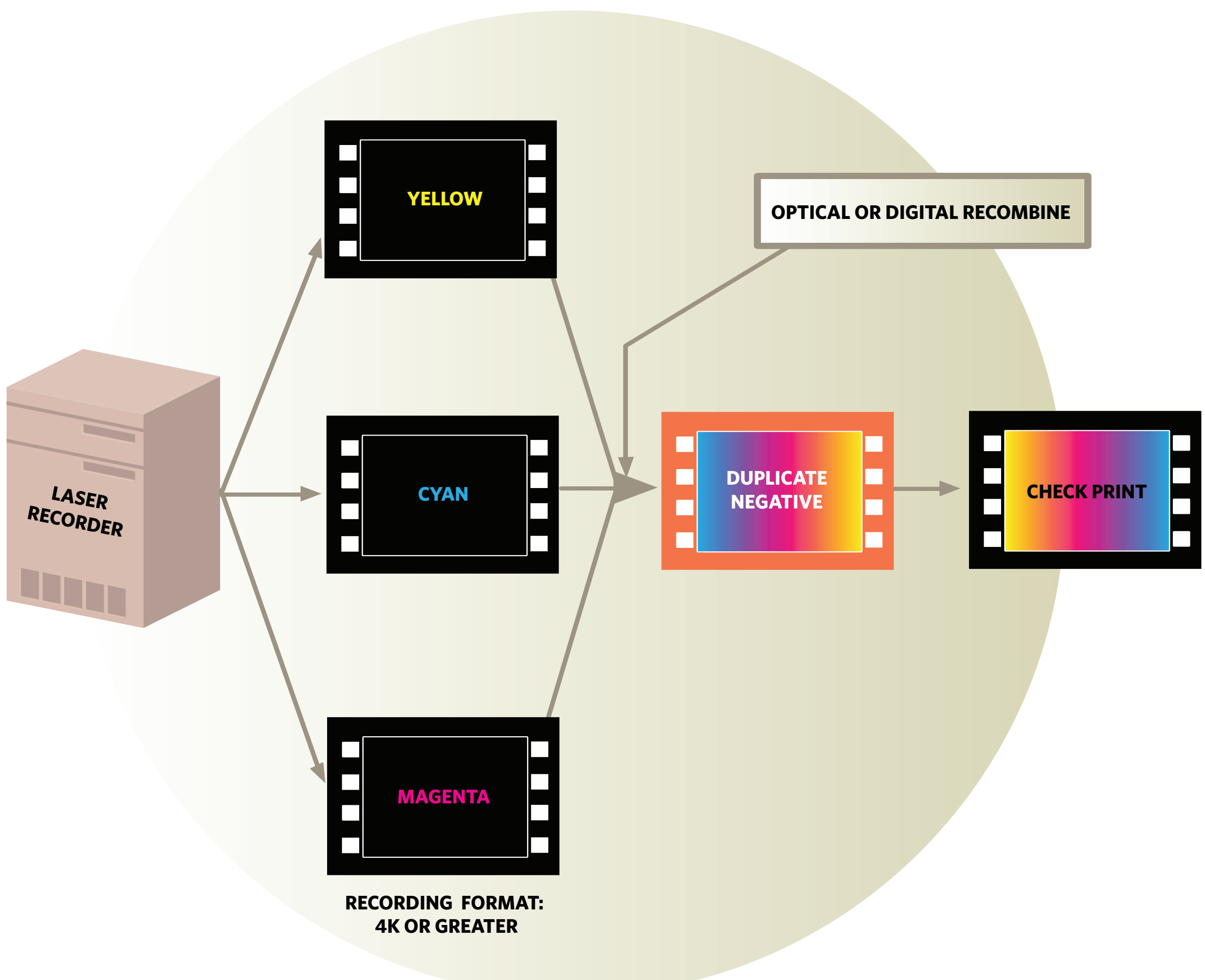
One senior technologist anticipates that the biggest challenges for digital archiving in a "studio culture" will be organizational, requiring long-term educational efforts, process re-engineering and self-discipline.

At another studio, a senior technology executive explained that, ideally, he would like to have all his archived assets available on online magnetic hard drives so that in 50 years the studio can use computing power to do things no one has thought of today. This would enable new video and audio search tools to automate the cataloging/metadata bottleneck. He believes that the studio ultimately wants instant accessibility to everything.

Another studio executive explained that in addition to wanting to archive new HDCAM SR tapes as original footage from digital cameras, his company wants to archive all of the scripts and the shooting logs. But everything on paper goes into the "all paper" archive storage facility, while anything to be saved from photo shoots is sent to a different library than the videotapes themselves. The hope is that everything will eventually be connected through the use of metadata and databases. But today there are still a lot of cardboard cartons just filled with un-inventoried "stuff."

On the other hand, a very senior studio executive explained that archival storage of major motion pictures is both a financial obligation and a cultural obligation. He feels a heavy responsibility to protect the studio's ongoing archival preservation activities against all risks, internal or external. In his view, this makes it imperative that any strategy for archival preservation be able to

Making a Digital YCM Separation Archival Master
On Black-and-White Polyester Film Stock



survive even the potential risks of global economic bad times or investor-driven corporate budget cuts from "on high." He emphasized the critical requirement that truly archival assets be able to survive even if someday there is no money for the next data migration. For him, "archival" means "store and ignore," with the belief that survival for 20+ years without worry, and playability 50 to 100 years from now without major additional investment and in the face of "benign neglect" are essential ingredients of any sustainable archiving strategy. The problem, he believes, is that no modern corporate institution will or can commit to "forever" funding. His studio today also holds data tapes, but does not consider them archival. They would like to be able to archive data – treat it as an archival asset – if and when there are good solutions to the problems of data migration and similar digital preservation strategies. The only thing that meets this studio's definition of "archival" now is 35mm film, so "archival preservation" at this studio depends on film prints and positive/negative elements, plus YCM black-and-white separations on film that go into deep (underground) storage. His company's film IP typically goes bad every 6 to 7 years due to repeated use. If the OCN is in good shape when the IP goes bad, the IP can be remade from the OCN. If not, it can be reconstructed from the black-and-white separations. If digital files need to be reconstructed in the future, the black-and-white separations can be re-scanned using the higher-quality, faster digital scanners of the future. He would like to find a digital alternative to film archiving, but has not seen it yet. And until then, the only thing the studio can truly depend upon is film for archival preservation.

Hollywood will probably continue to archive new motion pictures on film as long as film stock and film processing remain available and economical. The simplicity and dependability of film's "direct view" access compared to the software-based "interpreted view" of digital content continues to be attractive to many in Hollywood. The economics of film archiving are well understood compared to those of digital archiving, and film archiving requires little new investment. Furthermore, old motion pictures already in the film archives are expected to survive intact for the next 50 to 100 years, assuming the temperature and humidity in the film vaults are maintained under proper film preservation conditions. If for no other reasons, institutional inertia and the natural conservatism of studio management will tend to extend the use of film for archiving of motion pictures. Interestingly, the cataloging and indexing systems for film archives, especially the crucial metadata databases needed to implement any enterprise-wide DAM system, have already gone fully digital in many cases, although there is no commonality of implementation among the studios.

Cinema motion picture archiving must encompass digital archiving; "born digital" assets have no film elements to preserve

Like all the other media industries that have adopted digital technologies before it, the cinema industry is starting to generate an increasing percentage of important media assets that have no analog version – that is, they are not created on film in the first place. These assets are "born digital." The growing use of digital cameras for principal motion picture photography on "A title" studio movies means that instead of original camera negative at the end of a day's shooting, there are boxes of HDCAM SR videotapes or terabytes of camera-generated data files on disk-packs and data tapes. The move away from shooting film is also associated with a reported trend to higher shooting ratios, yielding more boxes of videotape or more terabytes of data, depending on the production. The output of the Digital Mastering process is no longer a cut negative in many cases, but rather terabytes of uncompressed digital frames on magnetic data tape. And with the spreading deployment of Digital Cinema to theaters in the coming years, the use of release prints overall is likely to decline in favor of Digital Cinema Packages (DCP) for digital distribution to theatres via hard disk, fiber or satellite.

Based on interviews, it appears that within the major studios there is no clear strategy yet for dealing with these new digital assets. Born-digital assets are being generated in growing amounts. Without a clear plan or direction, managers at production companies, post houses and the studios themselves generally seem to be taking the safest short-term approach, which is to continue the conventional practice of saving everything for possible future use, and keeping the assets in their original format – that is, putting digital camera originals on HDCAM SR tapes, magnetic hard drives and LTO data tapes on a shelf in a cool, dry place until further notice. Some studios are

3.6 The Impact of Digital Technology on Motion Picture Archiving *continued*

recording "digital master" files of the completed motion picture to LTO data tape cartridges and putting them on the shelf next to the HDCAM SR videotape cassettes. It is a reasonable and prudent interim tactic, but not a long-term strategy.

New types of content are not suited to film preservation

Even studio executives who firmly believe in the wisdom of film-based "store and ignore" archives realize they must eventually embrace digital archiving and reduce their exclusive dependency on "film in the freezer" for long-term preservation of corporate media assets. Their own marketing and sales teams, tracking new demand trends and innovating to generate new revenue opportunities for their business units, are driving changes in the formats and mix of the commercial media products manufactured by the studios. Full-length motion pictures for theatrical release will continue, of course, but the non-theatrical release versions sold by studios account for much larger percentages of their global media business [Galloway]. Some of the non-theatrical release products will still be derivative of the cinema original, but many will not. This will affect the choice of elements that studios need to put in their libraries and archives and how they will be used in the future.

One studio executive explained that he expects change in archival rationales will largely be driven by the changing form factor of content. Eighty percent of this studio's business now is about feature-length cinema, 90 to 135 minutes per title, and television-show length: 22 minutes and 44 minutes per episode. Naturally, these are the primary "content" being archived today by studios. But a growing volume of material created at the desktop level is neither for feature movies nor television programs. There is growing demand for short videos and animation. New digital formats for Internet distribution, elements produced originally for the World Wide Web, and content targeting small portable media players are not yet being consistently archived. As customers get more of their entertainment from more types of digital channels, the media elements that are created by the studio will be smaller, more varied, and more numerous. A decreasing percentage will be film-based or even suitable for film recording. An increasing percentage of a studio's productive output of commercial media assets will be "born digital" and cannot be preserved through traditional film archiving practices. This is presenting the studios with new archiving questions.

The theatrical release of a new movie is increasingly accompanied by the simultaneous release of companion video games by the games division of the studio in order to get greater customer awareness and higher sales within the demographics targeted by both theatrical marketing and games marketing. This raises new questions about how the studio should archive the digital assets created for the game such as the digital characters, computer models, scenery and software programs that determine the game's interactivity and "play value." Preservation on film is not even an option in this case.

Long-term viability of film as a preservation medium is also at risk because of overall film market trends

Today's large-scale "day-and-date" release patterns have actually increased the use of intermediate and print film stocks. However, the accelerating conversion of cinema to digital distribution following the recommendations of Hollywood's Digital Cinema Initiatives (DCI) consortium will erode the market for film release prints and intermediate film. In parallel, the emergence of Hollywood-grade digital cinema cameras will likely cut into sales of camera negative film. All these market trends will put downward pressure on sales volumes for film manufacturers, film laboratories and suppliers of film-processing chemicals. As demand shrinks for any consumable technology, manufacturing loses economy of scale, product availability decreases, prices increase and quality control suffers, further depressing customer demand.

The manufacturing base for high-quality 35mm entertainment film is already shrinking and has consolidated to just three remaining vendors, Kodak, Fujifilm and Agfa, although Agfa produces only print film and specialty black-and-white stock for sound applications. All are big companies with proud histories as technical innovators and market leaders. But none are likely to make significant new investments in film R&D or manufacturing improvements, given the sinking demand for their film products. While Kodak, Fujifilm and Agfa continue to supply high-quality, reliable film stocks and film chemicals, their managements will not commit their firms to the entertainment film business "always and forever." Nor would their shareholders welcome such a commitment to a "sunset" market.

The consumer film market is also collapsing due to the enormous popularity of digital cameras. According to IDC market research, worldwide manufacture of consumer film peaked at 80 to 90 million prints a year in the late 1990s, but had declined to about 40 million by 2005, and is continuing to drop 20 to 30% per year [Hogan]. This weakens yet another pillar of the film business that has historically offered valuable economy-of-scale advantages to the largest film manufacturers.

Several people interviewed for this report acknowledged that the demise of film is a long-term eventuality, but do not expect film to disappear in the next decade. However, studios planning their long-term archiving strategies must recognize the risk that analog film archiving of new titles may become more expensive and/or stop being a viable option altogether in the future. Is it prudent to build long-term preservation infrastructure based on a medium that, if not totally obsolete, may only survive in niche markets – like motion picture archiving? When YCM separations reach the end of their archival life, archivists entrusted with these valuable corporate assets must consider whether it will be better to migrate them to another generation of film stock, or to migrate them to a future digital format with its to-be-determined preservation methodology.

Some old film assets will be selectively added to digital archives

The wholesale migration of major film archives to digital storage is such a large and expensive undertaking that no studio appears to be considering this currently, at least for archival preservation in the strict sense. It seems likely that the studios will start to consider digitally scanning some older content to protect irreplaceable film elements when they become dangerously fragile or deteriorated. Some old film assets will also be digitized for commercial exploitation on a deal-by-deal basis, because once converted from analog film to digital files the content can be more easily manipulated and re-purposed to generate new revenues.

3.7 Television

Since shifting from 16mm film acquisition to videotape recording, the broadcast industry has repeatedly chosen to adopt tape formats to take advantage of technical advances that offer near-term operational, economic and/or quality improvements.

ALTHOUGH THIS REPORT IS AN INVESTIGAtion of digital archiving and access issues from the motion picture industry's perspective, there is no denying the intimate connection between production and consumption of theatrical motion pictures and television programming. In fact, every major Hollywood studio also has significant television activities, and it is worthwhile to look at what has happened and is happening in that area.

History of Television Archiving

According to a 1997 report by the Library of Congress [US, LC, NFPB, Television/Video], historically, few television programs held by the major studios and networks have been destroyed due to deliberate decisions or policies. In fact, the growth of non-broadcast distribution channels, consumer packaged media sales, and overseas markets for American TV programs, has encouraged systematic preservation. All the major studios, even in 1997, had implemented asset-preservation programs for their prime-time programming that included both film and videotape assets. The reason for preserving these programs was explicit: they represent real assets of value to their corporate owners.

Network news divisions, even in 1997, were having difficulty preserving all their programs because of the sheer volume they produced every day. They focused on preserving what they judged was valuable for the daily production needs of their reporters and editors, more than on keeping historically complete archives of all the news they broadcast.

The oldest television archives are on 16mm and 35mm film. 16mm film was phased out after Electronic News Gathering (ENG) cameras and compact videotape recorders were introduced in the 1980s. 35mm film is being replaced by HDTV acquisition even for premium programs. Many local stations simply discarded their 16mm film libraries when they converted to U-matic videotape, leading to gaps in the public archives of local news between 1950 and 1975. Even today, most local news content is not saved more than a few weeks before the videotape is recycled. But the largest, most progressive broadcasters have been migrating their archives from film to videotape, and from videotape to all-digital archives using general-purpose Information Technology (IT) infrastructure for the past 5 to 10 years, still a "work in progress" according to many in the field.

Since shifting from 16mm film acquisition to videotape recording (which is recognized to be a much less durable medium than film), the broadcast industry has repeatedly chosen to adopt tape formats to take advantage of technical advances that offer near-term operational, economic and/or quality improvements.

Videotape recorder vendors have engineered many improvements since AMPEX introduced the first commercial videotape recorder in 1956, the VRX-1000 with its proprietary 2'' Quadruplex tape format. Since then, there have been more than 60 different videotape formats. Scanning methods, signal encoding and image formats have evolved rapidly. Image and sound recording quality has gone up steadily while size and cost have come down just as steadily.

Over the years, competing vendors have fought "format wars" for market share, churning out new and better devices that users have sequentially adopted to their own advantage. This has left video archives full of many incompatible formats that run only on obsolete devices, requiring migration from old tape formats to new formats to access the value of the assets in the archives.

Modern digital videotape, such as HDCAM SR, are claimed to have a shelf life, under recommended environmental conditions, of up to 30 years according to manufacturers' commercial literature. The HDCAM SR format is still far too young to confirm this longevity empirically, and there is no assurance that functioning HDCAM SR tape drives will still be available 30 years in the future.

Today, professional videotape manufacture is limited to a few very large companies that have, in the past, been able to effectively leverage the consumer market for videotape to achieve economies of scale in manufacturing and R&D. But the consumer market for videotape has declined greatly in the past decade as Hollywood-packaged media shifted from VHS to DVD. When this consumer trend is combined with the accelerating conversion of broadcast infrastructure and workflow to tapeless, file-based operations [Kienzle, "Breaking," 1], it is likely that videotape as a recording medium will itself become obsolete in the not-too-distant future.

The scale of Hollywood's archiving requirements is not so different from that of institutions in other domains which also require long-term preservation of very large volumes of valuable pictures, sounds, text and other types of data in support of their missions.

ONE OF THE MOST OFTEN-ASKED QUESTIONS WHEN DISCUSSING Hollywood's digital archival issues is, "What are other industries doing?" Many industries have already adopted strategies for the preservation of their digital data collections, and while Hollywood's digital assets are large in number and size, they are not unique in these respects. Potentially, Hollywood does not need to invent digital archiving from scratch, so the risks of trying new approaches can be kept relatively low by studying others.

There are several areas of modern society that have large collections of media assets of various types with archival requirements similar to the needs of the cinema industry. While the digital motion picture assets that the Hollywood studios want to protect are exceptionally large on a per-title basis, the scale of Hollywood's archiving requirements is not so different from that of institutions in other domains which also require long-term preservation of very large volumes of valuable pictures, sounds, text and other types of data in support of their missions.

Advanced visualization technologies have always been supported by three "pillar" industries that drive the state of the art: entertainment (cinema and publishing), defense/intelligence, and science/medicine/education. Historically, they have all used analog imaging techniques developed for their particular needs, with little dialog or technical cross-fertilization between them. However, with the widespread adoption of high-quality digital imaging, all three pillars are moving away from film to common digital platforms applied to their different purposes. All three have a growing need for digital archiving of still and moving images. All are facing similar challenges in terms of infrastructure, workflows and requirements for long-term preservation. The defense/intelligence and science/medicine/education communities are already operating large digital image archives and can provide valuable reference for Hollywood studios initiating their own digital archiving programs. All of these large-scale, long-term digital archives have refined their system designs over the years to accommodate data ingest, search and retrieval, as well as relatively efficient and reliable data migration, file format updating, auditing, quality control and (when used) discard/transfer processes needed to insure the accessibility and integrity of their digital assets. They also anticipate long time horizons: 50 to 100 years, or even "permanently" in some cases; that is, for the life of their particular enterprise, assuming adequate funding.

Other media archives are already transitioning from analog to digital. Many large public libraries and archives with extensive collections comprising many media types are preserving their most recently acquired content digitally because more and more modern media are originally produced digitally, "born digital" assets that are delivered to the library on some kind of digital storage media or directly as data files via a computer network. Users have generally appreciated the faster, easier accessibility of the digital collections offered by the libraries. In response, librarians and archivists are digitizing their most important (most popular) analog assets to make them more accessible, too. Other assets are being converted from analog to digital when the analog media are deteriorating, putting survival of the content at risk in the absence of a reliable digital preservation strategy. Similar trends can be seen in several commercial media industries as well.

4.1 Corporate America

The most significant difference between corporate record archiving and Hollywood archiving is the intended duration of archival preservation.

4.1.1 Sarbanes-Oxley Act Requirements

IN TERMS OF THE TECHNICAL DIFFICULTIES involved in long-term preservation of digital assets, the most significant difference between corporate record archiving and Hollywood archiving is the intended duration of archival preservation. For example, the Sarbanes-Oxley Act of 2002 (SOX), passed as a result of the corporate accounting scandals at the turn of this century, only requires preservation of certain types of corporate data for seven years, a term marginally within the life cycle of a single generation of digital storage technology. SOX affects mostly transactional data, which means that the archive period starts at the time of transaction. The archived data is always rolling over as new data replaces old data that can be discarded after the seven-year mandatory archival period. There is a statutory requirement for "protecting the unalterability" of the archives, so implementation emphasizes the use of "Write Once, Read Many" (WORM) storage, audit trails, rigorous access control, data authentication techniques and legal compliance.

This contrasts to the Hollywood goal that digital motion picture archives be preserved for 50 to 100 years, comparable to existing film archives. That is a longer period of time than any digital technology available today can reasonably support without using specialized digital preservation strategies such as data migration, discussed later in this report. Nonetheless, properly designed corporate data archival systems implement the defensive data preservation strategies described in *Requirements for Digital Preservation Systems* and the National Research Council's *Recommendations for a Long-Term Strategy* [Rosenthal 3; Natl. Research 59-69].

A further contrasting parameter is the storage volume of corporate data that requires this level of preservation. Many of the studies on corporate IT practices reviewed for this report were sized in the gigabyte and terabyte range, which is substantially less data than is generated by a single digital motion picture production. This has significant impact on overall system costs, from both initial capital investment and operating perspectives. This topic is covered in detail in Section 6.

Finally, the tight integration of business systems and IT infrastructure, as well as the relatively common data storage requirements across corporate America, enables significant economies of scale and close collaborations with IT vendors that are not easily achievable in the motion picture industry, given the specialized nature of motion picture production.

4.1.2 Oil Exploration

IN INTERVIEWS WITH A LARGE OIL COMPANY'S data system manager and a data storage service company, it was learned that the oil and motion picture industries have something in common: the derived data is more valuable than the raw captured data. That is, captured geological data must be heavily processed before it has any immediate value – that value being the location and size of oil deposits to be extracted. The processing algorithms improve over time, and that is the incentive to preserve the original captured data: new oil deposits can be identified using "old" data.

The oil industry has another characteristic in common with the motion picture industry: a typical raw geological data set can be 200 terabytes (the size of about 25 uncompressed 4K digital motion picture masters), and a typical survey of the Gulf of Mexico can generate hundreds of data sets, and is normally stored on hundreds of magnetic data tapes.

According to the people interviewed, raw geological data began to be archived more than a decade ago, and they are experiencing problems that sound familiar: a heavy and undesirable reliance on vendor-specific solutions which limit future freedom of choice, a lack of standard file formats (which enforces single-vendor reliance), ad hoc archiving procedures, no experience with data migration, and a need to maintain working versions of old hardware and software to guarantee access to valuable data. Therefore, standardized archival practices and data formats are a long-term goal for this industry.

4.2 Government and Public Archives in the U.S.

The task of saving digital assets is too large for solitary efforts.

THE LIBRARY OF CONGRESS' NATIONAL DIGITAL INFORMATION Infrastructure and Preservation Program (NDIIPP) initiative and the National Archive and Records Administration's Electronic Records Archive (ERA) program have both put emphasis on collaborative problem solving, drawing on the opinions of experts from many fields and providing forums for valuable exchange of information that serves the purposes of these institutions and contributes to general understanding of the challenges and possible solutions to large-scale institutional preservation of digital assets. The Academy participates in both of these federal government initiatives, as a member of NARA's Advisory Committee on the Electronic Records Archive (ACERA) and a partner in the Library's Preserving Creative America program under NDIIPP.

4.2.1 Library of Congress

ACCORDING TO ONE VETERAN AT A LARGE STORAGE MEDIA manufacturer, in the late 1980s the U.S. Library of Congress said it wanted a 200-year archival life for its digital assets. The vendor's engineers went to work on accelerated aging tests to try to meet the Library's goals. However, after several years it became clear that no digital storage scheme available then (or now, or in the foreseeable future) can be sustained for 200 years. The Library realized that digital media are so ephemeral and digital technology changes so rapidly that long-term digital preservation was going to need a new approach.

In December 2000, Congress appropriated $100 million for the NDIIPP collaborative project in recognition of the importance of preserving digital content for future generations. Led by the Library of Congress, NDIIPP has generated digital archiving guidelines that are useful for any organization that is formulating its own strategy to collect, archive and preserve growing amounts of digital content for current and future generations, especially materials that are created only in digital formats. NDIIPP set five initial goals for the Library of Congress and, by extension, for any organization faced with digital archiving challenges [US, LC, Digital Preservation]:

1. ***Identify and collect at-risk "born digital" content that is created only in digital form, before it is lost, misplaced, goes obsolete or becomes corrupted.***
2. ***Build and support a network of partners working together to preserve digital content. The task of saving digital assets is too large for solitary efforts.***
3. ***Develop and use technical tools and services for digital archiving.***
4. ***Encourage development of strategic policies to support efficient and reliable preservation of digital information. Document the rules, and educate the staff; technology is only part of the problem.***
5. ***Show why digital preservation is important for everyone in the enterprise. Saving information, especially the right information, has to become everyone's task.***

The Library of Congress has an additional motivating factor for the development of digital preservation technologies and practices: its National Audio-Visual Conservation Center (NAVCC), located in Culpeper, Virginia, will house the entire collection of the Library's Motion Picture, Broadcast and Recorded Sound Division. The collection contains increasing amounts of digital materials, and the NAVCC's digital storage system is expected to ingest over 8 petabytes (equivalent to about 1,040 uncompressed 4K digital motion picture masters) per year when fully operational [US, LC, Natl. Audio 15]. Furthermore, as the repository for mandatory copyright deposits, the NAVCC system must consider the copyright term of 120 years or longer.

It is clear that the digital preservation concerns of the Library of Congress are quite similar to those of the Hollywood studios.

4.2 Government and Public Archives in the U.S. *continued*

4.2.2 National Archives and Records Administration

THE U.S. NATIONAL ARCHIVES AND RECORDS Administration (NARA) is responsible for preserving all official government records, both to protect the records as official history and to make them available for future reference. NARA operates both classified (secret) and unclassified (open) archives. By NARA's estimate, only 1 to 3% of the documents generated by the federal government are significant enough to be added to the archives.

NARA's current digital holdings are diverse. A few data files were originally created as early as World War II and reflect punch card technology in use since the 1800s. An even smaller number contain information from the 19th century that has been converted to an electronic format. However, most of the electronic records in NARA's holdings have been created since the 1960s.

As the 21st century began, NARA planners realized that going forward more and more official government business will use electronic records that NARA itself will have to accept, catalog, search, give access to and preserve "permanently" in a new kind of digital archive capable of handling thousands of formats and trillions of data objects. NARA recognized that these are very complex problems requiring long-term planning, and therefore established the Electronic Records Archive (ERA) program to meet their visionary goals: preserve any type of record, created using any type of application, on any computing platform, from any entity in the federal government or any donor; and provide discovery and delivery to anyone with an interest and legal right of access, now and for the life of the Republic.

NARA archiving responsibilities are mandated by statutory requirements to preserve all official government records, with rules that compel record creators to deliver assets to the Archivist of the United States within certain time limits. The ERA team at NARA realized they could not do or think of everything themselves, so ERA established a network of partnerships with computer scientists, engineers, information management specialists, archivists, industry experts and professionals. Through workshops, symposia and funded research projects, the ERA's strategy has been to attack the critical preservation problems as the first priority, defining the requirements in terms of the "lifecycle" management of records. They want to use commercially viable, mainstream technologies being developed to support e-commerce, e-government and the next-generation national information infrastructure, aligning NARA with the overall direction of IT in the U.S. government, and in the process perhaps leading the U.S. government's IT practices to align better with NARA's essential archiving mission. As such, NARA/ERA is in a position to dig deeply into the issues of long-term digital archiving and help build consensus in the field.

The ERA project is phased, with the initial goal of accepting the electronic records of President George W. Bush's administration when he leaves office in January 2009. NARA expects to process and ingest over 800 million email messages and attachments into the ERA system at that time, and is currently working with four other federal agencies to develop and test the base system. Although the system is still in development, NARA believes that the use of standardized file formats and metadata, as well as automated ingest and metadata harvesting, are critical to the system's long-term success.

4.2.3 Department of Defense

THE ACADEMY'S PARTICIPATION ON THE National Archives' ACERA committee provides a view into the digital data management activities of other government agencies, including the Department of Defense (DoD). One of the most relevant presentations from the motion picture industry's perspective was an overview of the DoD's Advanced Distributed Learning Project (ADL), which is a system designed to make all of the department's audiovisual training materials accessible and reusable across the entire department. The volume of training materials is large, as is the variety of media types, so the ADL system has had to address a number of challenging issues ranging from basic issues of digital storage to more complex topics such as metadata and digital object registries.

While one might consider the ADL system more of a digital library than a digital archive, there is a tremendous overlap between technologies and practices used in this system and those the motion picture industry will need to implement in the future.

Also interviewed were representatives from the Office of the Director of National Intelligence, who serves as head of the Intelligence Community. They expressed much interest, both on the part of this office and the DoD, in collaborating with large producers and consumers of digital media to develop standardized file formats, especially with respect to metadata. They strongly believe that it is not economically feasible for single organizations, even as large as the DoD and those that are part of the Intelligence Community, to develop these technologies on their own. As it is, for various historical, organizational and technical reasons, the DoD and intelligence communities are populated with many large digital-imagery archives that have emerged without comprehensive planning for long-term preservation and suffer from interoperability barriers between archives and agencies. Their stated recommendation was, "Don't let this happen in Hollywood."

4.3 Medical

Digital medical images are "born archival"... all of a patient's identifying data is collected *before* the diagnostic image is captured, and the metadata is forever associated with a captured image.

THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT of 1996 (HIPAA) requires hospitals, clinics, medical equipment rental companies, physicians' networks, dentists, drug stores, medical insurance companies, medical billing companies and nursing homes to preserve and protect the privacy of electronic medical records, including diagnostic medical images. HIPAA requirements are, in practice, designed more to protect privacy and promote operational efficiency than comprehensive archiving. The motivation for archiving medical data comes from its benefits for research and education. According to a report by the National Science Foundation's Science Board:

> *"It is exceedingly rare that fundamentally new approaches to research and education arise. Information technology has ushered in such a fundamental change. Digital data collections are at the heart of this change. They enable analysis at unprecedented levels of accuracy and sophistication and provide novel insights through innovative information integration. Through their very size and complexity, such digital collections provide new phenomena for study. At the same time, such collections are a powerful force for inclusion, removing barriers to participation at all ages and levels of education."*

And according to the Mayo Clinic in Rochester, Minnesota, "cine [filming] for radiography has served the medical industry's needs well since the early 1950s. For the first time, it allowed recording of motion studies of the cardiac structures on film. The cine technique has been standardized over the years, both the camera and the display. Cine filming techniques, however, have not advanced, except for new film products with faster emulsions and better-quality films. Video imaging for cardiology has made rapid advancements.…With the advent of interventional procedures in the cardiac catheterization laboratory, the need to assess images immediately cannot be fulfilled by cine filming because of the requirement for the processing of the film with its inherent delays" [Holmes, Wondrow, and Gray 1].

According to the Cleveland Clinic, one of the largest hospitals in the U.S. that maintains a substantial digital image archive supporting both its clinical and research activities, much of the medical field began its conversion from film imaging to digital imaging a few years after the Holmes paper was published in 1990. Currently, all of the Cleveland Clinic's medical imaging (including radiology) is done digitally, and its plan is to keep all digital images forever for historical trend analysis and research. The current film holdings are not digitized because it is not cost-effective, and the old films are stored in a cool and dry warehouse.

The Cleveland Clinic's archive currently stores 1 petabyte of digital data, composed of objects such as chest images that are about 20 megabytes per image, and motion clips that are 500 gigabytes per patient. Image data compression is not used because life-or-death medical decisions are made based on this data. The archive is growing at a rate of 3 terabytes per week, which will double its size in the next year. Image pixel counts have increased and more images per patient are being made, so this growth trend is expected to continue.

Their storage strategy is to store the most recent data on an array of magnetic hard drives, and there is currently 100 terabytes of such online storage. The biggest problem with this system is that the disk lifecycle is only three years – every disk drive must be replaced after that interval. Hardware is not the only cost; all of the

4.3 Medical *continued*

data must be copied when the hardware is replaced, and as the archive grows, the time required for copying is getting longer and longer. Once the data ages to a certain point, it is automatically transferred to a data tape library, with every tape automatically evaluated every 90 days. The reliability requirement is zero errors per 100,000 operations. A secondary archive is located 12 blocks away from the primary archive at the Clinic, and the two are linked by a fiber-optic connection.

The medical industry has some experience with file format standardization, but it was not particularly positive. When digital medical imaging devices were introduced in the 1970s, all vendors had proprietary file formats that were designed to lock their users into their technology, and it was a successful strategy for the vendors. When the users wanted to move data between vendors, the American College of Radiology and the National Electrical Manufacturers Association collaborated on the development of the Digital Imaging and Communications in Medicine image format standard, or DICOM [Digital Imaging 5]. The problem with DICOM, reportedly, is that the standard is implemented differently by each manufacturer, and proprietary extensions have been added by manufacturers, so that interoperability is still not achieved. For this and other reasons, archiving of medical images is still done in proprietary image formats.

It is interesting to note that digital medical images are "born archival" with respect to essential metadata. That is, all of a patient's identifying data – name, address, date of birth, attending physician, billing account, etc. – is collected ***before*** the diagnostic image is captured, and the metadata is forever associated with a captured image. This contrasts with the motion picture industry's practice of generating metadata ***after*** image capture, and the associated downstream difficulties with metadata management.

4.4 Earth Science

EROS stores film today in archives for the same reason Hollywood saves film: film can be reasonably preserved for 100 years or more.

THE CENTER FOR EARTH RESOURCES OBSERVATION AND SCIENCE (EROS) is a data management, systems development, and research field center for the U.S. Geological Survey's (USGS) Geography Discipline. Organizationally, the USGS is a bureau of the U.S. Department of the Interior. The archive contains aerial photography and satellite remote sensing data of the Earth's land surface. The EROS mission is to preserve this data "permanently" and make it easily accessible and readily available for study. Residing in the USGS' EROS Data Center near Sioux Falls, South Dakota – one of the largest computer complexes within the Interior Department – is the National Satellite Land Remote Sensing Data Archive (NSLRSDA), a comprehensive, permanent, and impartial record of the planet's land surface derived from 40+ years of satellite remote sensing. Aside from the larger question of change at the global scale, NSLRSDA permits scientists to study water, energy, and mineral resource problems over time; to help protect environmental quality; and to contribute to prudent, orderly management and development of our nation's natural resources.

Over the past three decades the U.S. government has invested money to acquire and distribute data worldwide from the Landsat series of satellites – more than 630 terabytes of which are held at the EROS Data Center. The archive also includes more than 28 terabytes of data from the Advanced Very High Resolution Radiometer (AVHRR) carried aboard National Oceanic Atmospheric Administration's polar orbiting weather satellites, and more than 880,000 declassified intelligence satellite photographs.

The primary objective of NSLRSDA is to preserve entrusted data records "permanently" and to distribute this data on demand to a worldwide community of scientific users. As a result, the EROS Data Center has become a world leader not only in techniques of archiving remotely sensed data, but also in getting the data to end-users quickly, in forms they can use, at costs they can bear. According to the archivist at EROS, every advance in online distribution, in storage media, in applications research, or in cost-saving delivery technologies means more people can use the data. As demand increases, user expectations about delivery times and efficiency rise.

EROS archives also contain approximately 4 million satellite images of global scale and 8 million aerial images of the U.S. Images held are stored both on film and digital media, but almost all new images are digital. In 2004, the archive included 80,000 pieces of film, and by early 2007, the film archive had grown to 110,000 pieces as other agencies send their collections to EROS for archival preservation and scientific access. Film assets are preserved in climate-controlled film vaults that have been inspected by the National Archives, which estimated a 100+ year shelf life for these film assets.

In 2004, EROS archives held roughly 2 petabytes of digital image data "nearline and online" in robotic data tape library systems and magnetic hard drive arrays. It took 30 years for the archive to reach this size. In 2004, the archive was growing at the rate of 2 terabytes per day, and EROS forecast a doubling of the digital archive in just four years. As of this writing, EROS archives hold more than 3 petabytes, on track to match the 2004 forecast, and the archives continue to grow at the rate of 2 terabytes per day.

EROS stores film today in archives for the same reason Hollywood saves film: film can be reasonably preserved for 100 years or more. The question is: at the end of 100 years' life of film in archives, does one make a new film copy or make a digital copy? Currently, EROS does not keep high-resolution scanned data from original film images; it scans on demand from the film archives, with full-resolution scans at approximately 7904 x 8512 pixels, approximately 800 megabytes per frame, at a cost of $20 to $30 per frame.

EROS provided some interesting observations from its experience building and operating a large digital archive: the operating cost is proportionate to the number of times the data is read, and the risk of losing data is proportionate to the number of

4.4 Earth Science *continued*

times data tape is accessed. EROS expressed similar opinions to other digital archive operators that optical disks appear attractive from a cost perspective but they do not have suitable long-term reliability characteristics, and power consumption of magnetic hard drives is a growing cost. Furthermore, EROS said it was important not to become dependent on a single technology. They related a story about two different, large digital archives that chose a technology called CREO which had but one supplier. The supplier went out of business, requiring an immediate and unplanned migration for both of these organizations – one European and one Canadian – at a cost of millions of dollars. They emphasized that this was a lesson not to be forgotten.

EROS also has experience with data migration. They prefer to invest in the migration costs of a collection rather than potentially experience the "cost" of losing a collection, and they believe migration works as a possible strategy to ensure long-term access. Their early data migrations were very expensive, took a long time to execute and were performed at 7- to 10-year intervals before 1992. Since 1992, migration has taken place every 3 to 5 years, and based on lessons learned and investments in robotic library systems (they are moving toward the SUN T10000 tape technology) and other efficiency-enhancing technologies, migration at EROS is getting easier, faster and less expensive, even after including 100% read-after-write data verification during migrations.

Since the events of September 11, 2001, a higher priority has been placed on a full offsite system – a third archive – for disaster protection. But EROS also learned as a result of the 9/11 attacks that air travel can be disrupted for extended periods. Therefore, it is more desirable to build offsite data storage within driving distance. Today, EROS has only a few terabytes offsite, but is planning to establish a 100% redundant offsite archive, such that EROS will have three archives:

- *First copy – near-line/online, on a robotic tape library or on magnetic hard drives*
- *Second copy – offline "basement tapes"*
- *Third copy – physically offsite*

Their goal is to keep the best-quality data, and fewest versions of the data in the archives.

In the opinion of the people interviewed at EROS, it is inevitable that distribution library and long-term preservation functions/systems will merge.

4.5 Supercomputing

The primary causes of digital archive loss are human error and magnetic disk hardware failures.

THE SAN DIEGO SUPERCOMPUTER CENTER (SDSC) operates three supercomputers for the national research community and provides supercomputing services for a range of extreme computing needs such as astrophysics visualization, bioinformatics and other science and engineering disciplines. The SDSC also operates a large hybrid storage system with 2.5 petabytes of online magnetic disk storage, and 5 petabytes of near-line data tape storage, with a capacity of 25 petabytes for the robotic tape storage system. For perspective, 25 petabytes is enough to store over 3,000 uncompressed 4K digital motion picture masters, or approximately 5 to 6 years of MPAA-rated motion pictures. SDSC's storage volume is doubling every 14 months, and they expect it to grow to 10 petabytes by the end of 2008.

According to those interviewed at the SDSC, data migration, or the copying of data to new storage devices and media from those becoming obsolete, is a fact of life for them, "like painting the Golden Gate Bridge," in that one must continually repeat the process to avoid decay. SDSC's migration period is five years, and they say one of the benefits of migration is that the newer storage technology is faster and denser than the old technology being replaced.

One of the factors determining the migration period is concern about "bit error rate" (BER), which is a fundamental measure of a storage medium's reliability and integrity of the recorded data. BER generally increases with the age of the storage media, and therefore the oldest data tapes in SDSC's collection are only 6 to 7 years old. However, they say the BER is nearly irrelevant to long-term preservation and that the primary causes of digital archive loss are human error and magnetic disk hardware failures. SDSC relies on data migration and a program of ongoing data verification to ensure data preservation. But there is a level of uncertainty to data verification because the act of reading data for verification increases the probability of error, but *not* reading data for verification can allow latent errors to accumulate unnoticed. This topic was covered in detail at the 2006 Eurosys Conference [Baker, et al. 3].

SDSC's biggest concern is not with technological obsolescence – they think the biggest risk to data preservation is gaps in funding for system maintenance and data migration. To address this concern, SDSC expects to begin allocating storage costs to system users, instead of just billing for computing power as is currently the practice.

Computer networking is a factor in the SDSC archiving picture. In a collaboration among SDSC, the National Center for Atmospheric Research located in Boulder, Colorado, and the University of Maryland, these three organizations have agreed to share archive resources that will only be accessed if one party loses its primary copy. Despite some opinions that archives should not connect to networks, those interviewed at SDSC recommend that operators of digital archives learn how to use networking because they feel it is faster and more efficient for debugging and problem solving.

SDSC is also collaborating with the Library of Congress on a pilot "third party repository" project designed to address certain risk scenarios, e.g., streamlining the number of file formats and file systems to support. In general, they think very-long-term data preservation and "cold storage" digital archiving are unsolved challenges suitable for future research.

4.6 Summary

"Don't make the same mistakes we made by letting all the different vendors create proprietary formats."

THIS SURVEY OF OTHER INDUSTRIES WITH LARGE DATA STORAGE and long-term preservation needs revealed a set of common issues that also arise in the motion picture industry's conversations about digital archiving, and it is worthwhile to summarize these issues as well as the advice offered by those who by now have a substantial amount of firsthand experience.

There is both consensus and disagreement among those interviewed on key issues of long-term digital data preservation.

4.6.1 Consensus View

There is general agreement on the part of those interviewed from outside the motion picture industry that:

- *Multiple copies of important digital data should be maintained*
- *The stakeholders, not the vendors, should drive requirements and standards*
- *The total cost of ownership is much more than just media costs*
- *The cost of labor, and secondarily the cost of electricity, not technology, are the limiting economic factors in digital archives*
- *There is definite economy of scale in digital archiving systems*
- *The number of file formats and file systems used should be minimized (and they should be chosen carefully) to keep labor costs down*
- *An extensible file system should be chosen to keep down long-term management costs*
- *Economics force an ongoing assessment of future value of assets each time a major data migration is done*
- *Good project management is essential in all migrations*
- *Unproven or exotic technology should not be used*
- *Biodiversity, or spreading technological risk across several different technologies, is important; i.e., do not be dependent on a single vendor*
- *It is very difficult to change vendors after even one year of using a new system, so the initial choice is critical. Make sure to negotiate up front in order to continue using archival system software after its license expires*
- *Currently there is no digital alternative to analog film archiving if the goal is "store and ignore" long-term preservation for 50 to 100 years*

Of particular note is that everyone interviewed for this report agreed that no one can make a perfect choice of what to save and what to discard, or how valuable an asset will be in the future. This issue, as it applies to the motion picture industry, is discussed in much more detail in Section 6.3.

4.6.2 Unresolved Issues

In discussions with people outside the motion picture industry, a wide range of opinions surfaced on a number of issues:

- *Is data migration most effectively done in-house or on an outsource basis?*
- *Is data compression an important technical consideration? (Although those interviewed all supported "no compression" unless an asset is "born compressed")*
- *What data should be saved, and what should not be saved?*
- *What level of geographical separation should be achieved; i.e., how far apart is far enough?*
- *Should a primary and backup archive system be connected, or not be connected, by a network?*
- *Are standardized file formats necessary?*
- *What is the best digital preservation strategy?*

The issues of standardization and policy-setting merit further discussion.

Standards and Policy

With respect to standardizing file formats, some recommend converting everything to a normalized or universal archive file format upon ingest, usually based on a widely accepted standard of the day. But there are many contra-indicating examples of standards not staying "standard" or going obsolete. Some recommend archiving in the original file format submitted to the archive. This avoids having to normalize formats upon ingest. This approach requires the archive to be capable of holding many different file formats, including, potentially, original data creation (acquisition) formats, intermediate data processing (postproduction) formats and final data delivery (distribution) formats.

Returning to the motion picture industry, one of the major Hollywood studios stated quite clearly that it is not concerned about picking an eternal universal archiving format, nor is it worried about compatibility between different archives. When they need to, they say, they will convert, or "transcode," formats. They say it is better if converting can be avoided, but it is not an insurmountable obstacle when necessary. It "just costs money and time." This studio recognizes that it does a lot of format transcoding today, and they assume this transcoding requirement continues as a "fact of life," but probably will become even easier and faster in the digital future.

Some of the people interviewed for this report believe the stakeholders – the owners and the archivists – should try to influence any standardization process to their advantage because they will be paying for the long-term preservation costs and thus should have a large say in what is created. They said, "Don't make the same mistakes we made by letting all the different vendors create proprietary formats."

Setting archive policy involves utilizing common records-management practices (Association of Records Managers and Administrators and the Society of American Archivists are good places to start) and utilizing them along with as many others as possible. Having stakeholders involved will increase the chances of gaining consensus. Implementation will involve an ongoing educational element. This is never an easy task, and although it becomes less onerous over time, it never goes away. Gaining acceptance and support from the highest management levels is essential for success. Without upper-management support, standardizing archival policies will be very difficult.

5 Archiving in the Changing Environment

Archiving digital data requires a more active management approach, and a more collaborative partnership among producers, archivists and users to exploit its full benefits.

DIGITAL ARCHIVING IS NOT JUST A MATTER OF ARCHIVING DIGITAL assets by putting digital storage media (magnetic hard drive, magnetic data tape or optical disk) on a shelf next to existing analog archives (film). The long-term accessibility of digital assets on magnetic data tape or magnetic hard drives or optical disks cannot be reliably protected for the long term just by keeping the humidity and temperature of the archiving environment within an acceptable range. Archiving digital data requires a more active management approach, and a more collaborative partnership among producers, archivists and users to exploit its full benefits.

Accessing the data stored on digital media requires access to the digital tools that "go with" the archived data. For example, early digital data from the NASA Viking probes launched in 1975 was transmitted from Mars back to the Jet Propulsion Lab in Pasadena, California, where it was recorded on magnetic data tape, analyzed by scientists at the time and then archived in a cool, dry data warehouse and left undisturbed until 1999 when USC neurobiologist Joseph Miller asked NASA to check some of the old Viking data. NASA found the tapes he requested, but could not find any way to read them. It turns out the data, despite being only about 25 years old, was in a format NASA had long since forgotten about. Or, as Miller puts it, "The programmers who knew it had all retired or died." Luckily, Miller was able to cobble together about a third of the data and get some useful results off the Viking tapes, but only because he also found a partial set of reference notes and records printed on paper that had been put away with the tapes [Kushner 3]. Overall, this incident with NASA's Viking data was an important warning bell about the dangers of what some have called "data extinction" and stimulated the development of a data reference model called the Open Archival Information System (OAIS), designed to protect data assets within the U.S. federal government through systematic data migration.

Interactive media, especially when designed for use on custom-made hardware and software, are exposed to another type of long-term threat for digital archiving. For example, the BBC Domesday Project was a pair of interactive videodiscs made by the BBC in London to celebrate the 900th anniversary of the original *Domesday Book.* It was one of the major interactive projects of its time, involving the work of 60 BBC staff, a budget of 2 million pounds and the volunteer efforts of thousands of British schoolchildren and teachers. The modern Domesday contained text, photographs, video, maps, data and a controlling computer program to bind it all together. The final package was published on two custom-designed laser disks with the special controlling software designed for the BBC Micro, a popular microcomputer. This software program was composed of 70,000 lines of custom code written in BCPL, a forerunner of the widely used C programming language. Within 15 years, it was impossible to use the "digital" Domesday, as compared to the original *Domesday Book* which was handwritten, probably by a single monk in 1086 and which is still readable (in Latin) if one goes to the UK's National Archives, where it has been preserved. However, in 2002, a research project by the University of Leeds and the University of Michigan managed to successfully emulate the original BBC system using modern hardware and software, one of the pioneering efforts in digital "archaeology" that enabled continuing access to old, nearly "extinct" digital media assets.

These ominous examples epitomize the difficulties faced in maintaining digital data accessibility over a long period of time. There are several similar stories circulating in the motion picture industry that ultimately had happy endings, but they foreshadow the possibility of more dire consequences in the absence of adequate digital preservation practices. To understand the underlying reasons for these difficulties, it is necessary to understand certain technical and operational aspects of digital storage technologies and the systems built around them.

5.1 Digital Storage Technology

Magnetic hard drives are designed to be "powered on and spinning," and cannot just be stored on a shelf for long periods of time.

MANY TECHNICAL TERMS AND ASSUMPTIONS are thrown about in any conversation about archiving and preserving digital data. The following section presents a condensed summary of practical information on various storage technologies, their reliability and other factors that affect the access lifetime of important digital data.

There are four primary digital storage media in professional use today: magnetic hard drives, digital data tape, digital videotape, and recordable optical disk. Solid-state memory devices, such as those used in digital still cameras and more recently in ENG and Digital Cinema acquisition, are not considered in this discussion because their storage densities (and therefore cost-effectiveness) are not likely to make them an influencing factor for motion picture archiving in the foreseeable future.

Magnetic hard drives

Also called "hard disks," "hard drives," or just "drives," magnetic hard drives have shown an impressive increase in storage capacity over the last 20 years and are the first choice for high-speed online storage. The earliest drives available for personal computers stored 5 megabytes (the size of a single digital photograph from today's consumer digital still cameras) and cost $1,500. As of this writing, 750-gigabyte drives are available for $269, and the annual 30% storage density increase continues.

Long-term magnetic disk storage capacity and cost trends are both favorable from the point of view of high-volume digital data producers. Conventional wisdom is that the cost per bit of magnetic storage is declining 40% per year or more. This is a long-term (40-plus years) trend that is expected to continue at least until 2025 or 2030. In other words, by 2020, if long trends continue unabated, a terabyte disk can be expected to cost $7.50 to $15 and a 1-petabyte disk (1,000 terabytes or enough to store over 100 uncompressed 4K digital motion picture masters) only $7,500 to $15,000. However, after 10 to 15 years, current magnetic recording technology could hit fundamental technical barriers, so it is not feasible to estimate hard disk trends beyond this period of time.

It should be noted that magnetic hard drives are designed to be "powered on and spinning," and cannot just be stored on a shelf for long periods of time. The drives' internal lubrication must be occasionally redistributed across the data recording surface through normal operation of the drive, otherwise they can develop "stiction" problems where internal components mechanically lock up. New power-saving strategies such as Massive Array of Idle Disks (MAID) attempt to address this problem at the cost of increased access time, although individual drive units still have a limited operational lifetime.

Digital Data Tape

The three leading data tape formats for digital archiving are Advanced Intelligent Tape (AIT), Digital Linear Tape (DLT), and Linear Tape-Open (LTO). Of the three, LTO, an open-format tape storage technology developed by Hewlett-Packard (HP), International Business Machines (IBM), and Seagate (which spun off its data tape business as Certance in 2000, subsequently acquired by Quantum in 2004), is the dominant format used in the motion picture industry and also has 82% market share in the mid-range tape drive segment [Mellor]. The term "open-format" means that users have access to multiple sources of storage media products that will be compatible. The high-capacity implementation of LTO technology is known as the LTO Ultrium format.

LTO Ultrium technology has evolved through several generations. The current LTO4, which became available in 2007, has a native capacity of 800 gigabytes per cartridge (1.6 terabytes using built-in data compression) and a maximum transfer rate of 240 megabytes per second. LTO5 and LTO6, still under development, are each expected to successively double LTO4's storage capacity and data transfer rate. LTO3 and its predecessor LTO2, each with lower capacities and transfer rates, are in wide use throughout the motion picture industry.

Technically, LTO offers faster access times than DLT. In addition, LTO features larger capacity per cartridge, higher transfer rates, multi-vendor interoperability, and a clear multigenerational technology roadmap that promises two generations of backward read-compatibility. For example, LTO1 tapes from

5.1 Digital Storage Technology *continued*

2000 are still readable on LTO3 drives introduced in 2004, but will no longer be readable on the new LTO4 devices due to practical limits of the physical media and electromechanical drive mechanisms.

Even executives of Quantum, the leading vendor of DLT tape, agree that LTO tape has won the format war for large-scale digital archiving. This is why Quantum acquired Certance in 2004 and publicly announced all its future investments in mid-range tape will focus on LTO [Global 10]. As a result, Quantum expects its own DLT tape will have a shorter commercial life than LTO.

According to Sun Microsystems, its Titanium 10000 (T10K) data tape system has been developed to meet the needs of "enterprise-class" storage applications. Enterprise-class storage is better for mission-critical applications, such as data centers and valuable archives because the data transfer rate is higher, the cartridges are more durable, and all components are manufactured to tighter specifications for more reliable, consistent operations and longer life cycles. They also point to slightly superior bit error rate for the T10K format. On the other hand, according to some of Sun's competitors, the enterprise market segment is being eroded from below by mid-range LTO which has been steadily improving in terms of reliability, durability, bit error rate and error detection/correction to the point that the advantages of enterprise-class products are less meaningful. They say mid-range LTO is good enough for digital archiving applications, and significantly less expensive. A 2006 study on data tape technologies prepared for the United States Geological Survey appears to confirm this point of view [Science Applications 13].

Some industry executives interviewed for this report worry that the collapse of the consumer market for VHS tape will weaken research and development investment in magnetic tape in general, and therefore slow the historical downward dollar-per-bit trends enjoyed by professional data tape for the past 20 years. They suggest that prices for professional data tape stock and its ingredients such as magnetic coatings, binders and lubricants will go up as consumer tape volumes decline. To keep up the pace of progress, data tape makers may have to invest more in their own fundamental R&D, to be amortized through higher professional tape prices.

According to an executive with Imation, a large provider of data storage products spun off from 3M in 1996, tape manufacturing operations for audio/visual applications were discontinued by Imation at the time of the spin-off because it had become a low-profit commodity business. Imation saw no danger, however, to its growing and profitable data tape business.

One final note about the large "robotic library" storage systems built around any data tape format: some postproduction facilities interviewed for this report state that the benefits of a standardized data tape format are lost when the systems that control the data tape drives write custom, library-specific data to the tapes. This locks the facility into a single vendor's storage library product, and makes interchange impossible with customers and other facilities that may use another vendor's library system.

Digital Videotape

HDCAM SR and D5 are the only high-end professional videotape formats being used in motion picture mastering today, although HDCAM SR is the dominant format currently in use, especially for digital motion picture acquisition. Introduced by Sony in 2003, HDCAM SR can record HDTV images (1920 x 1080 pixels), which is slightly less than Digital Cinema's "2K" pixel count (2048 x 1080), and uses MPEG-4 Studio Profile image compression. The D5 format, introduced by Panasonic in 1995, is also an HDTV system, although the format was recently upgraded by Panasonic to full "2K" pixel count and JPEG-2000 image compression, the same compression scheme used for Digital Cinema digital "prints." There are several technical differences between the formats, but these are beyond the scope of this report.

There is general agreement in the industry that there will be little or no new development of professional and consumer videotape formats as broadcast television continues to go "tapeless," although that transition is not without its digital storage issues [Kienzle, "Taking," 12]. The consensus is that although digital videotape stored in proper environmental conditions can last for at least 5 to 10 years (or longer), there may be no new videotape format to migrate to when the medium nears the end of its shelf life.

Optical Media

Optical storage technology in general is not keeping up with magnetic storage technology in terms of areal density, capacity per unit, or transfer rates. Optical disk is primarily a consumer technology, so cost per bit is very inexpensive – much lower than magnetic disk or data tape. But the rate of progress of new optical storage technologies is actually slower than that of magnetic tape and disk because of the need for broad standards to insure interoperability among many vendors' products, and because consumers are reluctant to commit to any technology if they think it is likely to become obsolete in just a few years.

In the motion picture industry, recordable DVDs (DVD-R) are currently preferred over rewriteable Magneto-Optical (MO) disks because they are less expensive and have higher capacity per unit. DVD-R offers an attractive cost per unit, but the relatively small capacity per unit of optical disks – between 4.7 and 8.5 gigabytes, depending on how they are used – relative to data tape that holds 400 to 800 gigabytes per cartridge, is a disadvantage for handling the large amounts of data generated in digital motion picture production.

The new generation blue-laser DVDs, capable of holding 35 to 50 gigabytes per disk, still have a much smaller capacity per unit than LTO4 tape cartridge. The related Ultra Density Optical (UDO) format, which uses a different recording technology than DVD, is capable of holding 30 gigabytes per disk cartridge today, and is expected to grow to 120GB by 2008, according to the manufacturer.

As a packaged media technology primarily targeting the large consumer market, the blue-laser DVD formats must fully standardize all aspects of their technology and stabilize as a storage medium in order to attract both content publishers and consumers in significant numbers to become profitable. This is not the case with magnetic data tape or magnetic hard drives, where technology advancement is continuing unabated and vendors win market share by being the first to offer higher speeds and higher capacities per unit.

WORM capability is one of the vaunted advantages of optical storage because with optical WORM there is no fear of electromagnetic interference (EMI) or accidental erasure. These are not particularly high-priority risks in most modern digital archiving applications except when the preservation of unaltered original data is legally mandated, as it is in so-called "compliance archiving" per the Sarbanes-Oxley rules discussed earlier. Magnetic tape and disk products with firmware-based WORM capabilities are also being introduced, which will further deflate the advantage of optical WORM for many users.

It has been difficult for users to get an unbiased forecast for the longevity of optical storage media. So NIST, together with the Library of Congress and with the support of the Optical Storage Technology Association, spent two years testing DVD-R disks (as well as the DVD's lower-capacity predecessor, the Compact Disc, or CD) from multiple manufacturers on multiple playback devices to understand their life expectancy characteristics. The study found that, generally speaking, both DVD-R and CD-R can be very stable, maintaining data availability for tens of years, although the measured data indicate that CD-R has a much better life expectancy compared to DVD-R: 100% of CDs tested have a life expectancy of over 15 years compared to 66% of DVDs with that life expectancy. The study also found that it is very difficult for users to identify which media on the market have better stability characteristics. The use of gold in a disk's recording layer significantly extends its life, but it also makes the disk five times as expensive as non-gold disks, which stifles market demand. Slowing the write speed or using stronger lasers for writing could also potentially increase the life expectancy of DVD-R, but this is considered unlikely, given the overwhelming pressure to reduce costs in order to grow consumer market share.

For all the reasons described, no large archives are known to use CD or DVD as their primary archival storage media. However, writeable CD and DVD are still widely used as non-permanent transfer and delivery vehicles for smallish amounts of digital media such as sound elements, photographs and oral histories. Furthermore, pressed CDs and DVDs are often submitted to the Library of Congress to satisfy copyright and mandatory deposit requirements. But this is not to say that optical storage technologies will never be adopted for large archival storage systems. At least one company has spent more than 12 years pioneering optical holographic storage, a fundamentally new technology with substantially greater density and theoretically longer life expectancy than CD or DVD. For archival applications, one of the potential advantages of holographic optical technology is that it is predicted to have 50-year longevity, based on the manufacturer's accelerated aging tests. Potentially, archives implemented with holographic optical storage will need to migrate data "only" every 20 years or so to accommodate changes in computer operating systems, file formats and application software. This is much less frequently than is currently recommended for magnetic tape archives, although due to the novelty of holographic storage technology, even its manufacturer's executives are hesitant to recommend it as a primary archival format. They agree that technology choices for archiving must be conservative and give priority to proven reliability and multi-vendor support.

As a point of reference, the LTO tape consortium claims their product has a life expectancy of 30 years based on accelerated aging tests. The National Media Lab has also estimated a 30-year life expectancy for magnetic tape based on its own testing [Van Bogart 34]. Nonetheless, leading tape vendors and even NARA recommend data migration of digital assets on magnetic tape as frequently as every 5 to 10 years.

5.2 Risks and Threats to Digital Data

TWO QUESTIONS ARE KEY TO UNDERSTANDING why digital archives cannot be preserved over the long term using a "store and ignore" management philosophy: "Is there any way to store a digital object for 100 years with no maintenance?" Secondly, "Is the bit density enough to hold what you want to preserve at a price you can afford?"

If one could make a "black box" with even 100-year lifespan components that could read data reliably without introducing any errors, required no maintenance, and offered sufficient bit density at an affordable price, everyone would buy it. After filling the black box with their most valuable "permanent" assets, one of the first things prudent archivists would do is create several replicas in multiple black boxes and geographically separate them to guarantee viability and enable the archive to become self-healing. If the archive format preserved both bits and needed application software together with contextual metadata, there would be no need for periodic data migration or system emulation. But there's a new danger inherent in this approach. If the 100-year-lifespan "box" fails at 99 years, no one involved in its development or capable of system repair is likely to remain alive. To avoid this risk, it would be necessary to continuously audit the integrity of the box to ensure that the archived assets can move to a new box before the old box fails. This points to the need to sustain a supportive human community around a digital archive with the requisite know-how in order to ensure its ability to preserve, renew and repair the system within which digital assets are stored.

Digital assets in the real world are not kept in "black boxes" with 100-year longevity. They are stored on physical media with longevities of 30 years or less, and are vulnerable to heat, humidity, static electricity and electromagnetic fields. The digital contents can be degraded by accumulating unnoticed statistically occurring "natural" errors, by corruption induced by processing or communication errors, or by malicious viruses or human action. Digital media cannot be viewed with the naked eye. As such, it is susceptible to misidentification, frequently poorly described (incomplete labeling and metadata), and therefore difficult to track. And digital assets are hard to maintain long-term because media, hardware and software can all become obsolete. This is commonly caused by the evolutionary loss of compatibility between data in the archive and the software applications that originally created the data. Sometimes proprietary formats in an archive are simply abandoned when a company goes out

Digital Archive Layers

LIFESPAN	HARDWARE	SOFTWARE
3→5 YEARS	HOST COMPUTER	• APPLICATION SOFTWARE • OPERATING SYSTEM • DEVICE DRIVERS
5→10+ YEARS	PHYSICAL INTERFACE	• INTERFACE FIRMWARE
3→5 YEARS	MEDIA DRIVE	• DRIVE CONTROL FIRMWARE
.5→10 YEARS	MEDIA	• FILE SYSTEM • DATA FILE FORMAT • PHYSICAL RECORDING FORMAT
VARIES	TRAINED PERSONNEL	
VARIES	FUNDING	

of business. A digital archive may have many "layers," each with its own finite lifespan as shown in the diagram on the facing page. When the end of the lifespan is reached, not only does the layer have to be replaced, but the adjacent layers may have to be modified to be compatible with the replacement layer. Thus, a digital archive built with today's digital technologies can only assure digital "permanence" via an ongoing and systematic preservation process.

The rapid and seemingly endless improvement in the price per bit of digital data storage tends to give the impression that storage is forever getting cheaper, so why worry about the "data explosion"? There are several reasons why the overall storage picture is not as simple as this might make it seem:

Increasing demand for storage offsets reduced media cost

Along with the increase in available storage comes a corresponding increase in the demand for storage. In the UC Berkeley digital data generation study discussed earlier, it was found that of the 5 exabytes of new data created in 2002, 92% was recorded on magnetic media, 7% on film, and the remaining 1% split between paper and optical media. Overall, UCB researchers estimated that new stored information grew about 30% from 1999 to 2002.

From the relatively narrow view of the motion picture industry, one only need consider the amount of data generated by the new generation of 4K digital motion picture cameras and digital postproduction process (in the petabyte range) to understand that there will always be a way to generate more data, usually in excess of available storage. The demand is compounded by the need to duplicate important data for backup purposes.

Data transfer rates do not increase at the same rate as storage density

As the storage density grows, the speed at which the data gets on and off the storage media (transfer rate, or throughput) becomes more important. The need for increased throughput drives up the cost of the physical interface, network connections and computers attached to the disk drives. As with the demand for increased storage, throughput requirements increase with the need to make backup copies of important data.

Longevity characteristics do not always meet advertised specifications

Recent studies by Google [Pinheiro] and the Computer Science Department at Carnegie Mellon University [Schroeder and Gibson 15] present evidence that hard drives are not as reliable as manufacturers' data sheets suggest, nor do they follow the conventionally accepted "bathtub curve"[9] failure characteristic. To the contrary, these studies observe that large numbers of drives fail well before manufacturer-specified "mean time before failure" (MTBF), and show a low correlation between drive failure rates and high temperatures, a commonly assumed failure predictor.

The manager of a large digital image archive interviewed for this report who has purchased a great deal of both tape and disk over the years said that in his experience, the biggest problem with a magnetic hard drive is its short device life cycle, supposedly five years according to manufacturers, but only three years in practice. He recognizes that disk technology is driven by personal computing and consumer electronics markets characterized by very short product life cycles, so there is naturally quite a bit of product churn. In contrast, data tape drives are industrial products, with multi-year life cycles, and with some degree of backward compatibility and forward-looking roadmaps from vendors.

These empirical observations raise questions about the "accelerated age testing" methodologies used by storage product manufacturers to determine the life expectancy of their products, and suggest that there is no way of knowing whether a storage device or medium will, on average, last for the advertised period of time without actually seeing what happens during that entire time frame. It is worth repeating that both storage technology suppliers and end-users significantly de-rate the published life expectancies of all digital storage systems, usually planning on wholesale equipment and media replacement after as little as three years, with five to ten years as the most often quoted migration period.

Economic, Technical and Human Threats

A recent report by the National Research Council written for the National Archives [Natl. Research 59-69], presents the notion of threat modeling and threat countering as a core consideration in the design of digital preservation systems. These threats are further detailed in a paper on digital preservation system requirements published by the Stanford University Libraries [Rosenthal 3], and are worth summarizing here for the benefit of those responsible for preserving digital motion picture assets:

Economic threat:

- Funding loss: Digital preservation systems require ongoing funding for equipment maintenance,

[9] *The bathtub curve, used in reliability engineering, predicts early "infant mortality" failures, followed by a constant failure rate during a product's useful life, followed by an increasing failure rate. A graphed curve of these failure characteristics looks like a bathtub – high at the ends and low in the middle.*

5.2 Risks and Threats to Digital Data *continued*

All of today's technology products, including storage media, hardware and software, have a finite lifetime, and the time required to migrate can exceed the data's lifetime.

replacement, operating staff and power, among other things. Every commercial enterprise has its good and less-than-good years, and the occasional "benign neglect" that film archives can tolerate may result in data loss in a digital archive. There is no known tactic to fully mitigate this threat, although factors that affect the economy of operating a digital storage system are discussed in Section 6.

Technical threats:

- Data integrity: At the most basic level, the 0s and 1s that represent digital images and sound must be reliably stored and retrieved. Common failure modes that affect the integrity of the 0s and 1s preserved in digital archives are latent errors (errors lurking undetected), ingest errors (translation errors when digital data is brought into a digital system), and network communication errors (errors caused when digital data is moved between computers on a network). Regular auditing and authentication of the data and rigorous quality control procedures are effective means for dealing with these threats.
- Monoculture vulnerabilities: Just as a single animal species can be wiped out from a deadly virus, individual storage media or technologies can be (and have been) seriously impacted in the same way [Herman]. Biodiversity, or the practice of utilizing several different media and technologies for digital storage, significantly reduces this threat [Baker 8; Science Applications 31].
- Single point-of-failure: Storing a single copy of data in just one location is dangerous. Storage solutions should include sufficient redundancy to protect from data loss resulting from the failure of media, hardware, software, network services and/or natural disasters.
- Obsolescence: All of today's technology products, including storage media, hardware and software, have a finite lifetime, and the time required to migrate can exceed the data's lifetime.
- Limited or no data compression: A popular technique for reducing storage and transmission bandwidth needs is to apply mathematical data reduction techniques to image and sound data. These techniques range from "mathematically lossless" (every single bit is recovered when decompressed), to "perceptually lossless" (not every bit is recovered, but one cannot see or hear the difference between the decompressed content and the original), to "lossy" (perceptual artifacts exist in the decompressed content). The effects of compression must be well understood if used.
- No risk of encryption key loss: There is much discussion today on safeguarding digital content through the use of data encryption methods. All encryption schemes require a digital key to "unlock" the encrypted content. If encryption is deemed necessary, then steps must be taken to eliminate the risk of losing the key, which is tantamount to losing the content it is intended to unlock. In general, there is broad consensus among those interviewed for this report that encrypting digital archives increases long-term complexity and risk.

Human threats:

- Operator error/malicious action: Today's technology requires human involvement in many aspects of digital storage system operations. And being human means mistakes can and will be made. Furthermore, systems can be attacked by disgruntled employees or hackers simply doing it for fun. Procedures for protecting against losing media, unauthorized internal and external system access, reliance on a single employee, and storing multiple copies of important data in separate locations not controlled from a single place can be effective in managing the human element. Documentation of procedures and system implementation details can also protect against organizational failures that often occur when companies are sold or merged, or when key employees move on.

5.3 Digital Preservation Strategies

Data migration is the most widely practiced digital preservation strategy today.

BROADLY SPEAKING, DIGITAL ARCHIVING experts have identified several preservation strategies that address either the general survivability of digital data or technical obsolescence. Two of those strategies are discussed here: migration and emulation.

Migration

Data migration involves the transfer of data from old physical media to new physical media, a process that often (but not always) includes updating file formats for currency with the latest-generation operating system and/or software applications. Older digital assets that are properly migrated will be accessible for some time into the future, until technological obsolescence motivates another migration cycle. Migration is designed to avoid having to preserve old devices to read the old storage media, old application software to interpret the old data, and old hardware to run the old software to use the old data. If everything goes smoothly, after migration the new data replaces the old data.

A major drawback to migration is that while copying data from one physical medium to another, or while converting digital assets from one file format to another, some data (or related metadata) might be lost. To make data migration a lossless, errorless process, migration procedures typically incorporate various quality control and auditing routines to ensure accuracy, integrity and completeness of the data throughout the migration process. Systemization of the migration process, including policy-driven automation routines, reportedly can be effective in reducing human errors and increasing the speed of migration. In practice, the emerging trend is to "migrate all the time" as a background task.

Migration of archived assets by replicating them on new media is a preservation strategy for both analog and digital assets. An advantage of migration as a digital preservation strategy is that digital assets will always be available in the form that is most widely accepted, and current hardware and software will be able to render these digital assets with little difficulty. In the case of analog assets, migration can cause the loss of image and sound quality over successive generations. In the case of digital archiving, data migration done correctly is lossless every time. Data migration can occur between instances of the same type of storage medium, from one medium to another, and from one format to another. Data migration can be effective against media and hardware failures. For example, the tape backup of the contents of a magnetic hard drive involves data migration between different mediums.

The goal of archival data migration is preservation of the full information content, not just the bits. For example, the Open Archival Information System (OAIS), pioneered by NASA and others, defines "preservation description information" that should be included in the data migration process. This includes provenance information that describes the source of content, who has had custody of it, its history, how the content relates to other information outside the archive, and fixity information that protects the content from undocumented alteration.

Data migration can be motivated by a variety of factors such as physical media decay, media or media drive obsolescence, even prior to complete system obsolescence. Older media drives may face escalating maintenance costs, there may be new user service requirements, or new media formats and/or file formats are introduced that are more compatible with users' technology and applications. The list of motivating factors goes on, and therefore data migration is the most widely practiced digital preservation strategy today.

Emulation

Emulation preserves the original data format, often on the original physical medium, and provides the user with tools that enable the data to be read even after the original file format, storage medium, application program or host hardware is no longer supported. Emulation refers to the ability of one system or device to imitate another system or device. In practice, emulation involves writing software that runs on new hardware to make it appear as if it is an old system, translating between the two, allowing old data on old media to be "tricked" into working on a new system after the old underlying system has become obsolete. For example, new storage devices added to existing digital storage systems are often built with the ability to emulate an older storage device, so

5.3 Digital Preservation Strategies *continued*

that the new storage technology can be integrated into the pre-existing software control and automation infrastructure of the system, thereby hiding the evolution of the infrastructure from the end-user. Emulation strategies for digital preservation are designed to minimize the need to copy, transfer, transform or otherwise "update" the digital assets in an archive. Digital archivists can use emulation strategies to reduce or even (theoretically) eliminate data migration. However, a serious drawback to emulation is the cost and complexity of developing and maintaining emulation tools. To avoid the risk that old emulation tools will not work on future computer platforms, software engineers must keep adapting and updating them.

While emulation has not been widely adopted as the primary digital preservation strategy for major digital archives to date, researchers at the University of Michigan and the University of Leeds in the UK, working with the BBC on the Domesday Project (as discussed earlier in this report), have demonstrated that emulation can preserve the consumer's experience of interactive multimedia based on older videodiscs and CD/DVD-ROM systems. They point to the need for emulation techniques in any effort to archive video games and hyperlinked rich-media documents.

This has led researchers, particularly some from IBM, to propose emulation strategies for long-term preservation based on the concept of a "Universal Virtual Computer" (UVC), a layer of software that remains the same on the "top side" facing the emulation tools while evolving as needed on the "bottom side" facing the hardware and operating system (OS) software to adapt to changes in technology. In this approach, digital asset data is archived with a very basic software program that decodes the data and returns the asset in a readable form using a future software application based on a logical view that is simple and self-contained enough to be interpreted without any specific software or hardware. Working with the National Library of the Netherlands, IBM has successfully shown a proof-of-concept of the UVC approach using electronic documents deposited in the library in the Adobe Acrobat electronic document format [Lorie 6].

Some argue that emulation, and its distant cousin encapsulation,[10] are just more complicated forms of data migration.

No one strategy is "best"

In considering emulation versus migration, experts agree that no one strategy is 'best' for long-term preservation of digital data. Both emulation and migration have pros and cons. In general, storage vendors have tended to promote migration, while computer and software vendors have tended to promote emulation. Some digital preservation researchers advocate a hybrid approach, combining both migration and emulation. For example, emulation uses a "root format" from which digital asset transfers and conversions can be generated even as hardware and software evolve. But sometimes new formats are just too attractive to pass up, so an archive might periodically migrate its data to the new better/faster format, which then becomes the new root format for subsequent emulation. Among operators of major digital archives we interviewed, migration is the overwhelmingly preferred strategy for digital preservation at this time. But these same experts recognize that emulation also has merit, and admit emulation has been under-explored as a strategy for long-term preservation. Perhaps migration is the more conservative strategy and emulation requires higher initial investment in software development.

[10] *Encapsulation is another digital preservation strategy that proposes "wrapping" a digital asset with instructions on how to be decoded.*

6 Digital Motion Picture Archiving Economics

THE ECONOMIC IMPACT OF USING digital technologies in motion picture mastering and acquisition cannot be fully understood without an understanding of the complete costs associated with depending on digital storage technology for the long-term accessibility of important digital data.

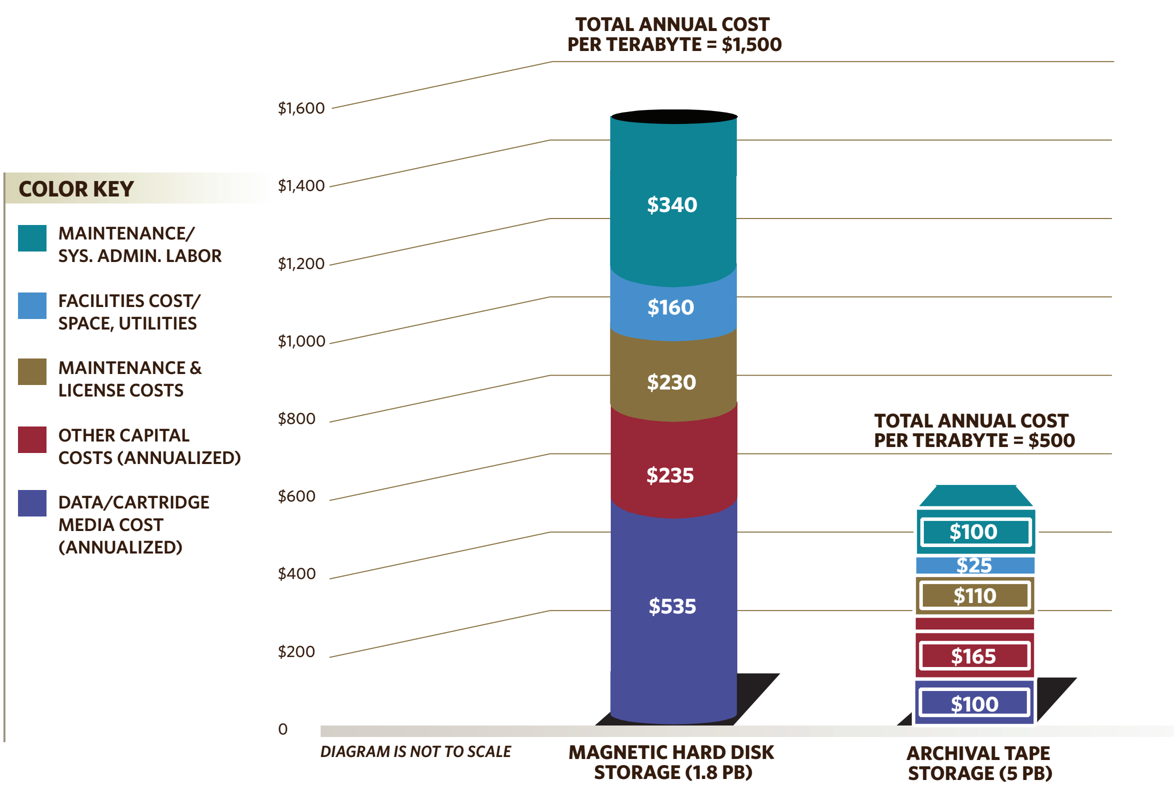
6.1 Digital Storage Economics

BASED ON CONVERSATIONS WITH SEVeral experts in digital archiving, it is clear that the economic model for digital archiving requires reconsideration of basic assumptions about both the costs and rewards of preservation. The total cost of ownership of operating a digital archive is typically expressed as $/terabyte/year. However, many vendors present their competitive advantages most favorably by simplifying the cost components to just the storage media and devices they sell, ignoring other costs that the user will inevitably face. Other vendors will compute costs based on an archive sized to fit their technology most economically. Inaccurate or incomplete cost analysis is not limited to vendors. Several archive operators interviewed for this report acknowledge that they have either not tried to compute a full cost analysis, or tried and gave up because expense information is hard to collect for administrative reasons when budgets and cost accounting are project-based and "stove-piped" due to organizational structures.

There are, however, several archival cost analyses from well-regarded organizations, breaking out all significant expense types and distinguishing between tape and disk storage. The San Diego Supercomputer Center recently published a paper that discusses a comprehensive cost model for its 25-petabyte capacity mixed magnetic disk/data tape storage system that currently holds approximately 7 petabytes [Moore 2].

The chart below shows the estimated normalized annual cost of delivering disk and tape storage at SDSC. The only costs not included are transaction costs; i.e., the cost of transmitting and receiving the data, networking/bandwidth costs, etc. Disk drive utilization is discounted from 100% to account for data overhead and operating efficiency – a consideration for any magnetic hard drive-based system.

Estimated Normalized Cost of Delivering Disk and Tape Storage at SDSC in 2006



6.1 Digital Storage Economics *continued*

The total cost-of-ownership calculation should include the cost of data replication – that is, multiple copies of data for protection against loss.

It is interesting to note that storage media costs are only 36% and 20% of the total annual operating costs for magnetic hard drive and data tape systems, respectively. It is also interesting to note that although data tape is one-fifth the cost of hard disks on a cost-per-bit basis, the annual expense for a large data tape storage system is one-third the cost of a hard disk-based system.

The numbers appear to scale in inverse proportion to the size of the storage system. The Cleveland Clinic reports its costs at $1,500/terabyte/year for a 1-petabyte mixed tape/disk system, and the Swedish National Archives' projected annual cost over five years for its 200 terabyte-capacity data tape storage system (less than 1/100th the capacity of the SDSC tape system) is over $11,000/terabyte/year, a 7-fold increase over the SDSC annual costs[11] [Palm 7].

As another reference point, Amazon.com, the large online retailer, recently introduced an online storage service called Simple Storage Service, or S3, targeted at software developers. Customers can upload data to the S3 service and pay a monthly storage fee of $0.15/month/gigabyte plus a transaction fee of $.10/gigabyte for data uploading and between $0.13 and $0.18/gigabyte for data access, depending on volume. This translates to $1,843/terabyte/year for data storage services, plus $102 per terabyte for initial data upload, plus between $133 and $184 per terabyte per access.

SOURCE	ARCHIVE SIZE	STORAGE TYPE	$/TB/YR	TRANSACTION $/TB
Amazon S3	Unknown	Hard drive	$1,843	$133 – $184
Cleveland Clinic	1 PB *(4 PB Capacity)*	Mixed tape/ hard drive	$1,500	Unknown
SDSC	7 PB *(25 PB Capacity)*	Mixed tape/ hard drive	$500 – $1,500	Unknown
Swedish Natl. Archives	40 TB *(200 TB Capacity)*	Tape	$11,344	Unknown

Representative Annual Total Cost of Ownership

The ultimate total cost-of-ownership calculation should include the cost of data replication; that is, multiple copies of data for protection against loss. For example, one hard drive copy and one tape copy at SDSC would be $2,000/terabyte/year. The S3 system, according to Amazon.com, replicates individual data objects across multiple storage "nodes" and physical locations, with at least two copies of data objects in existence at any one time.

Looking to the future, the SDSC study states that the cost differential between magnetic hard drive and data tape storage will likely diminish. One of the study's authors observed that data tape media costs are falling by half every 36 months, magnetic disk media costs are falling by half every 15 months, and a "crossover" in the two media costs is forecast in 2009-2010. But even if magnetic disk media costs less than data tape, it may be that data tape will remain best for "cold storage," given the higher power consumption per terabyte of magnetic hard disk storage and the expected long-term increase in the cost of electrical power.

[11] The Swedish National Archives numbers were calculated in 2005 and the SDSC numbers were calculated in 2007, so the relative difference might be somewhat less due to ongoing storage media cost-per-bit improvements.

6.2 Digital Motion Picture Storage Economics

IT IS IMPORTANT TO DETERMINE THE COST of archiving suitable elements for preserving and creating motion picture content well into the future. These elements are generally considered to be the master materials from which all downstream distribution materials are spawned, with an expected access timeframe of at least 100 years. This section of the report applies what was learned about the total costs of digital storage to the digital elements produced in typical motion picture productions.

To develop an understanding of the actual motion picture deliverables requiring long-term storage, two case studies were undertaken based on actual recent motion picture productions. One case study was a motion picture captured on film and digitally finished to distribution. The second case study was a motion picture that was captured digitally, or "born digital," and also finished digitally through to distribution. The scope of this analysis was limited to picture and sound elements created during production and postproduction that led to worldwide theatrical exhibition.

The film-capture production was chosen from among a number of average-length features (90 to 120 minutes) with an "average" budget (>$60 million) and few if any visual effects. The digital-capture production also met the same content criteria and was photographed using a modern digital motion picture camera generating digital frames at 1920 x 1080 pixel count. The camera output was recorded to HDCAM SR videotape (this case study preceded digital capture directly to magnetic hard disk), and the digital "masters" were also created at a pixel count of 1920 x 1080 and stored on magnetic hard drives and LTO data tape. The studios participating in the case studies provided complete inventory reports for each feature.

Each studio uses proprietary software systems to track archive and library assets. Since the two case studies contained inventory data from different studios, there was a need for a common reference of generated materials. This was accomplished by creating a generic hierarchy of materials for both picture and sound, which was then populated with elements described in the separate inventory data supplied by the participating studios. These picture and sound hierarchy charts are in the Appendix. The charts are color-coded to indicate to which storage category (archival or working library) each element is assigned. The source information for the hierarchy charts is an amalgamated (albeit typical) delivery schedule used by studios in third-party production agreements. The delivery schedule is a legal approach to describing the known "common sense" results of the production process and specifying how the studio expects to receive these items at the point of final delivery. A successful delivery is usually tied to the final payment, so producers are keen to understand the exact delivery requirements from the studio as early in the process as possible. In the end, a production company will deliver a mountain of film, paper, magnetic hard drives, DVDs, data tape and videotapes according to the schedule.

Executives from participating studios were extremely helpful in validating the accuracy of the results. One studio opened several cartons of aggregated materials to provide an average count of like media for the analysis. This average count is conservative, and it is used in the digital capture case study to estimate the average number of HDCAM SR camera original videotapes that are stored in a single carton.

The case study information is presented in a series of tables in the Appendix that summarize the number and type of elements, and their estimated annual storage costs. Because of varying practices between studios and production workflows, several assumptions were necessary regarding the calculation of the number of elements and the "byte count" of the digital elements. As with any case study, the results represent a snapshot of time, and while production practices continue to evolve, the data presented are still considered valid at this time, although the summary data presented in this section incorporates further assumptions (described later in this section) to reflect current industry trends.

In the summary cost analysis, two key definitions are used:

- ***"Archival"*** *is defined as storage of the master elements from which all downstream distribution materials can be created over a 100-year timeframe.*
- ***"Working Library"*** *storage is a broad term for elements that are generally kept on hand for distribution purposes.*

The participating studios save their elements in either archival storage conditions or working library conditions, depending on their preservation and near-term access policies.

The only picture element that continues to achieve broad consensus as the indisputable archival master picture element for a major motion picture is the YCM separation master on black-and-white polyester film stock. The current cost of creating a complete set of archival separation masters is estimated to be between $65,000 and $85,000, depending on service options.

With respect to base storage costs, physical element storage cost information was obtained from several companies engaged in that business, given the difficulty in determining accurate on-the-studio-lot storage costs.

6.2 Digital Motion Picture Storage Economics *continued*

The cost of storing 4K digital masters was found to be enormously higher – 1,100% higher – than the cost of storing film masters.

The cost figures for data storage came from the San Diego Supercomputer Center study discussed earlier in this section, which describes the lowest observed cost for a fully managed digital storage system with both online magnetic hard drive storage and near-line data tape storage. It is important to restate that this baseline cost represents only a single fully managed copy of the data.

The baseline storage costs used for this study are:

- *$4.80 per physical item per year for archival storage*
- *$1.80 per physical item per year for working library*
- *$500 per terabyte per year for near-line data tape storage*

Initial inspection and access costs are not included in the baseline film storage costs, nor are access or ingest costs included in the baseline digital storage costs because reliable information for the latter is not available. Nonetheless, these costs should be taken into account when considering the type and quantity of assets being stored.

The table on the facing page summarizes the annual storage costs, exclusive of ingest, inspection and access costs, for five common scenarios:

1. ***An "all film" production that generates no digital assets***
2. ***A film-captured, digitally finished production at 4K***
3. ***A digitally captured, digitally finished production using HDCAM SR videotape as the capture medium at 1920 x 1080***
4. ***A digitally captured, digitally finished production using an uncompressed digital data capture system at 2K***
5. ***A digitally captured, digitally finished production using an uncompressed digital data capture system at 4K***

The film-capture/digital finish production and 4K-captured productions produce 4K masters, and the 2K-captured productions produce 2K masters.[12] Three copies of the digital master are assumed, given the recommended practice of data replication, and this is consistent with the practice of archiving between two and five film masters, although three is most typical: a YCM, a finished negative, and an interpositive. The cost of producing the three film masters ($80,000 amortized over 100 years) is added to the annual storage cost of film. The finished master is assumed to be 120 minutes in duration for all scenarios; a shooting ratio of 25:1 is assumed as an industry average to calculate the amount of source material, and two copies of all digital source material is assumed to reflect current industry practice and insurance requirements.

Using current preservation methodology, the cost of storing 4K digital masters was found to be enormously higher – 1,100% higher – than the cost of storing film masters. The overall costs increase further with the use of magnetic hard drive data capture systems at 2K, and further still with 4K digital capture systems.

Although 1920 x 1080 capture using HDCAM SR videotape appears to be a cost-effective alternative to 4K, it is worth repeating that this cost reduction comes with a corresponding reduction in certain performance characteristics relative to film. These

[12] *For this report's calculations, a 4K frame is composed of 4096 x 2160 pixels, 48 bits per pixel; a 2K frame is composed of 2048 x 1080 pixels, 30 bits per pixel; and a 1920 x 1080 frame is composed of 1920 x 1080 pixels, 30 bits per pixel.*

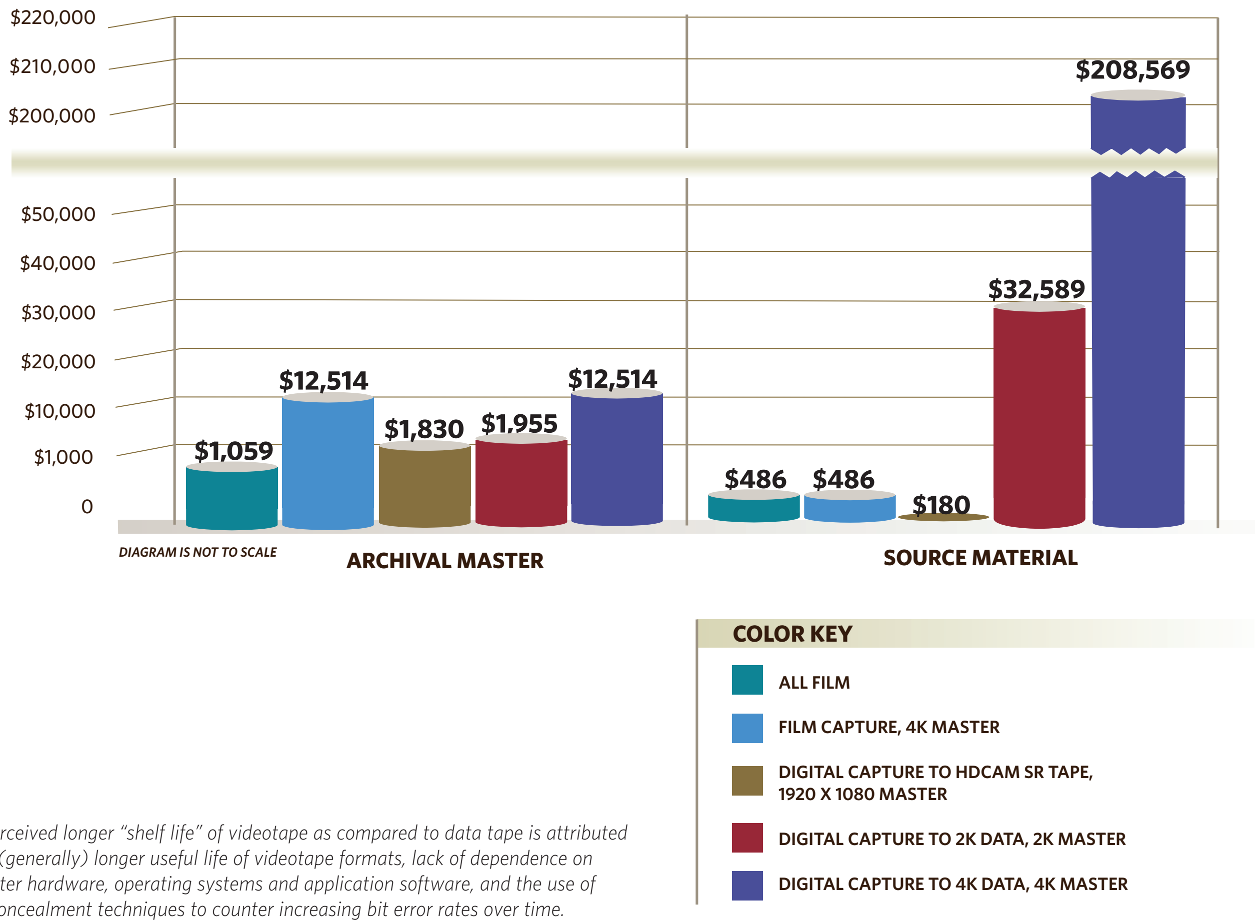
performance characteristics, and their effect on perceived image quality, are beyond the scope of this report, although there is significant and ongoing debate about these tradeoffs. Furthermore, the decision to migrate HDCAM SR source material would likely be made in approximately 10 years.[13] If the choice is made to copy the tapes to some newer videotape format (assuming such a videotape format is developed in the future), the cost to do so is estimated as follows:

- *Total number of original production videotapes (from case study): 5,347*
- *Cost of new tape stock: $100 per tape*
- *Cost of copying to new videotape: $400 per tape*
- *Total cost of migration: 5,347 x ($100 + $400) = $2,673,500*[14]

Again, there are longer-term costs to consider beyond those associated with the initial creation of a digital motion picture.

With an understanding of the new storage cost realities of digital motion picture data, the questions that must be asked now include: What materials should be stored for commercial exploitation on some new distribution technology or platform as yet unseen? What bonus materials will be needed? What about a potential "director's cut" or newly edited version? Is it sufficient to protect only the finished master? Should "mild" or "mathematically lossless" data compression be considered to reduce digital storage requirements by one-half or more? Is the image quality of the archived master sufficient for future display technologies? The new economics of digital motion pictures require a careful look at the complete asset picture.

Annual Storage Costs of Motion Picture Materials



[13] *The perceived longer "shelf life" of videotape as compared to data tape is attributed to the (generally) longer useful life of videotape formats, lack of dependence on computer hardware, operating systems and application software, and the use of error concealment techniques to counter increasing bit error rates over time.*

[14] *Tape stock and copy costs based on current high volume dubbing rates for HDCAM SR tapes.*

6.3 What This Means for the Motion Picture Industry

The ongoing costs of storage technology trend down while the costs of data management services, labor and power increase as a percentage of the total cost of ownership.

6.3.1 Economics of Archiving are Changing

TRADITIONAL ANALOG ARCHIVING COST structures have high initial delivery and archive accession costs, followed by low storage maintenance expenses until such time as the analog asset needs to be accessed and utilized, when there may be substantial additional costs. On the other hand, digital archiving of born-digital assets has lower initial delivery and accession costs and lower costs associated with access and utilization of the asset, but requires higher levels of investment to support the ongoing digital preservation process which may include digital migration. This increases the importance of organizational continuity and sustained funding. To date, storage (disks and tape) has been the biggest expense category, but as seen at the San Diego Supercomputer Center, EROS, and Swedish National Archives, as digital archives scale up and storage hardware prices decline over time, the ongoing costs of storage technology trend down while the costs of data management services, labor and power increase as a percentage of the total cost of ownership.

Traditional archiving is widely accepted as just another "cost of doing business" in Hollywood and elsewhere. For example, in every county in America (and other countries, no doubt) the tax assessor's office is responsible for maintaining accurate, up-to-date records of property ownership, property transfers, property definitions (surveyed plot lines) and taxes paid every step of the way. These archives are continuously growing. The assessor's archiving policy must be "save everything" because new records point to old records for authenticity and to describe changes. Records are never purged. This costs money – always has, always will. But the archiving of property ownership and tax payment records is a cost that society accepts as a necessary fact of life.

The economic model for traditional film archives incurs most of the expenses up front, in the form of one-time costs to acquire the collection, one-time construction costs for the building to hold the collection, and smaller ongoing expenditures for labor and power needed to catalog, copy and preserve the collection. There are also typically variable expenses to access/convert or restore film assets when they are retrieved, in order to make them useable/saleable. To reiterate, the longevity of film archives is primarily determined by media durability and proper use of media-specific conservation techniques, and secondarily by sustained funding and organizational continuity. Specialized skills required by traditional archivists are well-established organizational and managerial competence and a range of "white-glove" conservation techniques for specific media types.

Both traditional and digital archiving generally require investment "today" in the belief that some benefit will be realized in the indefinite "tomorrow." Historically, most archives have been operated as non-profit "public works" for religious or scholarly purposes, or the good of society as a whole. Many cinema archives around the world continue to operate as public archives. Digital conversion can expand potential access for future generations, conveying the cultural "patrimony" to more citizens. In Hollywood, private cinema archives have developed as valuable corporate assets that appreciate over time and can yield profitable commercial media products in the future. Digital archiving will enhance the potential for commercial exploitation of Hollywood's media assets and will play an increasingly central role in the business. But private digital cinema archives are not going to be cheap, they are not going to (immediately) eliminate the old costs of film archiving, and they require a new business model to sustain digital preservation activities.

As explained elsewhere in this report, longevity of the digital archive using current technology and procedures is primarily determined by digital migration or emulation rather than physical conservation of media objects. So digital archives will require recurring expenditures to support the regular procedures of data audit and exercise, data quality control, data migration and/or emulation required for long-term preservation of digital assets. These processes can be done as periodic batch jobs, which leads to peaks and valleys in the workload of the staff, bandwidth requirements for the system, operating expenses and capital investment, all depending on the periodic frequency of technology obsolescence-driven data migration. But the more modern approach at larger digital archives has been to automate data migration so it can occur as an ongoing background task in order to smooth out the workload and the budgetary expenses over time. This increases the importance of organizational continuity and sustained funding.

At the same time as the cost structures of archiving are changing with the transition from analog to digital, it is clear that access to archives (even analog/film) has become increasingly valuable over the past few decades. Digital archives are potentially more accessible than analog/film, which implies that digital archives will become potentially more valuable than analog archives. Digital archives offer the benefits, compared to analog archives, of faster search and asset retrieval, easier local and remote access via networks, lower cost of replication and distribution, and easier and faster format conversion, including the ability to extensively "slice and dice" old digital assets into new content that can be commercially exploited for new markets via new distribution channels.

6.3.2 Save Everything

CURRENT PRACTICE IN HOLLYWOOD IS to "save everything" on film in the film archives or warehouse storage facilities. This ensures future users will have maximum flexibility to pick and chose what they want, when they want it. As an archiving policy, "save everything" is fast, comprehensive and simple to understand. And it is safe, because no one has to take responsibility for deciding what not to save. This practice has been extended to include most every element not on film, which includes paper documentation, digital data tapes, videotapes, optical disks and hard drives.

At one studio, a senior technologist recognized the potential for future value of all assets, but said the decisive reason to save everything is that it is just too troublesome to sort out the desirable materials from the undesirable. The easiest thing to do is throw it all in the vault. And, in the view of this technologist, ideally the digital "vault" should be online magnetic disk storage because this will make the digital assets more accessible for both re-purposing and data mining, including new techniques to extract contextual and descriptive metadata automatically.

Historically, motion picture elements were and continue to be stored in several locations: on the lot in the studio archives, at independent film archives and storage warehouses, and in many cases, at film labs and postproduction houses at no cost as a courtesy to their studio clients. This model has well served the studios when the elements are completely film-based, but the situation changes dramatically when digital elements are involved.

6.3.3 Don't Save Everything

THE "SAVE EVERYTHING" POLICY, whether motivated by concerns about future sales opportunities or adopted because it is the path of least resistance, must confront the practical reality described by several studio executives, which is that studios are producing so many bits and pieces of digital content that they cannot afford to save everything forever, but instead must learn what to discard. This is especially true for feature-length motion pictures because, as we have seen, long-term digital archiving incurs ongoing preservation costs that are significantly higher than archiving film – on an annual basis, $8.83 per running minute to archive a film master versus $104.28[15] per running minute to archive a 4K digital master.

It is useful to look again at what is being done in television. ESPN – arguably the largest cablecaster of sporting events – is awash each weekend in data tapes. ESPN covers professional, college and local sports in most categories, and by Monday morning, unless the weekend's captured material is cleaned out, there may be no physical room to ingest the next week's national and international feeds. ESPN's guiding rule is to save only assets that cannot be reasonably reconstructed. In their case, the sheer volume of data forces decisions on what to save, and the same sheer volume of data makes it impossible to save everything. The decision as to what to discard and what to save has been described as "triage on the fly." In the motion picture industry, the practice of deleting digital non-"circle takes" on set has been reported, but until the cost of digital archiving is considered, there has been no compelling reason to do this or some other culling process on a regular basis.

6.3.4 Who Decides, How, and When?

THE VALUE PROPOSITION OF "SAVE everything" is changing as the media business goes digital because for the first time in history it is becoming feasible to create digital distribution libraries and digital archives capable of exploiting

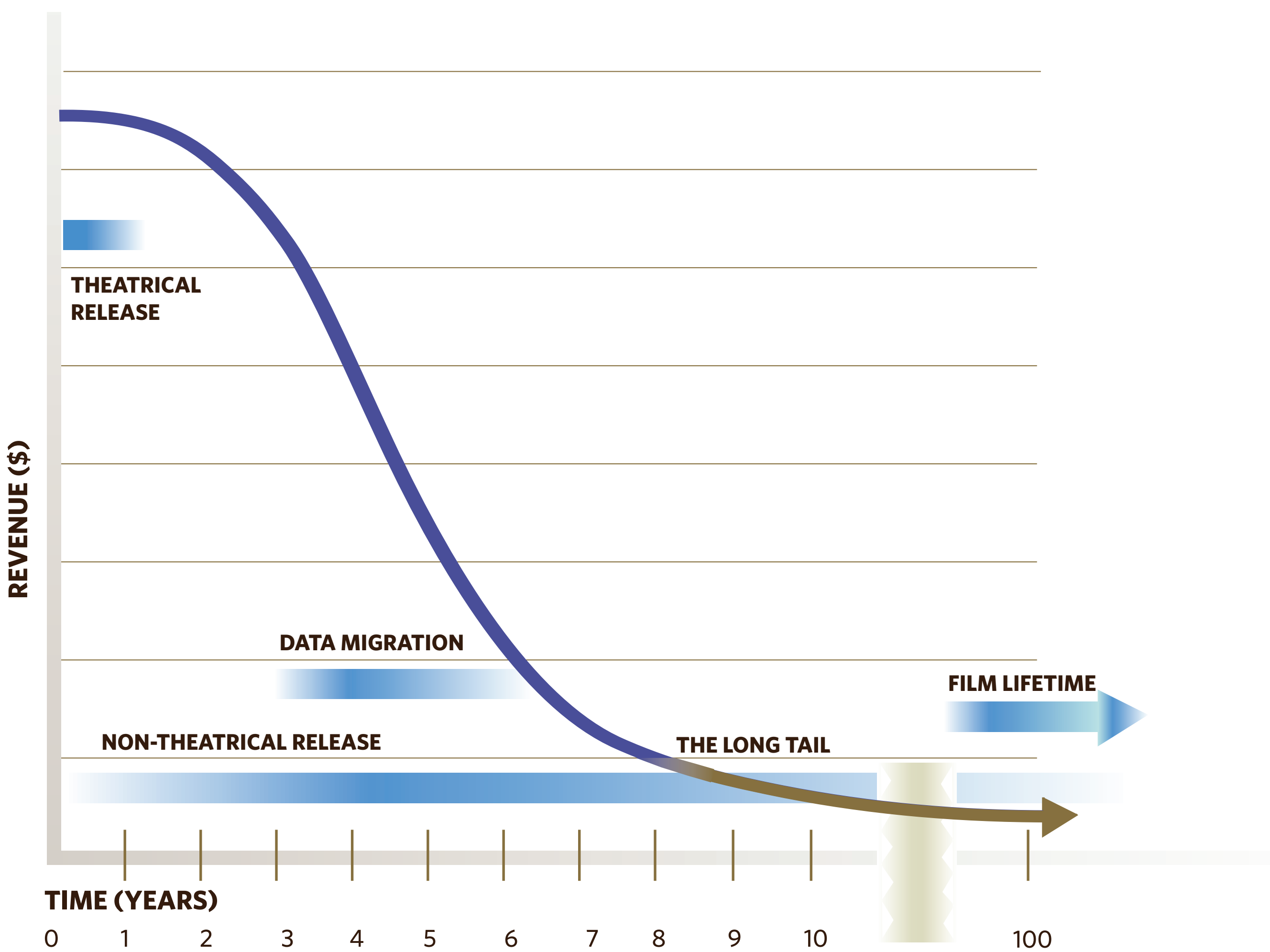
[15] *Assumes 3 copies of digital data and a YCM separation/negative/interpositive set for film.*

6.3 What This Means for the Motion Picture Industry *continued*

the marketing theory of "The Long Tail." While the economic aspects of this theory present an interesting marketing concept, its application and attendant cost of saving everything for future and unknown uses are not necessarily practical for today's motion picture content owners. This theory came up in several interviews for this report, and therefore deserves some discussion.

The Long Tail theory was first articulated in 2004 by Chris Anderson of *Wired* magazine to explain changing sales trends for digital media over the Internet. Anderson argued that products that are in low demand or have low sales volume can collectively make up a market share that rivals or exceeds the relatively few current bestsellers and blockbusters, if the store or distribution channel is large enough. The Long Tail acknowledges that sales volumes for a unit of digital media are highest at the time of initial release due both to novelty and promotional activities typical of the modern "hits" media business. After the initial period, sales derived from a digital media asset will decline, just as for traditional packaged media. But the Long Tail theory asserts that the low cost of delivery (but not of storage and access) of digital media over the Internet allows customers to continue buying a given title for a longer period with little incremental expense to the owner of the media asset because no physical duplication or delivery is required to complete a profitable transaction. The Long Tail market extends past the initial peak of revenue and subsequent decline, to a new, third phase of commercialization when the value of digital assets is extended because of rarity or nostalgia or unexpected opportunities to re-purpose assets for new distribution channels. Long Tail theorists argue that digital media and digital distribution will lead global media companies to expand their business

Content Lifetime



by first selling a few products in large volumes to mass audiences, and then selling small volumes of many products to thousands of niche markets, reaching new customers who are willing to dig deeply into a studio's media catalog.

The key supply-side factor that determines whether a sales distribution has a Long Tail is, of course, the cost of inventory storage and distribution. Where inventory storage and distribution costs are insignificant, it becomes economically viable to sell relatively unpopular products; however, when storage and distribution costs are high, only the most popular products can be sold.

Netflix is often referenced as a Long Tail business success story. A traditional movie rental store has limited shelf space, for which it pays facilities overhead; to maximize its profits, it must stock only the most popular movies to ensure that no shelf space is wasted. But Netflix stocks its movies in centralized warehouses, so its storage costs per unit are far lower and its distribution costs are the same for a popular or unpopular movie. Netflix is therefore able to build a viable business stocking a far wider range of movies than a traditional movie rental store. Those economics of storage and distribution then enable the advantageous use of the Long Tail. Reportedly, Netflix finds that in aggregate over time, "unpopular" movies are rented more than popular ones. And the people watching these sorts of movies are typically prepared to accept a delay of a few days between requesting a title and watching it.

The Long Tail theory of digital media marketing is consistent with new "data mining" practices emerging in other fields such as oil/gas exploration, medical imaging, Earth observation science, and even commercial credit card services. Data mining essentially involves analysis of old data (from the digital archives) using new algorithms running on more powerful computers than were available when the data was originally generated in order to extract new value from the old data. Data mining is being successfully used to find new oil and gas deposits, perform epidemiological studies of medical trends, track climate changes over time, and enable credit card default analysis by region, by time, or by individual. For the entertainment industry, the comparable techniques might include re-editing old content, or extracting certain types of scenes or dialog, or re-sizing and re-compressing for new distribution channels that did not exist when the media assets were originally created. For example, one studio executive described how his company had generated several million dollars in new revenue by extracting certain phrases from old television shows and selling these sound bites to consumers as downloadable ringtones for cellular telephones. The Long Tail implies that studios can maximize their profits by adopting a digital archiving strategy of "save everything forever," since everything will have value to someone someday. However, some feel that although the Long Tail is very long, it is also very thin, and therefore the cost, today, of saving everything, may preclude implementing this strategy.

Where to save – islands of archiving

The number of groups in a major studio capable of making their own digital media is growing. The creative talent and production facilities are becoming decentralized. Some or all of the content made by these teams may need to be preserved in digital archives. At the same time, the potential distribution channels served by the studio are proliferating, driving creation of more digital formats. One studio reported that without a "grand plan" for archiving, they are seeing spontaneous emergence of independent "islands" of digital archives in different business units and functional groups. These islands have been developed as stand-alone solutions, and often hold redundant inventory without consistent naming conventions (metadata). There is frequently no inter-operability, even for interchange within the enterprise itself.

According to some, digital archiving operations can add value to assets in terms of potential content repurposing far more quickly than analog archives. The question of future accessibility is the basic question behind all those decisions. Right now, access over extended periods (100 years or longer) is not guaranteed in the world of cinema except for YCM separation masters on film.

7 Industry Needs and Opportunities

A digital archival system should be at least as capable as the film preservation system it replaces.

THIS REPORT, THUS FAR, HAS SIMPLY PRESENTED A LARGE COLLECTION of relevant facts and informed opinions about the creation and preservation of film-based, hybrid and "born-digital" motion picture elements, digital storage technologies, and digital data handling practices occurring both within the motion picture industry and in other industries with similarly large and long-term data storage and preservation needs. As stated earlier, a primary goal of this work is to provide sufficient information and grounding so that the motion picture industry's needs with regard to the transition to a digital infrastructure are clearly defined, ultimately enabling sensible selection and implementation of appropriate technologies and practices that guarantee the long-term safety of and access to important corporate and cultural assets.

Armed with the perspective of a solid information base, the following sections state, in our view, the most fundamental industry needs regarding the archiving of and access to digital motion picture materials. Some of these requirements may appear obvious, but they simply articulate the needs met quite successfully by film technology for the entire history of our industry. Although there may not be an equivalent or improved replacement technology available today, the Science and Technology Council sees no reason to abandon these needs.

Herein lies the opportunity for the motion picture industry to break from the practice of accepting technologies and methods developed by other industries and business interests without regard to the most fundamental needs of motion picture production and preservation. We have the ability to define and communicate our particular needs, leverage the overlapping needs of other industries, and then, perhaps, to have a choice of solutions that solve as many problems as the new digital technologies seem to create.

To that end, what follows are the most basic needs of an archive for digital motion picture materials stated without regard for today's available solutions:

1. A digital archival system that meets or exceeds the performance characteristics of the traditional film archive

As a starting point, a digital archival system should be at least as capable as the film preservation system it replaces in the following respects:

Guaranteed access for at least 100 years: The single characteristic of a digital archival system universally requested by every studio and film archive we spoke with was that access to the content stored in the archive should be guaranteed for at least 100 years. Simply put, that is what they have with film, and that is what they want when and if film is no longer available.

Immunity from extended periods of neglect and financial hardship: Another characteristic of the film archive is that its contents remain accessible even if it were subject to, in the words of one studio executive, periods of benign neglect. That is, reductions in staffing or funding would not cause the content to disappear or become inaccessible. Although film may slowly degrade if funding shortfalls were to result in suboptimal environmental conditions, restoration would almost always remain an option.

Ability to create duplicate masters to fulfill future (and unknown) distribution needs: Film archival masters, when properly created and stored, have been of more than sufficient quality to generate any distribution master, whether for 4K Digital Cinema or handheld portable media players. Any replacement archival technology must be able to do the same, for both existing and unrealized distribution channels.

Picture and sound quality which meets or exceeds that of original camera negative and production sound recordings: There is no question that properly created film archival masters support the generation of distribution masters with little or no quality loss. The current use of 2K and HDTV mastering pipelines, and 2K digital cameras for theatrical motion pictures, as well as insufficient attention to image quality during the mastering process, together are generating archival elements that are of noticeably lower quality than films created more than 40 years ago.

The deployment of 4K Digital Cinema projection systems and the introduction of 4K consumer displays[16] are clear indications that future display systems will make greater picture quality demands on the archival master. At a minimum, image quality metrics regarding spatial resolution, color gamut and dynamic range as defined by SMPTE Digital Cinema standards should be the baseline quality standards, as well as the corresponding standards for audio.

No dependence on shifting technology platforms: Film stocks have changed over the years, generally with increasing quality and stability characteristics (with one or two notable exceptions). This technology evolution has not compromised the accessibility of the film archival master, and therefore a replacement archival technology should not subject the archival master to such a risk.[17]

2. *Standardized nomenclature*

The multi-studio case studies undertaken as part of this report uncovered a problem that has taken more than 100 years to develop: each studio has a different naming and identification system for the physical and digital objects they create in the manufacture of theatrical motion pictures. These differences developed for perfectly logical reasons: each studio's inventory management system developed organically, along with its internal business systems, so there is no uniformity across studios. Unfortunately, it is impossible to effectively leverage any digital solution with the inefficiencies this situation creates. Our attempt in the case studies to simply and accurately quantify the amount of film and digital materials generated during motion picture production was hampered by the wide variation of inventory management practices. Further refinement of the industry's needs in this area will be that much more difficult without uniform naming practices.

Individual studios will also benefit from such standardization efforts. As part of a recent internal review, one studio identified nine different ways its various business units referenced a single 60-year-old property [Solomon]. Rationalizing object names not only improves access capability, it also enables strategies that reduce the number of duplicated items taking up valuable digital storage space. One studio executive claimed that "de-duplication" of his studio's libraries reduced the overall inventory by 30 to 50%.

[16] Sharp Electronics 4K LCD shown at CEATEC 2006 and CES 2007.

[17] The discontinuation of older print stocks has, however, required compensating color timing and/or correction of older films, since the color characteristics of new print stocks have changed.

8 Consensus

There is no digital archival master format or process with longevity characteristics equivalent to that of film.

ALTHOUGH THE PRIMARY INTENT OF THIS DOCUMENT IS TO DEFINE the problem of digital archiving and shine a light on the important related issues for our industry, there is general agreement among the people interviewed with regard to what actions to take going forward. The consensus view answers two basic questions:

- *What should be done right now?*
- *What should be done over the long term?*

Given the conclusion that there is no digital archival master format or process with longevity characteristics equivalent to that of film, the emphasis is on protecting today's assets while work continues on developing suitable long-term solutions.

8.1 To Start

1. Create film separation masters

As stated earlier, there is virtually unanimous agreement within the industry that film separation masters, whether created using three-strip or successive exposure techniques, are a safe and affordable archival master. Some may argue that pure born-digital motion pictures (digitally shot or animated with computerized tools) are degraded when film grain, no matter how fine, is added to the images; but the film masters are still well above the historical notion of "highest quality," and are thus far more than capable of delivering the quality necessary and expected for all re-purposed distribution needs.

2. Enable the enterprise to develop a rational digital preservation strategy

While there are small groups within each studio that understand the issues presented in this paper, their influence is not exerted until well after the important decisions regarding digital asset creation are made. This is too late to ask the questions that must be asked when considering the huge number of choices presented by digital production and postproduction.

Although every studio ultimately manufactures its products (motion pictures) to identical delivery specifications (35mm film and Digital Cinema Packages), each organization has distinct internal structures and processes developed over its unique history, influenced by a wide and varying range of non-motion-picture-related business needs. The net effect is that each organization must consider the entirety of its own business goals in developing a long-term strategy for archiving and accessing digital materials.

That being said, there are some common elements to be considered:

Accept and understand that preserving digital motion picture materials is fundamentally different than preserving film, and as such, every assumption and practice in motion picture production (including corporate structure) must be looked at from this new perspective. The "save everything" practice used with film is cost-prohibitive with current digital storage technologies, given the huge quantity of data and ongoing preservation expense.

Identify the stakeholders in the enterprise and define their interests, roles and responsibilities with respect to the creation, preservation and access of digital motion picture assets. We have heard from several studios that the growth of alternate digital distribution channels – television, Internet, mobile, and so on – has fractured yesterday's relatively simple asset and inventory management process and corporate

Accept and understand that preserving digital motion picture materials is fundamentally different than preserving film.

structure, and expectations for future fulfillment do not always match up with the realities of current practice. Digital preservation and access must be defined for the enterprise as a whole; e.g.:

Σ

- *What is a long-term asset?*
- *What is considered perishable?*
- *What elements justify the cost of digital preservation?*
- *What are the methodologies under which these decisions are made?*

Enable collaboration among the stakeholders to develop a strategy for digital preservation. The following questions raised by using digital technologies cannot be answered by any single department or division:

- *What is the value of the content?*
- *Who determines the value of the content?*
- *What content will be archived?*
- *Who determines what content will be archived?*
- *How will the content be archived?*
- *Who determines how the content will be archived?*

3. This is an industry problem, and to solve it, the industry must work together

The founders of the motion picture industry knew early on that their business was about selling movies. The mechanisms needed to create their product were simply means to that end, and they generally went to great lengths to reduce the costs of production. Collaboration on solving technical problems dates back to 1916 with the creation of the Society of Motion Picture Engineers (now the Society of Motion Picture and Television Engineers), and 13 years later, at this Academy through its Producers-Technicians Joint Committee. The standardization of 35mm motion picture film, camera aperture and theatrical sound equalization, among other things, was the result of collaborative efforts by companies that otherwise competed with each other.

The issues of archiving and accessing digital materials are of the same nature: no studio or filmmaker will make any money from the technological solutions that enable the long-term preservation of and access to motion picture content. However, unless and until the issue of long-term access is solved, future revenue streams – and possibly the art form itself – are highly endangered. The motion picture industry must not necessarily accept solutions that fall short of what has been used successfully for 100 years. The technological solutions are likely to come from outside the industry, but it is vitally important that the industry speak with a common voice on its unique needs. There is also an opportunity to collaborate with other industries that share common aspects of long-term digital preservation and access, particularly with respect to influencing storage vendors and system solution providers to develop products that more closely match our requirements.

4. For the short term, actively protect important digital assets

There is no denying the reality that over a billion dollars[18] has been spent generating digital motion picture assets. Creating YCM separation masters on black-and-white polyester film stock protects the final theatrical product, but there may be tremendous value remaining in the multiple digital masters generated from a motion picture, and quite possibly in the original digital camera files and tapes. There may be both business

[18] *Based on the number of digital masters and digitally captured movies to date.*

opportunities and cultural obligations to maintain access to at least some of these digital objects while the bigger picture is assessed and long-term strategies are developed.

Design for evolution Unless and until there exists the digital equivalent of film, i.e., a "store and ignore" preservation medium, organizations will have to manage the realities that hardware and media will become obsolete, software applications will be upgraded, economic conditions will change, and personnel will come and go.

While we do not at this time accept data migration as a fait accompli for the industry's digital preservation future, there is no getting around it once one commits to creating valuable digital assets using commercially available information technology storage products. Whether this commitment happens proactively or by default, exclusively using today's digital technologies makes migration a necessary, if temporary, strategy to consider.

Design for low risk of technical obsolescence

It is worth repeating that modern technology products have finite usable lifetimes, in many cases as short as two years. However, there are some things that can be done to mitigate the impact of technology churn:

Standards: If standards exist, use them. File formats, image and sound encoding specifications, and metadata are important work items for the international standards development community, and many of them can be applied to today's needs. In the absence of relevant standards, the industry should organize to create the standards it needs for digital archiving, much as SMPTE is documenting Digital Cinema distribution specifications.

Open-source software: There is a large body of open-source software being developed specifically for large data storage problems. Base software technologies such as the Storage Resource Broker, the next-generation iRODs distributed storage system, and the LOCKSS (Lots Of Copies Keep Stuff Safe™) program offer interesting opportunities for minimizing the impact of changing vendor strategies and business goals, and proprietary single-vendor products.

Lower the risk from threats: economic, technical, human Maintaining digital data for the long term using today's technology demands perpetual funding. Most organizations want to minimize the total operating costs of a digital storage system. We want to re-emphasize that the **total cost of ownership** should be determined not only by counting the media costs or the initial purchase price of the hardware and software, but also the recurring costs. Furthermore, there are different cost factors to consider when building a digital storage system and/or outsourcing digital storage services:

In-house Systems: When building a digital storage system, total cost of ownership includes:

- *initial hardware, operating system and application software costs*
- *software and hardware maintenance contracts*
- *replacement costs of hardware, operating systems and software applications*
- *external network access costs for distributed systems*
- *initial and replacement media costs*
- *personnel costs, including ongoing training*
- *electrical power and cooling costs*
- *facilities and real estate costs, taxes and insurance*
- *increase in costs as digital asset collection grows*
- *data ingest and access costs*

Appropriately sizing the storage system will also affect total cost of ownership. Larger systems tend to reduce the cost per bit, although they require larger initial investments to construct.

Outsourced Systems: When outsourcing a digital storage system, total cost of ownership includes:

- *"rental" cost for storage – this can vary widely, depending on the service provider's level of service with respect to the threats described earlier*
- *data ingest and access costs*
- *risk mitigation of service provider failure*

For both scenarios, another factor that will impact total cost of ownership is data duplication. Many organizations we spoke with have the common problem of unintended multiple data copies. That is, there are many redundant copies of motion picture elements, and in the absence of sensible information lifecycle management policies, every bit of data is saved. This easily doubles or triples the amount of data managed (or not managed, as the case may be) by an organization and this drives storage costs up. "De-duplication" is the practice of eliminating unnecessary redundant files, which in turn reduces the amount of data to be stored and the associated cost. That requires another empowered decision.

8.2 Long Term Initiatives

The motion picture industry should organize itself to speak with a common voice on matters of digital archival technology and solutions.

THOSE INTERVIEWED FOR THIS REPORT AGREED THAT ACTIONS can be taken to produce better solutions for long-term digital preservation and access than we have today. The Science and Technology Council's goal is to move these notions from just being written about to being acted upon.

1. Collaborations

We stated earlier that the motion picture industry should organize itself to speak with a common voice on matters of digital archival technology and solutions, thus enabling it to effectively join forces with other industries that have similar needs with respect to digital preservation and access. This is not a problem that can be solved without great leverage – there needs to be a consortium of end-users, i.e., customers, who can economically scale their demands to make it attractive for vendors to agree to open standards. We point to the audiocassette, CD, 35mm film and for a while, the DVD, as examples of this. Many, many companies were successful in manufacturing, distributing and selling these standardized formats. They did not need proprietary "secret sauce" to be successful in creating and servicing their markets.

There are a number of examples of cross-industry collaboration, the most notable of which, for our purposes, is the National Digital Information Infrastructure Preservation Program (NDIIPP), created by the Library of Congress (discussed earlier in this report). The Library acknowledged that the scope of this problem is simply too large for any organization, even the United States government, to tackle on its own. The NDIIPP program currently funds more than 16 external partners working on digital preservation research and collections, and the Library is engaged in numerous digital preservation-related partnerships with notable institutions including the National Archives and Records Administration, the National Science Foundation, and the Digital Library Federation, as well as digital preservation initiatives abroad.

In August 2007 the Academy and the Library of Congress announced the Academy's participation in NDIIPP's Preserving Creative America project, a joint effort to address the issues of digital preservation as they relate to theatrical motion pictures. Participation in this program will bring increased visibility to the motion picture industry's needs, and it is hoped that we will also discover new ideas that will lead to better solutions for the industry. Topic areas of this joint effort include:

- *a report on the Digital Dilemma from the perspective of the independent filmmaker and smaller, public film archives*
- *development of a digital preservation case study system to investigate various digital motion picture archival strategies*
- *development of requirements and specifications for digital file formats that support long-term digital preservation*
- *education and research activities related to digital motion picture preservation*

This is just one example of the opportunities available to leverage the efforts of several organizations and industries toward a common goal.

2. Standards Development

While we have heard conflicting advice from other industries on the value of standards with respect to digital preservation, it is clear that the motion picture industry has benefited, and indeed would not exist, without worldwide standards for the interchange of motion picture content. International standards have the added

CONSENSUS

8.2 Long Term Initiatives *continued*

benefit of being automatically reviewed every five years, which provides a built-in mechanism for dealing with the constant churn of new technology.

We believe that standards are most likely to be successfully implemented and adopted when the user community of those standards takes an active and leadership role in their development. The Society of Motion Picture and Television Engineers (SMPTE) and the International Organization for Standardization (ISO) are the two accredited standards bodies that publish most of the motion picture industry's standards in use today, and both organizations are actively developing standards for Digital Cinema distribution and exhibition. It is interesting and important to note that these standards are based on specifications written not by equipment manufacturers and technology providers, but by a consortium of one segment of the user community: the Hollywood studios via the Digital Cinema Initiatives consortium. Much input on the specifications was taken from another important user group – the exhibitors – as well as from the equipment manufacturers, but the process was driven by a committed and influential user group.

Similar effort must be applied to the ad hoc world of archiving and access of digital motion picture materials. Image file formats, their associated "wrappers," filenames, metadata, and metadata registries all are of limited usefulness unless there is industry-wide agreement on what they are and how they are to be used.

Compared to motion picture film, motion picture digital formats are still in their infancy. There is no universally accepted standard for all phases in the life cycle of digital motion pictures assets – production, postproduction, distribution and archiving. The DCI recommendations and subsequent SMPTE DC28 standards efforts are building consensus around Digital Cinema distribution formats. But the format for the so-called Digital Source Master (DSM), i.e., the digital equivalent of the cut negative, is not standardized, nor is there even agreement on what a DSM is. Digital camera acquisition image formats are also not standardized. Digital film scanner output formats are not standardized. Technical innovation and market forces together are still influencing the evolution of various digital formats for Digital Cinema that might one day have to be preserved in a digital archive.

Based on the experience to date in television, the definition of a Digital Cinema archive master digital format will require a detailed evaluation of alternative file formats, wrappers, image and sound encoding formats, metadata formats and metadata registries. The subject needs a focused effort to build consensus around one or several digital formats that can be sustained for archival purposes.

8.3 Finally...

This is an issue that requires top-down examination, enlightened decision-making, and intra- and inter-industry cooperation for the benefit of today's content creators and tomorrow's audiences.

IN THE CENTURY SINCE CINEMA WAS FIRST INVENTED, MANY different perforation schemes, emulsions and sound track formats evolved and can be found in film archives around the world. Yet today, more than 100 years after its introduction, 35mm film is the shining example of a standardized and sustainable format that is widely adopted, globally interoperable, stable and well understood.

The bottom line is that any system proposing to replace photochemical film technology must meet or exceed film's capabilities. While it is true that end-users benefit from new features and cost efficiencies that generally come with new products and technologies, the economic benefits of technological obsolescence accrue primarily to the hardware manufacturers and software system developers. In exploring this digital dilemma, it becomes clear that if we allow the historical phenomenon of technological obsolescence to repeat itself, we are tied either to continuously increasing costs – or worse – the failure to save important assets. This is an issue that requires top-down examination, enlightened decision-making and intra- and inter-industry cooperation for the benefit of today's content creators and tomorrow's audiences.

The Academy was founded to, among other things, represent the viewpoint of the actual creators of motion pictures and facilitate technological progress among the creative leadership of the motion picture industry. It is therefore the proper role of the Academy to spotlight this issue by bringing together the resources that produced this report, and to lead in the actions necessary to solve this dilemma. In addition to initiating the activities discussed earlier in this report, in the coming months the Academy will bring together studio decision-makers and technology resources, as well as other experts, to further define the requirements and issues in the archiving of and access to digital motion picture materials. These efforts are a start, but what is also needed is commitment by the primary stakeholders, and objective overview of the manufacturers and system designers, to produce cooperation, standards, and guaranteed long-term access to created content.

Only then will we have solved this Digital Dilemma for the benefit of all the players.

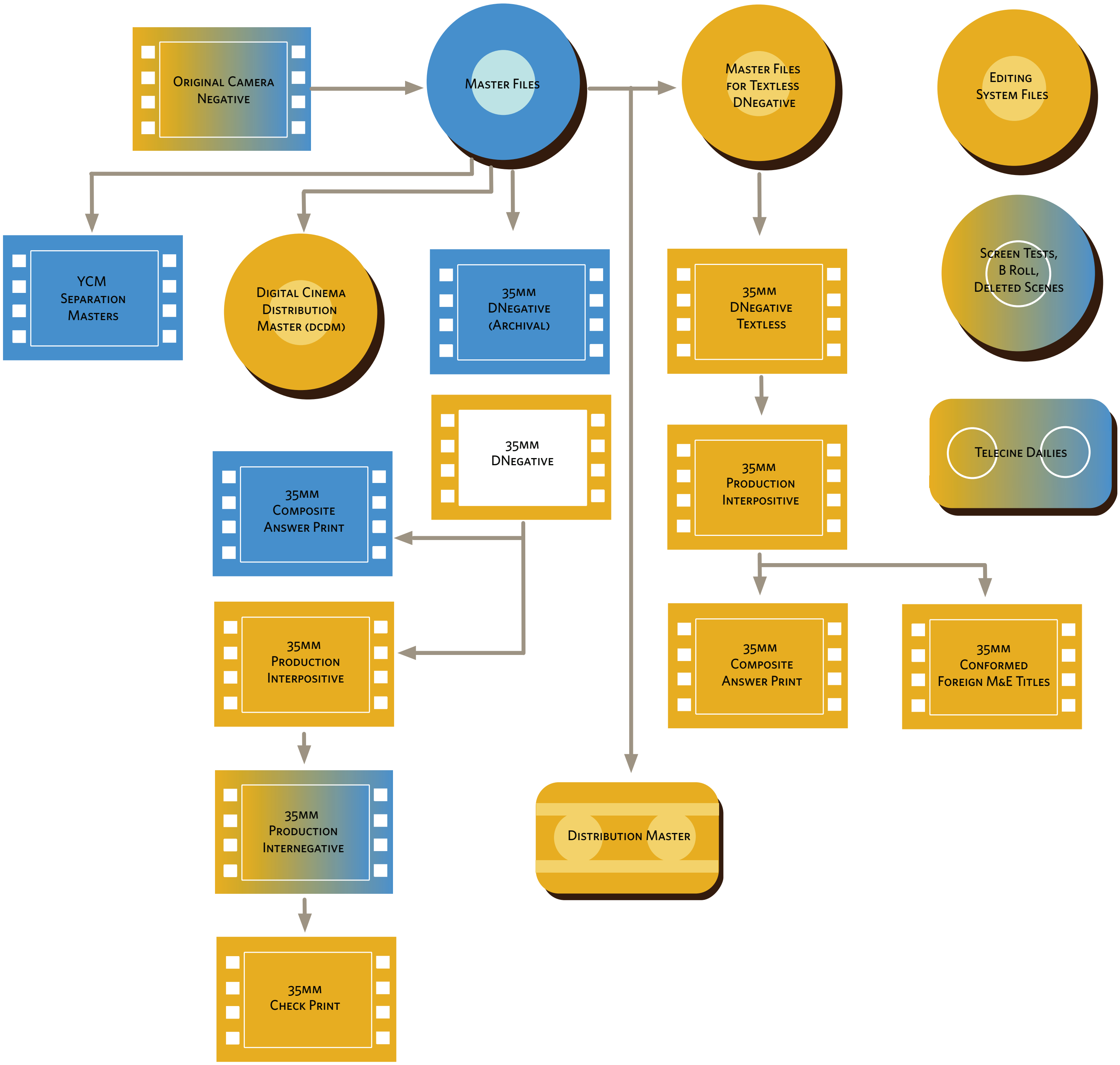
The place to start is here. The time to start is now.

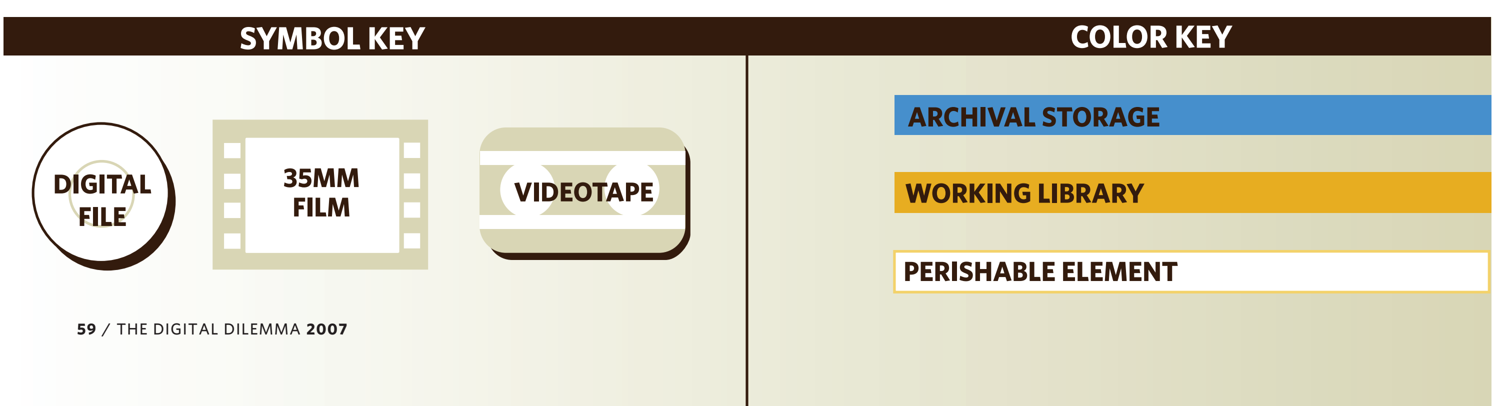
9 APPENDIX ▪ Case Study Data

The following sections contain summary information from the two case studies discussed in this report. The subject productions were captured either on film or HDCAM SR videotape and mastered at 1920 x 1080 pixel count with 10-bit precision per color component. This results in lower total byte counts than 2K/10 bit and 4K/16 bit encoding. However, the number and type of elements identified in the case studies are believed to be representative of those generated by both 2K and 4K productions.

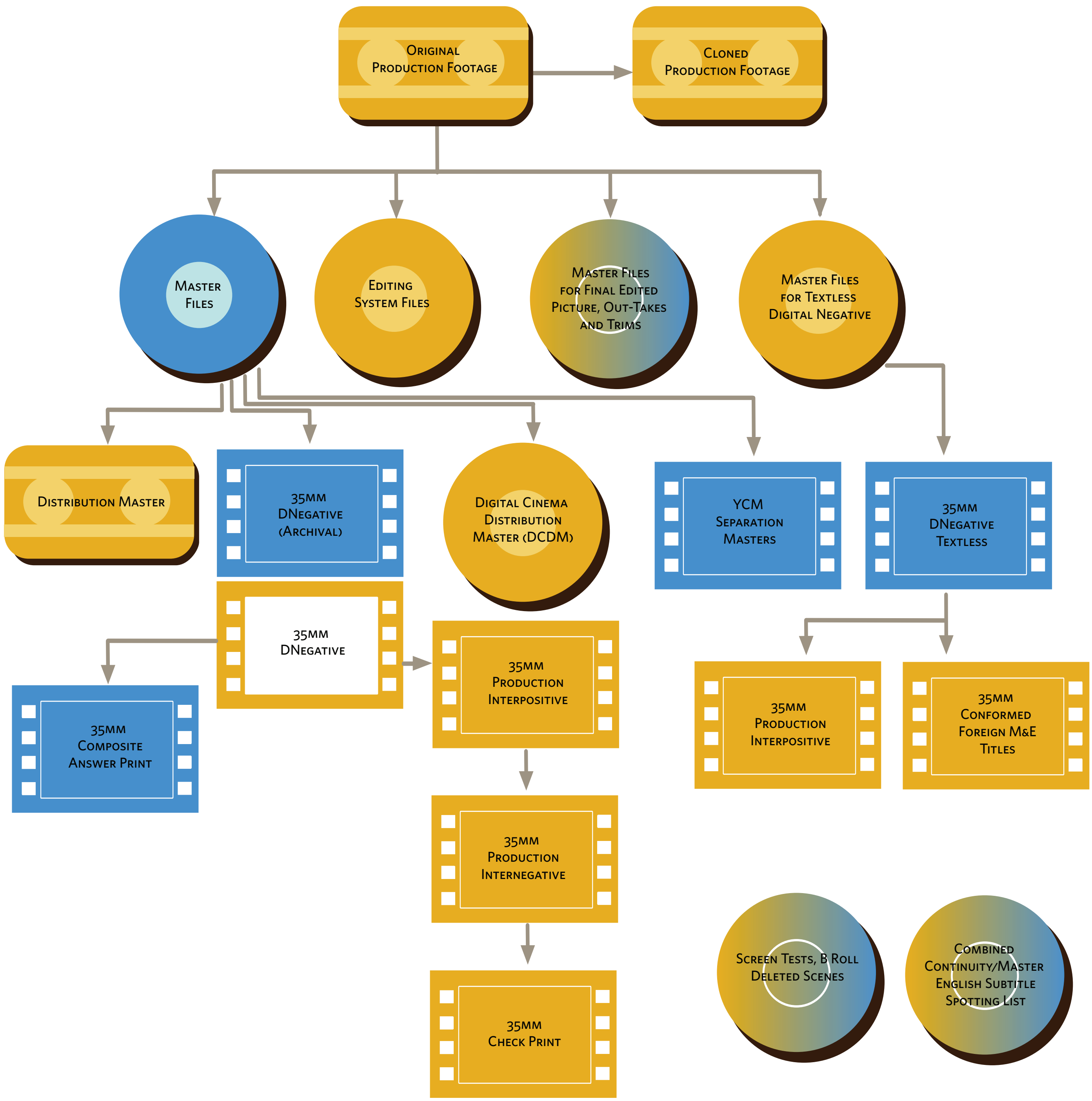
The Element Trees were derived from inventory data provided by the participating studios. The succeeding Case Study Data Tables contain the actual inventory data (with certain noted assumptions), using the identifying terms from the Element Trees.

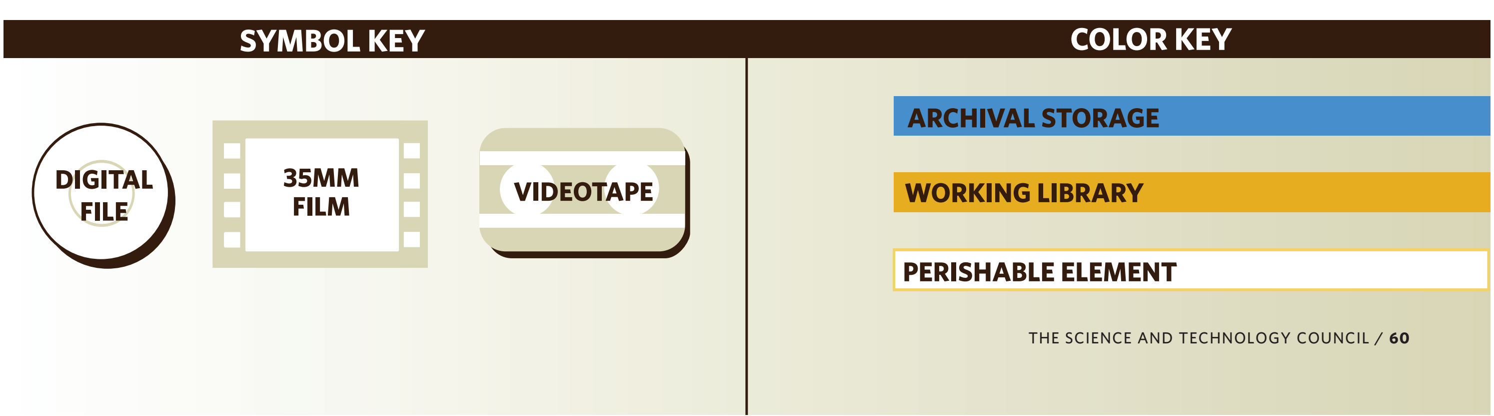
A.1 Generic Element Trees • Picture Element Hierarchy: Film Capture



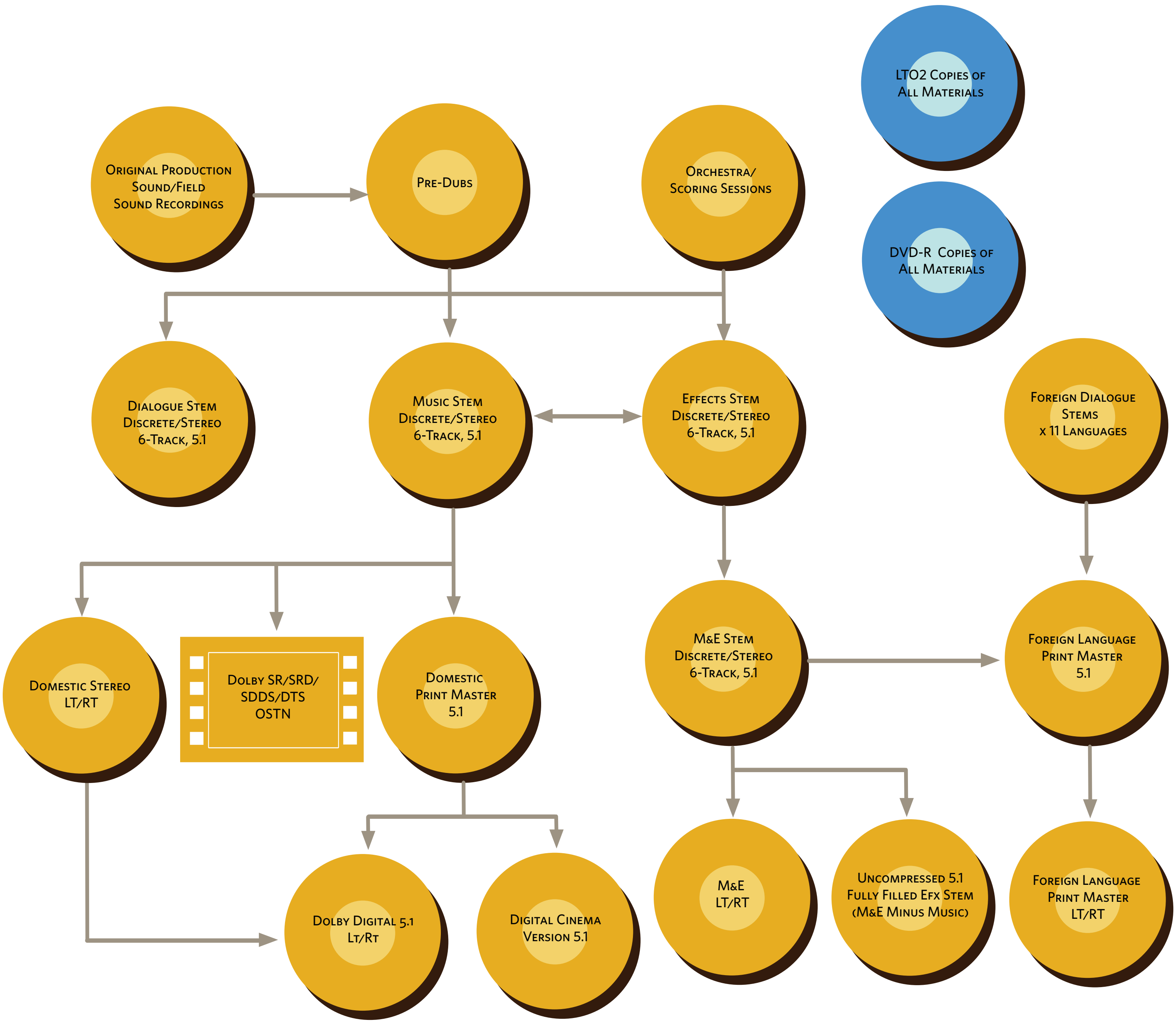


A.1 Generic Element Trees • Picture Element Hierarchy: Data Capture





A.1 Generic Element Trees • Sound Element Hierarchy: Film or Data Capture



A.2 Case Study Data Tables

Table A-1 lists the delivered picture elements identified in the film capture case study plus the medium of storage, the number of items per element and the estimated file size if digital storage is used.

STORAGE MEDIUM	ELEMENT	NUMBER OF ITEMS	FILE SIZE in TERABYTES
35mm Film	35mm Digitally Created Negatives 35mm Answer Print 35mm Production IP 35mm Production IN 35mm Check Print 35mm Textless DNegative 35mm Textless IP 35mm Textless Answer Print 35mm Foreign Language Main and Ends Negative 35mm YCM Separation Masters 35mm Original Camera Negative 35mm Trims and Outs	178 Cans or Cartons	NA
LTO2 Data Tape	1920x1080 Master Files 1920x1080 Master Files for Textless DNegative	15 LTO2 Data Tapes	3 TB
DVD-R Optical Disk	Editing System Files	1 Disk	.005 TB
HDCAM SR Videotape	Telecine Dailies	486 Tapes	173 TB[1]
D5 Videotape	Distribution Master	9 Tapes	.202 TB[1]

Table A-1 – Delivered Film Capture Picture Elements

[1] *Calculated.*

9 APPENDIX • Case Study Data *continued*

A.2 Case Study Data Tables *continued*

Table A-2 lists the delivered picture elements identified in the data capture case study plus the medium of storage, the average number of items per element and the estimated file size if stored on magnetic hard drives or data tapes.

STORAGE MEDIUM	ELEMENT	NUMBER OF ITEMS	FILE SIZE in TERABYTES
35mm Film	35mm Digitally Created Negatives 35mm Answer Print 35mm Production IP 35mm Production IN 35mm Check Print 35mm Textless DNegative 35mm Textless IP 35mm Textless Answer Print 35mm YCM Separation Masters 35mm Conformed Foreign Language Mains and Ends	129 Cans or Cartons	NA
Magnetic Hard Drives	1920x1080 Master Files 1920x1080 Master Files for Textless DNegative 1920x1080 Master Files for Outtakes and Trims Editing Files	42 Hard Drives	10.7 TB
HDCAM SR Videotape	Original Production Footage Cloned Production Footage/ Screen Tests/B-Neg	5,347 Tapes	3,257 TB[2]
D5 Videotape	Distribution Master	0 Tapes[1]	
DVCAM Videotape	Editing System Project Files	728 Tapes	24 TB[2]

Table A-2 – Delivered Data Capture Picture Elements

[1] *Items not yet delivered to archive.*
[2] *Calculated.*

A.2 Case Study Data Tables *continued*

Table A-3 lists the delivered sound elements for both film and data capture, the medium of storage, the average number of items per element, and the estimated file size if stored on magnetic hard disks or data tape. The delivered sound element types for both case studies were identical, although as expected, the quantities differed between productions because of variances in production and postproduction practices.

STORAGE MEDIUM	ELEMENT	NUMBER OF ITEMS (Data Capture)	NUMBER OF ITEMS (Film Capture)	FILE SIZE in TERABYTES
35mm Film	Optical Soundtrack Negative (OSTN)	34 Cans or Cartons	27 Cans or Cartons	NA
DVD-R Optical Disk	6-Track Dolby Digital Digital Cinema Version LT/RT Musefx 5.1 Efx Stems Foreign Language Dialogue Stems Foreign Language Print Master (5.1) Foreign Language Print Master (LT/RT)	71 DVD-R	371 DVD-R	.83 TB (Data Capture) .42 TB (Film Capture)
LTO2 Data Tape	Dolby LT/RT Print Master Dolby 5.1 Print Master 6-Track Dolby Digital	13 LTO2 Data Tapes	0[1]	.004 TB (Data Capture)
Magnetic Hard Drive	Music, Dialogue and Effects Stems Domestic LT/RT Domestic 5.1 Print Master 5.1 Musefx Orchestra/Scoring Sessions	43 Magnetic Hard Drives	3 Magnetic Hard Drives	2.6 TB (Data Capture) .73 TB (Film Capture)

Table A-3 – Delivered Sound Elements

[1] *Items not yet delivered to archive.*

9 APPENDIX • Case Study Data *continued*

A.2 Case Study Data Tables *continued*

Table A-4 lists the elements, the number of items and the storage category assignment for picture and sound elements from the film capture production. This breakdown may vary depending on the studio because of differing practices. Items in **bold** exist only in the film capture production and not in the data capture production.

As stated previously, "archival" is defined as storage of the master elements from which all down-stream distribution materials can be created over a 100-year timeframe, and "working library" storage is a broad term for elements that are generally kept on hand for distribution purposes.

STORAGE CATGEGORY	PICTURE ELEMENT	SOUND ELEMENT
Archival	35mm YCM Separation Masters LTO – 1920x1080 Master Files for Digital Negative 35mm Digital Negative 35mm Composite Answer Print (from Digital Negative) 35mm Digital Negative Textless	LTO or DVD-R Copies of all Working Library Sound Materials
Mixed Archival/ Working Library	**35mm Original Camera Negative** **35mm Production Internegative** HDCAM SR – Screen Tests, B-Roll, Deleted Scenes **HDCAM SR – Dailies**	NA
Working Library	DVD-R – Editing System Files LTO – 1920x1080 Master Files for Textless Digital Negative D5 – Distribution Master 35mm Production Interpositive 35mm Check Print 35mm Production IP, Textless **35mm Answer Print, Textless** 35mm Foreign Language Mains and Ends **35mm Trims and Outs** DVD-R – Combined Continuity/English Subtitle Spotting List	Original Production Sound Pre-Dubs Orchestra/Scoring Sessions Dialogue Stems Effects Stems Music Stems Dolby Stereo LT/RT Dolby SR/SRD/SDDS/DTS OSTN Domestic Print Master Dolby Digital LT/RT Musefx Stem Discrete 6-track, 5.1 Musefx LT/RT 5.1 Fully Filled Efx Stem Foreign Language Dialogue Stems Foreign Language Print Master 5.1 Foreign Language Print Master LT/RT

Table A-4 – Storage Categories for Picture and Sound Elements from a Film Capture Production

A.2 Case Study Data Tables *continued*

Table A-5 lists the elements, the number of items and the storage category assignment for picture and sound elements from the data capture production. Again, this may vary depending on the studio because of varying practices. Items in **bold** exist only in the data capture production.

STORAGE CATGEGORY	PICTURE ELEMENT	SOUND ELEMENT
Archival	35mm YCM Separation Masters Hard Drives – 1920x1080 Master Files for Digital Negative 35mm Digital Negative 35mm Composite Answer Print (from Digital Negative) 35mm DNegative Textless	LTO or DVD-R Copies of all Working Library Sound Materials
Mixed Archival/Working Library	**Hard Drives – 1920x1080 Master Files for Final Edited Picture, Outtakes and Trims** HDCAM SR – Screen Tests, B-Roll, Deleted Scenes	NA
Working Library	**HDCAM SR - Original Production Footage** HDCAM SR- Cloned Production Footage DVD-R – Editing System Files LTO – Master Files for Textless Digital Negative HDCAM SR – Distribution Master 35mm Production Interpositive **35mm Production Internegative** 35mm Check Print 35mm Production IP, Textless 35mm Foreign Language Mains and Ends	Original Production Sound Pre-Dubs Orchestra/Scoring Sessions Dialogue Stems Effects Stems Music Stems Dolby Stereo LT/RT Dolby SR/SRD/SDDS/DTS OSTN Domestic Print Master Dolby Digital LT/RT Musefx Stem Discrete 6-track, 5.1 Musefx LT/RT 5.1 Fully Filled Efx Stem Foreign Language Dialogue Stems Foreign Language Print Master 5.1 Foreign Language Print Master LT/RT

Table A-5 – Storage Categories for Picture and Sound Elements from a Data Capture Production

APPENDIX

9 APPENDIX • Case Study Data *continued*

A.2 Case Study Data Tables *continued*

Baseline Storage Costs

As stated previously, the baseline storage costs used for this study are:

- $4.80 per physical item per year for archival storage
- $1.80 per physical item per year for working library
- $500 per terabyte per year for near-line data tape storage (single copy)
- $1,500 per terabyte per year for online magnetic hard drive storage (single copy)

Initial inspection and access costs are not included in the baseline film storage costs, nor are access or ingest costs included in the baseline digital storage costs because reliable information for the latter is not available. Nonetheless, these costs should be considered when considering the type and quantity of assets being stored.

A.2 Case Study Data Tables *continued*

Table A-6 lists the estimated annual cost of storing the delivered picture elements from the film capture production. This includes digital elements that are created during postproduction and stored on LTO2 data tape, HDCAM SR (or equivalent) videotape, and DVD-R optical disk. Items in **bold** are stored in archival conditions and are so noted.

Today, the practice in Hollywood is to store digital media as physical items in either archival or working library conditions. Given the special handling requirements of digital data, and the associated costs, the following table calculates the estimated cost of storing the digital elements separately as data, on data tape, in a fully managed environment consistent with the archival intent.

STORAGE MEDIUM	ELEMENT	ANNUAL STORAGE COST OF DELIVERED ITEMS	ANNUAL FULLY MANAGED STORAGE COST IF STORED ON DATA TAPE
35mm Film	**35mm Digital Negative** **35mm Answer Print** 35mm Production IP **35mm Production IN** 35mm Check Print **35mm Textless Digital Negative** 35mm Textless IP 35mm Textless Answer Print 35mm Foreign Mains and Ends Negative **35mm YCM Separation Masters** **35mm Original Camera Negative** 35mm Trims and Outs	**$1,506[1] (Archival)** $290 (Working Library)	NA
LTO2 Data Tape	**1920x1080 Master Files for Digital Negative** **1920x1080 Master Files for Textless Digital Negative**	**$72 (Archival)**	**$1,465 (Archival)**
DVD-R Optical Disk	Editing Files	$2 (Working Library)	$2 (Archival or Working Library)
HDCAM SR Videotape	**Telecine Dailies**	**$2,333 (Archival)**	**$86,498 (Archival)**
D5 Videotape	Distribution Master	$16 (Working Library)	$96 (Working Library)

Table A-6 – Estimated Annual Cost of Element Storage – Film Capture

[1] *Includes amortized cost of YCM separation master manufacture, which is $800 per year.*

9 APPENDIX • Case Study Data *continued*

A.2 Case Study Data Tables *continued*

Table A-7 lists the estimated annual cost of storing the delivered picture elements from the data capture production. This includes both born-digital elements and digital elements created during postproduction, and stored on magnetic hard drive or HDCAM SR (or equivalent) videotape. Items in **bold** are stored in archival conditions and are so noted.

The estimated cost of storing a single copy of born-digital materials on data tape is also calculated to represent the use of uncompressed digital data recorders now in use.

STORAGE MEDIUM	ELEMENT	ANNUAL STORAGE COST OF DELIVERED ITEMS	ANNUAL FULLY MANAGED STORAGE COST IF STORED ON DATA TAPE
35mm Film	**35mm Digital Negative** **35mm Answer Print** 35mm Production IP 35mm Production IN 35mm Check Print **35mm Textless Negative** 35mm Textless IP **35mm Textless Answer Print** **35mm YCM Separation Masters**	**$1,102[1] (Archival)** $124 (Working Library)	NA
Magnetic Hard Drives	**1920x1080 Master Files for Digital Negative** **Complete 1920x1080 Master Files for Textless** **Complete 1920x1080 Master Files, Outtakes and Trims**	**$64 (Archival)**	**$5,127 (Archival)**
HDCAM SR Videotape	Original Production Footage Cloned Production Footage/ Screen Tests/B-Neg	$1,170 (Working Library)	$1,629,128 (Working Library)
DVCAM	Editing System Project Files	$100 (Working Library)	$11,245 (Working Library)

Table A-7 - Estimated Annual Cost of Element Storage – Data Capture

[1] *Includes amortized cost of YCM separation master manufacture, which is $800 per year.*

A.2 Case Study Data Tables *continued*

As stated earlier, the trend in the audio domain, where all delivered elements are born digital, is to copy all master audio files to DVD-R and LTO3 and geographically separate these materials for protection. The master files remain in the working library on magnetic hard drives for instant access. This approach is believed to be the most comprehensive attempt to create an archival process around motion picture sound elements, provided that the process includes a data integrity check and migration plan that outlives economic and labor factors.

Table A-8 lists the estimated annual cost of storing the delivered sound elements from the case study productions. The estimated annual cost of storing the sound elements on magnetic hard drives is also included to reflect current practice at certain studios.

STORAGE MEDIUM	ELEMENT	ANNUAL STORAGE COST OF DELIVERED ITEMS	ANNUAL FULLY MANAGED STORAGE COST IF STORED ON DATA TAPE	ANNUAL FULLY MANAGED STORAGE COST IF STORED ON HARD DRIVE
35mm Film	OSTN	$61 (Data Capture: Working Library) $49 (Film Capture: Working Library	NA	NA
DVD-R Optical Disk	Multi-Channel Master Stems 5.1 Domestic Printmaster Domestic LT/RT 6-Track Dolby Digital Digital Cinema Version 5.1 Musefx LT/RT Musefx Production Sound Pre-Dubs 5.1 Efx Stem Foreign Dialogue Stems Foreign Language Print Master (5.1) Foreign Language Print Master (LT/RT) **Copies for Geographic Separation**	**$144 (Data Capture: Archival)** $668 (Film Capture: Working Library)	**$414 (Data Capture: Archival)** $212 (Film Capture: Working Library)	**$1,242 (Data Capture: Archival)** $635 (Film Capture: Working Library)
LTO2 Data Tape	Dolby LT/RT, 5.1, 6 Track Print Masters	$4 (Data Capture: Working Library)	$2 (Data Capture: Working Library)	$5 (Data Capture: Working Library)
Magnetic Hard Drives	Music, Dialogue and Effects Stems Domestic LT/RT, 5.1 Print Master Musefx Orchestra/ Scoring Sessions	$79 (Data Capture: Working Library) $5 (Film Capture: Working Library)	$1,222 (Data Capture: Working Library) $366 (Film Capture: Working Library)	$3,667 (Data Capture: Working Library) $1,099 (Film Capture: Working Library)

Table A-8 – Estimated Annual Cost of Sound Element Storage

10 References

R E F E R E N C E S

Anderson, Chris. "The Long Tail." Wired 12.10 Oct. 2004. CondeNet Inc. 17 Aug. 2007 http://www.wired.com/wired/archive/12.10/tail.html.

Baker, Mary, et al. "A Fresh Look at the Reliability of Long-Term Digital Storage." Proceedings of the 2006 Eurosys Conference, April 18-21, 2006: Leuven, Belgium. ACM Press, New York, NY, 221-234. 22 Jan. 2007 http://www.lockss.org/locksswiki/files/3/30/Eurosys2006.pdf.

DCinema Today. "Digital Installations." Sept. 2007. WHM Cinema Consulting, LLC. 21 Sept. 2007. http://www.dcinematoday.com/dcdb/DCinemas.aspx?app=1&cnty=49&dci=1.

Digital Imaging and Communications in Medicine. "Part 1: Introduction and Overview." National Electrical Manufacturers Association PS 3.1-2007: 1-21. Rosslyn, VA. 15 Aug. 2007 http://medical.nema.org/dicom/2007/.

Galloway, Stephen. "Studios Are Hunting the Next Big Property." The Hollywood Reporter 10 July 2007. Nielsen Business Media, Inc. 16 July 2007 http://www.hollywoodreporter.com/hr/content_display/news/e3if727c623f03c782b8ad564866c828796.

Global Storage Technologies Inc. "GST's Annual State of Tape Report for 2004." GST Research and Engineering Group 2004: 1-15. Lake Forest, CA. 7 Sept. 2007 http://www.gstinc.com/white/report-2004.pdf.

Herman, Robert. Telephone interview with Laurin Herr. 1 Feb. 2007. Oakdale, MN, USA.

Hogan, Hank. "Film Fades Out." Photonics Spectra Mar. 2006. Laurin Publishing Co., Inc. 1 Oct. 2007. http://www.photonics.com/content/spectra/2006/March/business/68479.aspx.

Holmes Jr., David R., Merrill A. Wondrow and Joel E. Gray. "Isn't it Time to Abandon Cine Film?" Catheterization Cardiovascular Diagnosis 20.1 1990: 1-4. PubMed Abstract. 22 Aug. 2007 http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed.

Hurwitz, Matt. "Crime Scenes and Compression Schemes: The File-Based Workflow for David Fincher's Zodiac." Videography 15 Mar. 2007. 19 July 2007 http://www.videography.com/articles/article_14970.shtml.

Karagosian, Michael. "Multichannel Film Today." MPKE Consulting LLC Sept. 2003. 4 Sept. 2007 http://www.mkpe.com/publications/digital_cinema/misc/multichannel.php.

Kienzle, Claudia. "Breaking News . . . Disc at 11: The Migration to Tapeless News Production." TV Technology 11 July 2007: 1 & 12.

Kienzle, Claudia. "Taking the Tapeless Archive Challenge." TV Technology 11 July 2007: 12.

Kirsner, Scott. "Digital Cinematography: The Big Pixel." The Hollywood Reporter 13 June 2006. Nielsen Business Media, Inc. 19 July 2007 http://www.hollywoodreporter.com/hr/search/article_display.jsp?vnu_content_id=1002688109.

Kushner, David. "One Giant Screwup for Mankind." Wired 15.01 Jan. 2007: 1-3. CondeNet Inc. 15 July 2007 http://www.wired.com/wired/archive/15.01/nasa.html?pg=1&topic=nasa&topic_set=.

Lorie, Raymond. "The UVC: A Method for Preserving Digital Documents: Proof of Concept." IBM/KB Long-Term Preservation Study Report Series Number 4. International Business Machines Corporation and Koninklijke Bibliotheek. Dec. 2002: 1-47. Amsterdam, Netherlands. 18 Aug. 2007 http://www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf.

Lyman, Peter and Hal R. Varian. "How Much Information? 2003." School of Information Management and Systems 30 Oct. 2003: 1-112. University of California, Berkeley. 3 Aug. 2007 http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf.

Mellor, Chris. "Super-tape Market Edges Toward Monopoly." Techworld.com. 30 Jun. 2005. International Data Group. 17 Sept. 2007 http://www.techworld.com/storage/news/index.cfm?newsid=3951.

Moore, Richard L. et al. "Disk and Tape Storage Cost Models." San Diego Supercomputer Center, University of California, San Diego, CA: La Jolla, 2007.

Motion Picture Association of America. "2006 U.S. Theatrical Market Statistics: Worldwide Market Research & Analysis." mpaa.org 2006: 1-24. 24 Aug. 2007 http://www.mpaa.org/2006-US-Theatrical-Market-Statistics-Report.pdf.

National Research Council of the National Academies. Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for a Long-Term Strategy. Ed. Robert F. Sproull and Jon Eisenberg. Washington, D.C.: The National Academies Press, 2005.

Overfelt, Maggie. "Saving the Cineplex." CNNMoney.com 19 July 2007. DCinema Today. 22 Aug. 2007 http://www.dcinematoday.com/dc/extURLs.aspx?ID=290.

Palm, Jonas. "The Digital Black Hole." Swedish National Archives 2006: 1-14. Stockholm, Sweden. 19 July 2007 http://www.tape-online.net/docs/Palm_Black_Hole.pdf.

Pinheiro, Eduardo, Wolf-Dietrich Weber and Luiz André Barroso. "Failure Trends in a Large Disk Drive Population." Paper presented at the 5th USENIX Conference on File and Storage Technologies – Paper, February 12-13, 2007: 1-13 San Jose, California. 8 Aug. 2007 http://209.85.163.132/papers/disk_failures.pdf.

Rosenthal, David S., et al. "Requirements for Digital Preservation Systems." D-Lib Magazine 11, (11) 2005. 19 June 2007 http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.

Scherzer, Barbara. "Cookson Puts Focus on Classics." Variety 15 Apr. 2007. Reed Business Information. 20 Aug. 2007 http://www.variety.com/index.asp?layout=print_story&articleid=VR1117963134&categoryid=2222.

Schroeder, Bianca and Garth A. Gibson. "Disk Failures in the Real World: What does an MTTF of 1,000,000 Hours Mean to You?" Paper presented at the 5th USENIX Conference on File and Storage Technologies – Paper, February 12-13, 2007: San Jose, California. 26 Mar. 2007 http://www.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html.

Science Applications International Corporation. "Architecture and Technology Program: Offline Archive Media Trade Study." United States Geological Survey. 2006: 1-34. USA. Aug. 2007 http://wgiss.ceos.org/archive/archive.pdf/FY06%20Media%20Trade%20Study.pdf.

Solomon, Madi. Presentation. Digital Asset Symposium. Association of Moving Image Archivists. Pickford Center for Motion Picture Study, Linwood Dunn Theater, Hollywood, CA, 11 May 2007.

Taylor, Jeff and Tom Regal. Presentation. Digital Asset Symposium. Association of Moving Image Archivists. Pickford Center for Motion Picture Study, Linwood Dunn Theater, Hollywood, CA, 11 May 2007.

10 References *continued*

United States. Library of Congress. Digital Preservation. What the Library is Doing. The Library of Congress. 29 Aug. 2007 http://www.digitalpreservation.gov/library.

United States. Library of Congress. National Audio-Visual Conservation Center – Culpeper, Virginia: Computer Server and Storage Infrastructure RFP LOC-NAVCC-06-01. Washington, D.C.: Library of Congress, 2006.

United States. Library of Congress. National Film Preservation Board. National Film Preservation Plan: An Implementation Strategy. 22 Jan. 2004. 22 Aug. 2007 http://www.loc.gov/film/implemen.html.

United States. Library of Congress. National Film Preservation Board. Television/Video Preservation Study: Volume 1: Report. Oct. 1997. 28 Aug. 2007 http://www.loc.gov/film/tvstudy.html.

United States. National Science Board. National Science Foundation Long-lived Data Collections Enabling Research and Education in the 21st Century NSB-05-40 2005: 1-92. 3 Aug. 2007 http://nsf.gov/pubs/2005/nsb0540/start.html.

Van Bogart, John W. C. "Magnetic Tape Storage and Handling: A Guide for Libraries and Archives." Washington, DC: Commission on Preservation and Access; St. Paul, MN: National Media Laboratory, June 1995: 1-48. 17 Aug. 2007 http://www.imation.com/government/nml/pdfs/AP_NMLdoc_magtape_S_H.pdf.

The X Lab. "Optical Media Longevity." 2007. 18 Sept. 2007 http://www.thexlab.com/faqs/opticalmedialongevity.html.

Bibliography

Advanced Distributed Learning. 2007. Office of the Under Secretary of Defense for Personnel and Readiness. 28 Aug. 2007 http://www.adlnet.gov.

Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Information System (OAIS)." CCSDS 650.-B-1. Blue Book, Issue 1. Jan. 2002: 1-148. Washington, D.C. 8 Aug. 2007 http://public.ccsds.org/publications/archive/650x0b1.pdf.

Film Restoration & Conservation Strategies. 2004. European Film Heritage on the Threshold Of The Digital Era, The First Project's Final Report, Conclusions – Guidelines – Recommendations. Brussels: Gabrielle Claes, Royal Film Archive.

i Rule Oriented Data Systems. Aug. 2007. San Diego Supercomputer Center. 27 July 2007 http://irods.sdsc.edu/index.php/Main_Page.

Linear Tape-Open Technology. Ultrium Roadmap. 2007. Hewlett-Packard, International Business Machines Corporation, and Quantum. 19 Sept. 2007 http://www.lto-technology.com/technology/uroad.php?section=0&subsec=uroad.

Lots of Copies Keep Stuff Safe. Sept. 2007. Stanford University. 27 July 2007 http://www.lockss.org/lockss/Home.

Storage Resource Broker. June 2007. San Diego Supercomputer Center. 27 July 2007 http://www.sdsc.edu/srb/index.php/Main_Page.

11 Acknowledgments

This report would not have been possible without the support of the members of the Science and Technology Council's Digital Motion Picture Archival Project Committee, the consultants who performed the research and interviews, and the many experts from the motion picture industry and other sectors who generously provided information and insight on the Digital Dilemma.

This report, in its various draft forms, was reviewed by several of the individuals listed below to assist the Council with delivering a final version that meets Academy standards for overall quality and accuracy. The reviewers' comments and identities will remain confidential to protect the peer review process. The reviewers have not seen the final version of this report, nor do they necessarily support the report's conclusions and recommendations. The Academy of Motion Picture Arts and Sciences is wholly responsible for the contents of the final report.

Science and Technology Council

Ray Feeney, Co-chair
Bill Taylor, Co-chair
Andy Maltz, Director

Digital Motion Picture Archival Project Committee

Phil Feiner, Chair
Milt Shefter, Report Project Lead
Steven Anastasi*
Craig Barron
Elizabeth Cohen*
Grover Crisp
Clay Davis*
Carl Fleischhauer
Rob Hummel*
Brad Hunt*
Tim Kittleson
Phil Lelyveld
Edrolfo Leones
Gregory Lukow
Gary Morse
Michael O'Brien
Michael Pogorzelski
Eddie Richmond
Garrett Smith*
Madi Solomon
Charles Swartz* *(in memoriam)*

Laurin Herr, *Consultant & Lead Interviewer*
Von Johnson, *Case Studies Consultant*
***Report Project Committee Member**

The Council would also like to thank all of the individuals, studios and other organizations that participated in the creation of this report:

David Arctur
Mary Baker
Art Beckman
Peter Boer
Ken Boettcher
Richard Buchanan
Laura Campbell
Dave Cavena
Robert Cecil
Chris Cookson
Brian Davis
Jeff Davis
John Faundeen
Michael Friend
Edward Grogan
Jim Hannafin
Wade Hannibal
Margaret Hedstrom
Robert Herman
Scott Kelly
Glenn Kennel
Mark Kimball
Bob Lambert
Jim Lindner
Stephen Long
Howard Lukk
Tad Marburg
Brian McKay
Richard Moore
Liz Murphy
Bob O'Neil
Mark Ostlund
Peter Otto
Rick Picton
Jerry Pierce
Brian Saunders
Greg Servos
Tim Smith
Greg Thagard
Ken Thibodeau
Rick Utley
Rudy Valdez
Chris Wood
Bob Yacenda

Amazon.com, Inc.
Ascent Media Group
The Cleveland Clinic Foundation
Eastman Kodak
Hess Corporation
HP Labs
Imation Corporation
InPhase Technologies
Laser Pacific Media Corporation
Library of Congress
Microsoft Corporation
National Archives and Records Administration
National Security Agency
NBC/Universal Studios
Office of the Director of National Intelligence
Open Geospatial Consortium
Ovation Data Services, Inc.
Paramount Pictures
Pro-tek Media Preservation Services
Quantum
San Diego Supercomputing Center
Sony Pictures Entertainment
Sun Microsystems, Inc.
20th Century Fox Film Corporation
United States Geological Survey, Center for Earth Resource Observation and Science
University of Michigan
The Walt Disney Company
Warner Bros.

ABOUT THE SCIENCE AND TECHNOLOGY COUNCIL

The mission of the Science and Technology Council of the Academy of Motion Picture Arts and Sciences is:

- *To advance the science of motion pictures and foster cooperation for technological progress in support of the art*
- *To sponsor publications and foster educational activities that facilitate understanding of historical and new developments both within the industry and for the wider public audience*
- *To preserve the history of the science and technology of motion pictures*
- *To provide a forum and common meeting ground for the exchange of information and to promote cooperation among divergent technological interests, with the objective of increasing the quality of the theatrical motion picture experience*

For more information on the Science and Technology Council, visit http://www.oscars.org/council.

ACADEMY IMPRINTS

THE DIGITAL DILEMMA: STRATEGIC ISSUES IN ARCHIVING AND ACCESSING DIGITAL MOTION PICTURE MATERIALS