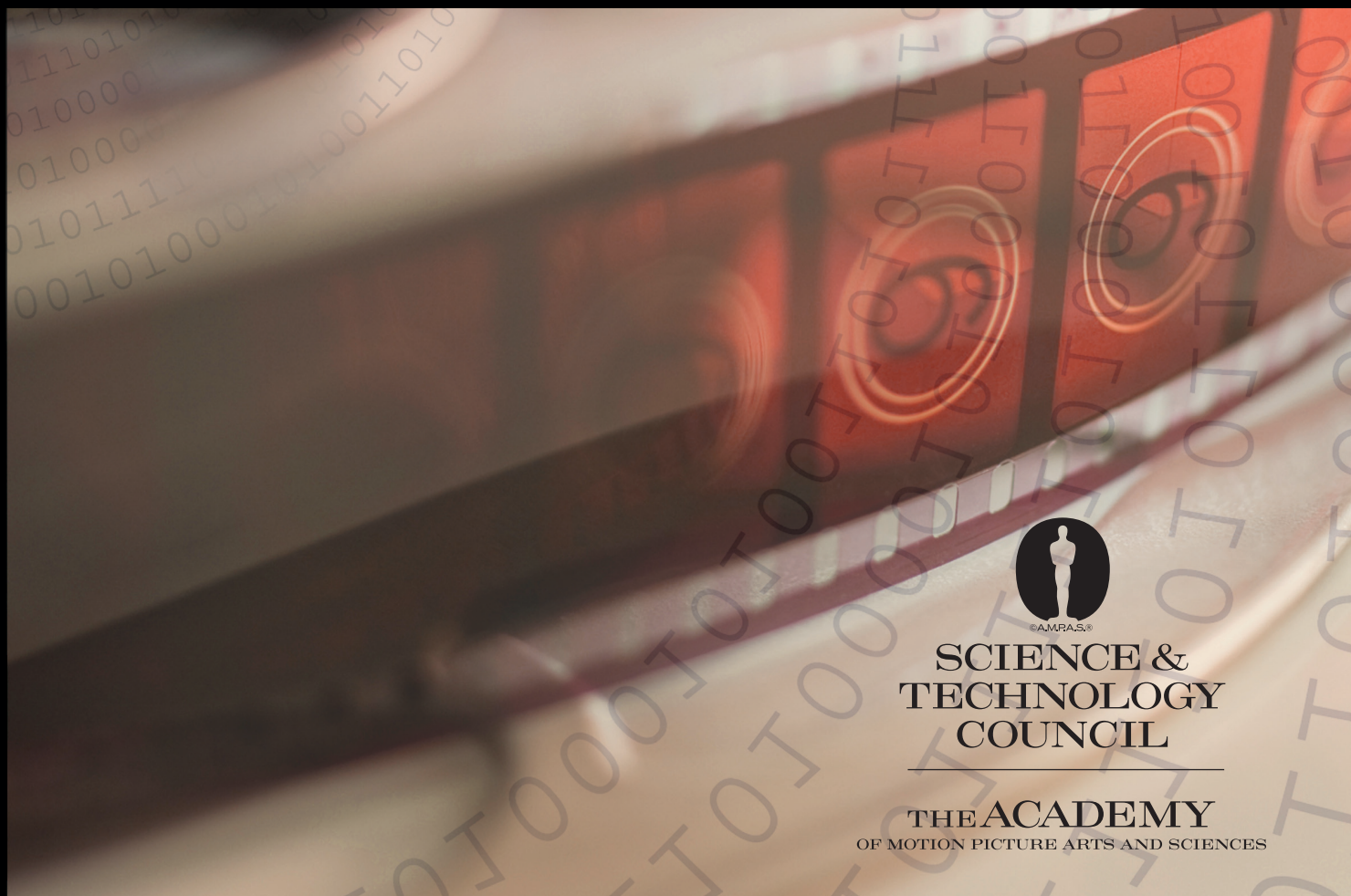
LONG-TERM MANAGEMENT AND STORAGE OF DIGITAL MOTION PICTURE MATERIALS

A DIGITAL MOTION PICTURE ARCHIVE FRAMEWORK PROJECT CASE STUDY



©A.M.P.A.S.®
SCIENCE & TECHNOLOGY COUNCIL
THE ACADEMY
OF MOTION PICTURE ARTS AND SCIENCES

LONG-TERM MANAGEMENT AND STORAGE OF DIGITAL MOTION PICTURE MATERIALS

A DIGITAL MOTION PICTURE ARCHIVE FRAMEWORK PROJECT CASE STUDY

SCIENCE & TECHNOLOGY COUNCIL

Published by the Academy of Motion Picture Arts and Sciences
Inquiries should be addressed to:
Science and Technology Council, Academy of Motion Picture Arts and Sciences
1313 Vine Street, Hollywood, CA 90028
310-247-3000
http://www.oscars.org

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Long-term Management and Storage of Digital Motion Picture Materials: a digital motion picture archive framework project case study/ Includes bibliographic references.
p. cm.

1. Digital preservation – Case studies. 2. Film archives – Technological innovations 3. Digital Cinematography I. Academy of Motion Picture Arts and Sciences – Science and Technology Council II. Nancy Lydon Silver III. Andrew Maltz

ISBN 978-0-615-39095-6
LCCN 2010911206

This publication is produced by the
Science and Technology Council of the Academy of Motion Picture Arts and Sciences
for the National Digital Information Infrastructure & Preservation Program:
A Collaborative Initiative of the Library of Congress

NATIONAL DIGITAL
INFORMATION INFRASTRUCTURE
AND PRESERVATION PROGRAM

TABLE OF CONTENTS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Motion picture film, when properly processed and stored, can last for more than 100 years at low cost and with minimal human intervention. Digital motion picture materials – which are rapidly replacing motion picture film in distribution, post-production, and principal photography – require substantial and perpetual expense and effort to preserve access. *The Digital Dilemma*, published by the Academy in 2007, reported that "there is no digital archival master format or process with longevity characteristics equivalent to that of film."[1]

The Academy of Motion Picture Arts and Sciences' Science and Technology Council, in collaboration with the Library of Congress under its National Digital Information and Infrastructure Preservation Program, undertook a case study project to discover the operational realities of various digital archiving strategies and technologies, as applied to digital motion picture materials, as a step toward finding a solution to the digital dilemma. This report discusses the Council's experiences in applying best preservation practices to an actual and historic digital motion picture collection in the Academy Film Archive.

The Standard Evaluation Material (StEM), co-produced by Digital Cinema Initiatives, LLC and the American Society of Cinematographers in 2003 for the testing and evaluation of digital projection equipment, was deposited in the Academy Film Archive in 2004. Shot on 35mm and 65mm film, the StEM was one of the earliest digitally mastered projects (a process also known as "digital intermediate"), and the collection includes both film and digital elements representative of a full-length theatrical motion picture. While the film elements were well documented and their associated long-term preservation practices are well understood, the **digital motion picture elements are fundamentally different from film with respect to general handling and long-term preservation.** There is much to be learned about the nature of digital motion picture materials from the study of general digital information management technology and practices. Many concepts from this field were applied to the design of the system developed as part of this project. Called "ACeSS," for Academy Case Study System, it is intended to be a learning tool and interim digital-collection management system that will preserve access to the StEM digital elements until the Academy adopts its own long-term digital preservation strategy.

EXECUTIVE SUMMARY

The key findings of this case study project are:

- Archival processing efforts and costs increase exponentially if digital materials are not "born archival." That is, metadata should be captured and created at the time of content creation, and organization of materials for archiving should be considered and implemented as part of the production process.

- There are no commercial digital asset management products designed for long-term archiving of digital motion picture materials. Therefore, the ACeSS project team adapted and integrated a suite of open source software tools, which are publicly available on a royalty-free basis.

- Implementing existing preservation metadata standards is complex, but crucial for maintaining long-term access to digital materials. Further development of digital motion picture technical metadata standards and file formats is required.

- Film archives need new staff skills to adequately handle digital motion picture materials. Content strategy, systems and software engineering, and technical project management were required for ACeSS system definition and development. Systems and database administration are also required for operational support.

- Developing and implementing a digital preservation strategy and supporting infrastructure from scratch requires substantial funding. ACeSS, designed to accommodate the 20-terabyte StEM collection and up to 76 terabytes of additional digital motion picture materials, cost approximately $600,000 in equipment and labor to develop.

ACeSS is now operational, and a future report will discuss the Council's experiences with using it. In the meantime, it is hoped that this report, and the technologies produced from this project, will serve as a framework and guiding path for film archives that need to manage and preserve their digital motion picture materials for an extended period of time.

1 INTRODUCTION AND BACKGROUND

A key deliverable of the Digital Motion Picture Archive Framework Project is building a system that organizes and stores digital motion picture materials using an actual and historic digital motion picture collection in order to discover the design challenges and operational realities of various digital motion picture archival strategies. The goal of this "learn by doing" project is to apply archival and library information management best practices to the long-term management and storage of digital motion picture materials. While "long term," as defined in *The Digital Dilemma*, is 100 years or longer,[2] the system built for this case study project is only expected to last long enough for the Academy to develop its own comprehensive long-term digital preservation strategy – probably on the order of several years. The key differences between this project and other efforts to build digital media storage systems are:

- **The design emphasis is on applying best film preservation and data curation practices, rather than forcing these practices to fit available technologies**
- **The resulting system is expected to serve only as an interim storage solution, and is likely to be replaced with a longer-term solution**

The project took place between 2008 and 2010 at the Academy's Pickford Center for Motion Picture Study in Hollywood, CA. The project team consisted of content strategists, film catalogers, metadata librarians, post-production specialists, software developers and computer network engineers. The result of this effort is the "ACeSS" system (Academy Case Study System).

The digital collection selected by the Academy for this project, known as the Standard Evaluation Material (StEM)[3], was deposited at the Academy Film Archive in 2004 by Digital Cinema Initiatives, LLC (DCI). The StEM was a collaboration between the major Hollywood studios and the American Society of Cinematographers (ASC), and was produced in 2003 as reference material for the testing of digital exhibition equipment. A team of ASC cinematographers designed and filmed two short pieces on the back lot at NBC/Universal Studios: the "Mini Movie," a 12-minute wedding sequence, and the "Display Reel," the film answer print from the cut negative. More than two hours of film were exposed in both 35mm and 65mm formats. This footage featured a number of scenes with a variety of lighting conditions, colors, textures and other variables of photographic definition including confetti, rain and fog. The final edited versions of the Mini Movie and Display Reel were digitally scanned at 6144 by 4168 pixel counts (known as "6K") to capture the detail contained in the film images. The 6K scans were then down-converted to both 4096 by 1714 ("4K") and 2048 by 857 ("2K") digital formats, which were subsequently used by DCI as a robust test suite of images for digital projectors, compression systems, and other elements of a digital cinema system.

INTRODUCTION AND BACKGROUND

Wedding Scene from the "Mini Movie" **FIGURE 1**

The StEM collection consists of 35mm and 65mm negatives, intermediate and print film elements, and digital data tape and hard disks containing digital versions of the StEM. The film elements consist primarily of 35mm and 65mm Mini Movie, Display Reel and related production elements. The digital components of the collection consist of CD-ROMs, DVDs, DAT (Digital Audio Tape), DTF2 (a digital data tape format), LTO (Linear Tape-Open), disk drive arrays and desktop disk drives.

The film and related digital materials were organized and prepared for deposit by a DCI production coordinator prior to deposit at the Academy Film Archive. A detailed written description of the film shoot and post-production process was also provided. This documentation consists of information about the cameras used, original logs and shot lists, floor plans, and camera assistants' notes. The written documentation also details the original lab and telecine reports as well as post-production information such as edit decision lists and credit and subtitle information. Finally, details of the StEM post-production schedule, the imaging chain for the DCI proof-of-concept test, requirements testing and the Display Reel were provided as part of the collection.

The Academy Film Archive completed a detailed physical inventory of the film elements and the digital motion picture materials. While the film elements were well documented and their preservation procedures well known, the digital portion of the collection was not only delivered with insufficient documentation for complete auditing and viewing purposes, but some of the data had already become inaccessible because of media obsolescence. Furthermore, at the time of the inventory, the Academy film archivists were not able to view or play back the digital files to see what actually was on the hard drives and data tapes. Viewing the StEM digital materials requires

INTRODUCTION AND BACKGROUND

sophisticated hardware and software that are not typically found in a film archive, and therefore archivists cataloged, packaged and stored the StEM collection to the best of their capability and resources.

In January 2008, the digital materials of the StEM collection were released to the Case Study team from the Academy Film Archive vaults, and a detailed content audit and assessment of the digital motion picture materials took place. The team consisted of content strategists, catalogers, network engineers and software developers who evaluated the contents of the disk drives and disk arrays after connecting them to appropriate computer hardware. In addition, detailed interviews of the original production and post-production teams occurred. Through this process it was determined that the core digital collection of the StEM consisted of the following:

- The 12-minute "Mini Movie," scanned at 2K, 4K and 6K, in both compressed and uncompressed formats
- The Display Reel, scanned at 4K and 6K
- Audio files
- Ancillary materials such as the paper archive (stored as Adobe Acrobat files) and still images documenting principal photography

The Academy used the StEM digital motion picture materials to examine data formatting, "packaging" (wrapping, preservation and technical metadata), cataloging (descriptive metadata), and storage issues. After several months of study, a computer network and storage system was designed and built to support the interim management of the StEM digital motion picture collection.

2 OVERVIEW OF FILM ARCHIVING

Introduction

Film archives are essentially collections of films, both complete works and, in many cases, fragments and source elements stored in environmental conditions that slow or stop chemical degradation processes. An effective archive integrates its holdings with up-to-date catalogs, indexes, and other tools needed to search and retrieve the assets stored in it. Archiving purposes vary, but in general, archiving is meant to systematically collect and protect assets valuable enough to keep "for the future." The term "archival" in this context is defined as storage of the master elements from which all downstream distribution materials can be created over a 100-year time frame.

Film archives exist to acquire and preserve motion pictures and motion picture elements. In addition to physical film media, most film archives preserve videotape, too. Film collections and policies vary, depending on the type of institution, its mission statement, and the user community it serves. The public nonprofit film archive generally tends to focus its collection on either a particular subject or a geographic locale. Theatrically released films, ethnographic footage, educational films, scientific films, news footage, home movies and avant-garde films are typical types and genres contained in collections that would be part of a nonprofit film archive. The mission of such institutions is generally to provide research and long-term access to their collections for the preservation of cultural heritage.

A commercial motion picture studio archive, however, focuses on the protection of its motion pictures for commercial repurposing. It is motivated to preserve its holdings by the financial benefit of maintaining and protecting its motion picture materials. Its collections consist of materials produced or purchased by the specific motion picture studio and its affiliates.

Whether employed by a nonprofit or commercial archive, the film archivist's work is devoted to saving and preserving motion pictures from decay. Film preservationists combat film's physical deterioration through an integrated strategy:

- Printing old film onto new, more stable film stock
- Storing film materials in proper environmental conditions
- Providing access through modern copies

When this strategy is successfully implemented, the public can study and enjoy access copies on film or electronic media, and archives can conserve the original source material and preservation masters so that they will be available for many years to come.[4]

OVERVIEW OF FILM ARCHIVING

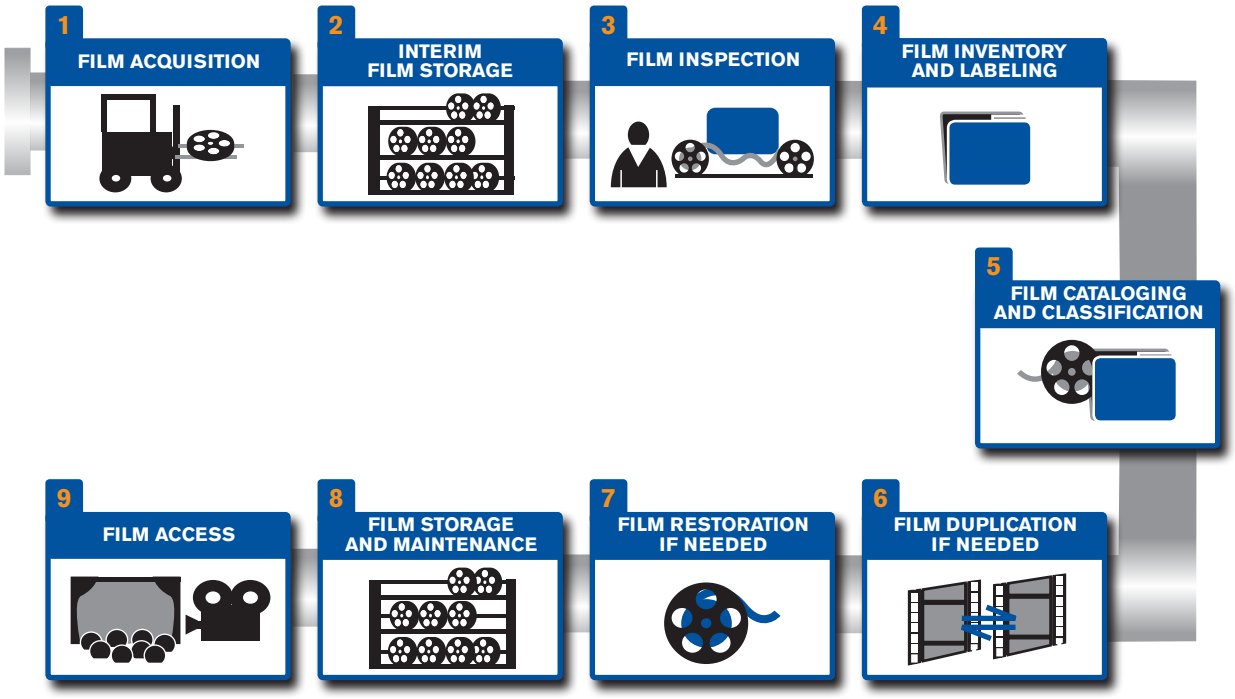
The film archiving process

The specifics of the archiving process of film elements may vary from institution to institution, but the core preservation tenets of a public film archive focus on:

- **conservation** – protecting the original film
- **duplication** – making a safety copy of the film
- **restoration** – the art of reconstructing a specific version of the film
- **access** – the process by which the film is shared with the community

Figure 2 shows a typical film archive workflow.

Film Archive Workflow **FIGURE 2**



OVERVIEW OF FILM ARCHIVING

There are several roles and functions at the public film archive for managing and preserving films. Some key job functions include:

Curator – This position is central to managing all aspects of the film archive's collection. The curator makes decisions regarding which films to collect, establishes policies and procedures for managing the collections, and oversees the overall management and operations of the film archive. Curators also direct and watch over the care and documentation for the films, conduct research based on the collection, provide proper paperwork for the transport and acquisition of the film, and share that research with the public and scholarly community through screenings and publications. Depending on the size of the institution, there can be one curator managing these activities or many curators managing several different collections and staff. Curators' educational backgrounds can vary: some are self-taught; others have undergraduate or graduate film archive, library information science, museum studies, or related arts degrees.

Archivist – Film archivists work closely with the film curator to arrange and care for film collections. Compared to the film curator, the film archivist has more specific roles such as hands-on processing of the collection or serving as the staff expert on a particular collection or collections. Film archivists' educational backgrounds can also vary in a similar fashion to those of curators.

Cataloger – The film cataloger is an information-management professional who assigns identifiers and subject headings, and otherwise creates the entire filmographic record of an archived asset. Catalogers are usually archive content subject-matter experts and maintain advanced degrees in either library information science or film archive management. Many have graduate degrees in both film studies and library information science. The cataloger maintains the database of information that the film archive uses to manage its collection.

Preservationist – The film preservationist's work focuses on the duplication, restoration, maintenance and physical handling of film. Preservationists' educational backgrounds can also vary in a similar fashion to those of curators and archivists.

Public Access Coordinator – As the liaison between the film archive and the public, the public access coordinator usually handles the loaning of a film to another institution and provides public outreach and promotion of the archive's collection. Most public access coordinators have graduate degrees in either film archive management or library information science.

OVERVIEW OF FILM ARCHIVING

Managing film collections – the process of cataloging and classification of films

According to *The Film Preservation Guide,* description is the key to managing film collections.[5] Description captures essential information about the film's physical characteristics and content and provides a textual link between the physical item and the end-user. In film archives, a basic form of description is cataloging. Cataloging and classification comprise the process of organizing information, resulting in the creation of a library or archive catalog.

A core function of a library or archive is cataloging and classification of its assets. The concept of cataloging started in France in the late 1700s.[6] Card catalogs were also popularized in the United States by Library of Congress (LC) cards. In the late 1960s, two developments changed the future of cataloging: The Library of Congress created the MARC record, enabling the machine readability of bibliographic records (a bibliographic record usually includes the description of the item, main entries, subject headings, and a call number),[7] and the Online Computer Library Center (OCLC) was developed in Dublin, Ohio, which started providing cataloging information via cable and terminal to all of its member libraries. These two developments paved the way for the creation of Online Public Access Catalogs (OPACs). Because of the considerable amount of cost savings, most libraries and archives converted to online catalogs and froze or discarded their print card catalogs.

A motion picture catalog describes the particular body of work, using cataloging codes such as AACR2 (Anglo-American Cataloguing Rules), RAK (die Regeln für die alphabetische Katalogisierung), RDA (Resource Description and Access), AMIM (Archival Moving Image Material) and the International Federation of Film Archives (FIAF) cataloging rules.

As part of the cataloging process, the film cataloger utilizes classification codes such as Library of Congress Classification, Dewey Decimal System or, most commonly, an in-house custom identifier usually generated as part of a database system. In addition, controlled vocabularies, taxonomies and name and title authorities are commonly used in the description cataloging process. Cataloging motion pictures can present more of a challenge than cataloging print items because books and similar publications are generally complete works that are produced separately, whereas motion picture archives may have several manifestations of a work as well as source materials, each of which can be incomplete, but when taken together, approximate a single whole work.

3 DIGITAL MOTION PICTURES AND FILM ARCHIVES

A brief introduction to digital motion picture materials

Digital motion picture materials are created in two different ways: they are either born digital, i.e., the materials originate in digital form from a digital motion picture camera, or they result from digitizing a film-based original. Today, all theatrical motion pictures have some digital image elements, and all have digital sound tracks.

Although some of the underlying technologies are similar, digital motion pictures are distinct from "video" or "television" for several reasons worth noting. For creative control and process reasons, digital motion picture image sequences are quite often generated as a series of individual image data files rather than as a serial "video clip." The range of colors, numerical precision and number of pixels in a theatrical digital motion picture are significantly greater than in television, even in its high-definition form. Image characteristics are also described using different nomenclature.

The term "digital cinematography" is usually applied to the principal photography process only in cases where digital acquisition is substituted for film acquisition. The term is not generally applied when digital acquisition is substituted for analog video acquisition, as with live-broadcast television programs.

Digital motion picture image data is quite often converted to a convenient "working" image file format, depending on the post-production facility, editing equipment and visual effects requirements and process. Near the end of the post-production process, all of the completed master image files are "conformed" to match an edit list created by the film editor,[8] and are then color-corrected in a process called "mastering." The end result is a set of master image files that are combined with the final sound files (and subtitles, if needed) into a Digital Cinema Distribution Master (DCDM). Image compression is applied to reduce the size of the DCDM, and a digital "print" is created for theatrical distribution to digital cinemas. These Digital Cinema Packages (DCPs) will also have secure data encryption applied to prevent unauthorized use or theft.

These distinctions are important for film archives that have accepted analog video materials in the past, not just because the terminology is different, but the handling procedures for digital motion pictures are fundamentally different from those for video. And while many television productions are moving to tapeless, digital file-based workflows, they generally do not have the same level of technical complexity as digital motion pictures. Both cases, however, face the same underlying challenges of preserving large numbers of large digital files.

DIGITAL MOTION PICTURES AND FILM ARCHIVES

How film archives are being impacted by digital motion picture materials

Today, many film archives are receiving digital motion picture materials, and most film archives are receiving digital files of some kind. Film archives receiving digital motion picture materials are facing a plethora of complicated digital image formats: TIFF, MOV, MXF, DPX, WAV, TXT and PDF,[9] with and without image compression, and in varying physical conditions. Compliance with existing preservation standards requires fundamentally different approaches to archiving than the established and reliable processes developed for film over the last 100 years. Furthermore, the expanding use of DCPs presents another new challenge for archives because, as mentioned earlier in this report, a DCP is likely to be encrypted and requires a special "decryption key" to unlock the file for viewing. If a key is lost by the archive, or is not delivered to the archive with the content, the content is not accessible, and therefore not preservable.

4 DIGITAL INFORMATION MANAGEMENT FUNDAMENTALS

Digital collections, libraries and repositories

Because important digital content has been around for many years in other application spaces, the field of library and information science has studied many of the issues related to the management and preservation of digital data. While digital motion picture materials bring their own set of unique requirements, it is useful to understand some of the fundamentals of digital information management.

Digital materials are best organized as collections. A digital collection is a set of related items, all in an electronic form. Collection management for digital materials involves the same principles as traditional print collection management. Like print collections, digital collections must be selected, organized and managed to meet the needs of the particular archive's mission. A set of one or more digital collections is often referred to as a digital library. Digital libraries are often organized by collections that are defined by the communities they serve. Digital libraries are created using collections-management software that catalogs and organizes the digital materials. There are many benefits to providing electronic access to digital collections, especially via the Internet. Researchers no longer need to physically travel to libraries or archives to find the information they need. Digital libraries provide quick and easy access to material that was once difficult to obtain.[10]

Digital repositories offer a way of storing and retrieving digital material. A digital repository is simply a "place" to store, access, and preserve digital materials on a single computer or computer network. Digital repositories are very similar to digital libraries, and they are sometimes referred to interchangeably. The focus of a digital repository, in addition to providing search functions and access to digital materials, is the safekeeping of the digital object and related metadata. Some key features of digital repositories are content versioning, relationships, event histories, extensible metadata management, audit trails of modifications to the digital material deposited in the repository, and monitoring and alerting services.[11]

With the increased development of digital libraries and repositories, there grew a need to develop standards in support of the long-term archiving of digital information.[12] The Consultative Committee for Space Data Systems (CCSDS) coordinated the specification of the reference model for an Open Archival Information System (OAIS). Originally designed for data obtained from observations of the terrestrial and space environments, the model has found application in other communities such as universities, archives, libraries and museums. *"The OAIS model describes a conceptual framework for a complete, generic archival system. Positioned at a high level, it is defined as 'an organisation of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community.'"*[13] The OAIS model today serves as a framework for institutions developing digital libraries, repositories and related archival databases.

DIGITAL INFORMATION MANAGEMENT FUNDAMENTALS

Another theoretical framework for digital archiving and preservation comes from the American Library Association (ALA). As defined by the ALA:

> *Digital preservation combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change. Digital preservation applies to both born digital and reformatted content.*
>
> *Digital preservation policies document an organization's commitment to preserve digital content for future use; specify file formats to be preserved and the level of preservation to be provided; and ensure compliance with standards and best practices for responsible stewardship of digital information.*
>
> *Digital preservation strategies and actions address content creation, integrity and maintenance.*

Content creation includes:

- *Clear and complete technical specifications*
- *Production of reliable master files*
- *Sufficient descriptive, administrative and structural metadata to ensure future access*
- *Detailed quality control of processes*

Content integrity includes:

- *Documentation of all policies, strategies and procedures*
- *Use of persistent identifiers*
- *Recorded provenance and change history for all objects*
- *Verification mechanisms*
- *Attention to security requirements*
- *Routine audits*

Content maintenance includes:

- *A robust computing and networking infrastructure*
- *Storage and synchronization of files at multiple sites*
- *Continuous monitoring and management of files*
- *Programs for refreshing, migration and emulation*
- *Creation and testing of disaster prevention and recovery plans*
- *Periodic review and updating of policies and procedures*[14]

DIGITAL INFORMATION MANAGEMENT FUNDAMENTALS

Content audits

A key step in managing a digital collection is completing a content audit. Content audits are also known as content assessments or inventories. The content audit helps uncover the full scope of what exists in the digital collection. It enables the user to evaluate, see patterns in, and learn about the collection. Most content audits are documented in either a spreadsheet or database, with each entry representing attributes of the content.[15]

In addition to content audits, special focus is given to the cataloging and classification of the digital materials. This enables proper and easy retrieval of the digital materials.[16]

Cataloging and classification for digital materials

The 21st century ushered in a wave of new technologies that forged the need for digital collection management for digital materials. This new technology brought major developments in cataloging, digital libraries and the creation of metadata. "Metadata are structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities."[17] There are also different types of metadata: descriptive, administrative, preservation, and technical. Dublin Core, MODS, EAD, MARC 21, and PREMIS are some notable metadata standards and frameworks for describing and organizing digital materials.

Metadata development and maintenance are generally driven by companies or other organizations to serve the needs of their specific fields. For example, PBCore is a metadata schema for public broadcasting content developed by the Corporation for Public Broadcasting, and has proved to be useful for organizing moving-image materials. PREMIS (Preservation Metadata: Implemention Strategies) is a preservation metadata dictionary and data model maintained by the Library of Congress that consists of five interrelated entities: Intellectual, Object, Event, Agent, and Rights. SMPTE (the Society of Motion Picture and Television Engineers) maintains RP210, which is a metadata dictionary for moving images. RP210 use is limited for digital motion picture metadata; it focuses primarily on television broadcast metadata. There is current activity within FIAF, the International Federation of Film Archives, and CEN/TC372, the European Committee for Standardization, to develop metadata standards for cinematographic works. To date, a minimum set of information elements for the unambiguous identification of film works has been completed. Further work is in progress to develop a specification for structuring machine-processable metadata describing cinematographic works.[18]

5 REQUIREMENTS FOR MANAGING THE StEM COLLECTION

Approach to requirements development

With an understanding of the general approaches to managing digital collections, the next step in the process was to develop workflows and system requirements that extended these approaches to include the unique aspects of digital motion picture collections, specifically the StEM collection. The primary characteristic that differentiates digital motion picture collections from typical digital libraries is the large number of individual large number of individual items. Digital motion picture collections are composed of hundreds of thousands, perhaps millions, of individual digital objects, and the large size of the objects (tens of megabytes to hundreds of gigabytes per object), the variety of metadata (sometimes proprietary) and file formats, as well as the unique organization of collection elements (raw scans, reels, intermediate and final versions, etc.), make digital motion picture materials uniquely difficult to manage. To aid organization, this material can be grouped into three levels: Collections, Works, and Assets. Assets are the individual objects that are the components that form a whole Work. Usually the descriptive and technical metadata is captured at the Asset level. A Work can be composed of many Assets. Intellectual property and rights metadata is usually captured at the Work level. And finally, Collections are formed through groupings of similar Works. Their grouping is usually based on institutional and user needs, such as provenance or classification.

After studying the StEM collection and evaluating the various digital formats and condition of the media, the project team determined the operational and system requirements and specifications for a system to ingest, catalog, and manage the StEM digital materials over the course of several years. This system would initially manage the StEM digital materials, but would also have the capability to manage a limited number of other digital motion picture collections, such as DCPs and digital restoration elements currently in the Academy Film Archive, digital motion picture test materials from a variety of Council technology projects and future digital collections that might need to be managed until the Academy Film Archive adopts its own long-term digital preservation solution. The system requirements were based upon a variety of information science and archiving best practices (for both traditional film and electronic media), strategies from seasoned information management professionals and engineers, and models such as the OAIS Reference Model and the American Library Association's definition and framework for digital preservation, discussed earlier in this report.

REQUIREMENTS FOR MANAGING THE StEM COLLECTION

Based on the analysis of the StEM collection, the initial digital workflow that the requirements were based on consisted of:

1. Perform content audit to inventory, quality check and view the StEM digital materials.
2. Restore corrupted StEM digital materials if needed.
3. Harvest metadata from existing StEM digital materials.
4. Organize the StEM digital materials into the corresponding collections:
   - Mini Movie 2K, 4K and 6K versions
   - Display Reel 4K and 6K versions
   - Related digital audio materials
   - Digital ancillary production material
5. Catalog the StEM digital materials by collection, work and asset.
6. Create a descriptive metadata schema for the StEM digital materials (with a standard such as PBCore, and use SMPTE and PREMIS metadata dictionaries if applicable) and utilize unique DCI nomenclature and Library of Congress subject headings.
7. Create a catalog record for the Mini Movie, Display Reel and related production elements, and populate the record with the appropriate descriptive metadata.
8. Ingest appropriate StEM digital material into the catalog and associate it with corresponding StEM collection record.
9. Create thumbnails of media associated with descriptive metadata record.
10. Transfer the StEM digital materials and associated metadata records to the digital repository.
11. Export the StEM digital materials and associated metadata records to LTO-4 data tape and store in an environmentally controlled vault at the Academy Film Archive.
12. Export the StEM digital materials and associated metadata records to a trusted offsite repository.
13. Monitor the integrity of the stored digital materials and corresponding metadata.
14. Maintain proper security of the stored digital materials.
15. Maintain and update digital preservation policies and procedures, which includes a disaster recovery plan that is periodically tested.
16. Refresh and/or migrate the stored materials until a long-term digital preservation solution is identified and adopted.

REQUIREMENTS FOR MANAGING THE StEM COLLECTION

Functional system requirements

The functional requirements for the case study system were broken down by:

1. Viewing of incoming and stored digital motion picture materials
2. Cataloging and long-term storage management of digital motion picture materials
3. Storage requirements: size and lifetime
4. Technical infrastructure requirements

Viewing requirements

The system needed to provide the ability to view digital motion picture materials in the following file formats, each with its own image encoding and/or compression specifications: DPX, OpenEXR, TIFF, MXF, WAV, JPEG, and PSD. System users need to view incoming digital motion picture materials as part of the quality assessment, which is a component of the content audit. Viewing the incoming materials is also important for developing an understanding of the content so a record of the materials can be created.

Cataloging requirements

To ensure sustainability and interoperability with other catalogs, the system needed to implement an accepted cataloging standard for moving-image materials. After a number of approaches were considered, a combination of PBCore (an audio/visual metadata standard developed by the Corporation for Public Broadcasting for their archives) and the Library of Congress's PREMIS preservation metadata standard was selected. The pairing of standards enables the system to extend the descriptive flexibility and pragmatic design of PBCore with the well-developed preservation metadata structures of PREMIS, creating a hybrid particularly well-suited for digital motion picture preservation.

Storage requirements: size and lifetime

After assessing the types and sizes of the collections to be preserved, the team determined that a three-tiered storage system would be required:

Tier 1: high-performance storage tier for active processing of a digital motion picture collection. Initially, 12 terabytes of tier 1 storage would be needed, likely growing to 24 terabytes over the life of the system.

Tier 2: high-reliability, lower-performance storage tier for temporary storage of in-process digital motion picture materials that might be needed to complete collection processing. Approximately 24 terabytes of tier 2 storage would be needed, likely growing to 48 terabytes over the life of the system.

REQUIREMENTS FOR MANAGING THE StEM COLLECTION

Tier 3: high-reliability, lowest-performance, lowest-cost "dark" storage tier for ingested materials. The project team concluded that the initial set of collections to be ingested was approximately 12 terabytes, which would likely grow to 96 terabytes over the useful life of the system.

Although a goal of this case study project is to apply best preservation practices as a priority in the design of a long-term digital motion picture preservation system, the practical reality is that the case study system is an interim solution, designed to reliably store Academy assets until a 100-year archiving solution is designed and implemented. The project team specified an operating lifetime of the system of two to five years, which is believed to be sufficient time for the Academy Film Archive to develop and adopt a long-term preservation strategy.

Technical infrastructure requirements

Digital storage systems that meet the size and lifetime requirements stated above resulted in a further set of technical requirements: a high-performance computer network (10Gbit or faster) that would support the software components developed during this project requiring minimal network latency (a performance reduction generally caused by the addition of excessive network components such as hubs and routers) and compatibility with the range of devices that would be connected to the network. Ingest and viewing workstations were required, as well as remote access capabilities for cataloging activities via a web browser interface and system administration via the Internet for 24-hour technical support.

6 TECHNOLOGY SELECTION PROCESS AND INFRASTRUCTURE DEVELOPMENT

Product and technology survey

As with any digital information management system development project, user requirements drive workflows, which in turn drive software and hardware selection. As discussed earlier, the project team iterated through several drafts of proposed workflows for the various stages of collection management.

A product and technology survey was conducted in tandem with workflow development so that each effort would inform the other, resulting in a more complete analysis. Digital motion picture archiving workflows have unique requirements that may or may not be supportable by commercial or open source digital library and "archiving"[19] systems, but there were things to learn from existing state-of-the-art systems as well.

After surveying the market for digital asset management systems, digital library systems and open source digital preservation frameworks, the project team invited several leading commercial solution providers and open source solution providers to demonstrate their products' capabilities. Each of these systems had strong points, some of which included:

- User-friendly software interfaces
- The ability to view proxies of cataloged content
- The ability to create frame-accurate shot lists
- Built-in security tools

While these systems all enjoy varying levels of commercial success and, in the case of the open source frameworks, varying levels of adoption, most of them had at least one substantial deficiency relative to the case study system requirements. These deficiencies included:

- Licensing fees or up-front costs in excess of the project budget or what a typical archive would be able to afford
- Proprietary data formats with no export mechanism to a standard or open format
- Insufficient cataloging and collection management capabilities
- Lack of support for reading and transcoding required file formats such as OpenEXR, DPX and DCP
- Inability to view, ingest and manage large numbers of large moving-image files
- Lack of support for repository functionality
- Lack of integrated support for LTO-4 and similar data-backup devices

TECHNOLOGY SELECTION PROCESS AND INFRASTRUCTURE DEVELOPMENT

The project team also found it difficult to establish a suitable research-oriented working relationship with commercial vendors. For understandable reasons, the commercial vendors were not interested in implementing features for which they did not see significant future revenue possibilities, and there is no market survey data quantifying the size of the market for long-term digital preservation products.

The project team ultimately selected an open source cataloging application, called **CollectiveAccess**, as the "point of departure" for software development. For a preservation test system, open source-licensed software is an attractive alternative to proprietary commercial software. With open source, the application code may be freely modified and customized without restriction. This flexibility allowed the project team to rapidly adapt and modify the cataloging application to best address evolving project requirements. With proprietary commercial software, such modifications would have required coordination with and permission from the author vendor, significantly slowing development and increasing cost. CollectiveAccess's features will be discussed at length later in this report, but there were three key factors leading to its selection:

1. CollectiveAccess's impressive base feature set included a modular and extensible media-handling architecture as well as configurable support for several metadata and cataloging standards.

2. As an open source project, there would be no initial or ongoing license fee, and all computer source code is available under the Educational Community License (ECL) 2.0, an open source license acceptable to the project team.

3. CollectiveAccess is capable of importing digital media in bulk into the cataloging system and repository, allowing for mass import of pre-existing data files.

The project team also selected several repositories that would each be integrated with CollectiveAccess and evaluated once the overall system was completed. The repositories selected were:

- **iRODS**[20] **(integrated Rule Oriented Data System)** is a data grid or "intelligent cloud" software system developed by the Data Intensive Cyber Environments research group (developers of the SRB, the Storage Resource Broker) and collaborators. iRODS management policies (sets of assertions user communities make about their digital collections) are characterized in iRODS rules and state information. At the iRODS core, a rule engine interprets these rules to decide how the system is to respond to various requests and conditions. iRODS is an open source project available under a BSD license.

TECHNOLOGY SELECTION PROCESS AND INFRASTRUCTURE DEVELOPMENT

- **Fedora Commons**[21] **(Flexible Extensible Digital Object Repository Architecture)** is a modular architecture built on the principle that interoperability and extensibility are best achieved by the integration of data, interfaces, and mechanisms (i.e., executable programs) as clearly defined modules. Fedora is a digital asset-management (DAM) architecture, upon which many types of institutional repositories, digital archives, and digital library systems might be built. Fedora is the underlying architecture for a digital repository, and is not a complete management, indexing, discovery or delivery application.

- **ZFS file system**[22] While technically not a repository per se, CollectiveAccess connects directly to any file system, and ZFS has several features that enable good preservation practices, e.g., audit trails and flexible storage system configuration and performance monitoring. ZFS is a combined file system and logical volume manager designed by Sun Microsystems (now part of Oracle). ZFS includes support for large storage capacities, integration of the concepts of file system and volume management, snapshots and copy-on-write clones, continuous integrity checking and automatic repair. ZFS is implemented as open source software, licensed under the Common Development and Distribution License (CDDL).

Once the base software was selected, the project team was able to specify a hardware platform on which to run the software. Sun Microsystems, Inc.'s Sun Fire™ X4540 "Thor" storage server and accompanying J4400 JBOD Arrays and 10Gbit Ethernet were selected for several reasons:

- This hardware configuration is quite common among the digital library and high-performance storage research community, and the project team felt it would be well supported.

- The X4540 server, when properly configured and coupled with a 10 Gbit network "backbone," would have sufficient data throughput for the types of operations expected to be performed on the system.

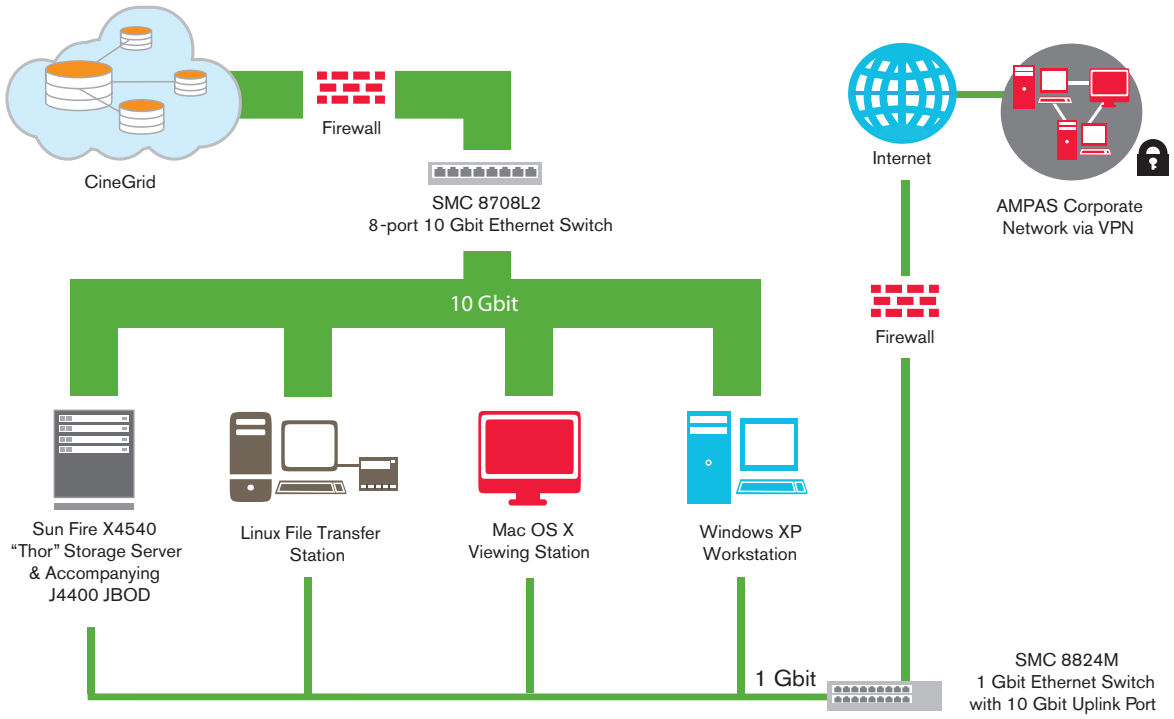
The details and challenges of implementing these choices into the Academy Case Study System follow:

Building a network infrastructure to view, ingest, process and store the StEM digital materials

Figure 3 is a simplified block diagram of the network, storage and processing infrastructure built to support the case study system. Nicknamed "CouncilNet," the network also supports other Council activities including development of the Image Interchange Framework (IIF),[23] an architecture for the exchange and unambiguous interpretation of digital motion picture images.

TECHNOLOGY SELECTION PROCESS AND INFRASTRUCTURE DEVELOPMENT

CouncilNet Overview **FIGURE 3**



The key components of CouncilNet are:

- **SMC 8708L2 stand-alone 8-port 10 Gbit Ethernet switch:** this forms the backbone of the network, and all high-performance components connect to this switch. Connection to the CineGrid research network (described below) is also achieved via this switch.
- **SMC 8824M 1 Gbit Ethernet switch with 10 Gbit uplink port:** enables connection of non–10 Gbit devices such as cataloging workstations to CouncilNet.
- **Firewall:** provides secure connection between CouncilNet and the Academy Corporate network. This enables a "virtual private network" connection to be established among catalogers in the Academy Film Archive and CouncilNet, as well as for system administrators from remote locations over the public Internet.
- **Storage server:** the Sun Fire X4540 server, incorporating tiers 1 and 2 of the tiered storage system described earlier. CollectiveAccess and some repository software are hosted on this server as well.
- **Transfer Station:** a general-purpose workstation for transferring digital motion picture materials into and out of the storage server. The LTO-4 tape drive, which requires a host computer, is connected to this workstation.
- **Viewing Station:** a high-performance workstation for viewing full-fidelity digital motion picture materials on a digital cinema projector.

TECHNOLOGY SELECTION PROCESS AND INFRASTRUCTURE DEVELOPMENT

- **Primary Cataloging Station:** a high-performance workstation for viewing and cataloging digital motion picture materials.

While the project team would have preferred to use only one computer operating system, the functional requirements of the overall system necessitated the use of a variety of them:

- **Sun Solaris 10:** chosen for the Sun Fire X4540 server because it supports the ZFS file system.
- **Linux:** chosen for the transfer station because it is fully configurable.
- **Mac OS X:** chosen for the primary cataloging and viewing station because of its reliability, ease of use, and its application software requirements for viewing digital motion picture materials.
- **Windows XP:** chosen for cataloging and accessing the Academy corporate network for email and related applications.

To meet the requirement for multiple, geographically diverse copies of the digital motion picture materials to be ingested, the project team decided to collaborate on a "grid storage" research project with CineGrid, a nonprofit international membership organization administratively based in California, but composed of a globally diverse community of research institutions. CineGrid's mission is to build an interdisciplinary community focused on the research, development and demonstration of networked collaborative tools, enabling the production, use, preservation and exchange of very high-quality digital media over high-speed photonic networks. The research project consisted of implementing the iRODS data grid within CineGrid's own content-management system, known as CineGrid Exchange (CX). The CX distributed repository would then be available to the case study system as an offsite repository, and CX would provide additional end-user input for the overall design of the case study system.[24]

7 BUILDING A SOFTWARE SYSTEM TO PROCESS AND STORE THE StEM DIGITAL MATERIALS

With the key software components selected, the project team focused on the details of integrating these components and extending their feature sets where required.

Selected software tools for cataloging and digital collection management

As mentioned earlier in this report, the core collection management software selected for development and customization is an open source project called CollectiveAccess. Developed by Whirl-i-Gig, Inc., CollectiveAccess is a highly configurable cataloging tool and web-based application for museums, archives and digital collections. Available free of charge under the open source Educational Community License (ECL) 2.0, the same license used for other widely used open source products, CollectiveAccess supports a variety of metadata standards and customizable schemas, external data sources and repositories and many popular media formats.[25] It also has comprehensive support for multilingual cataloging. Use of a wide variety of roman and non-roman character sets is possible via CollectiveAccess's support for Unicode.

CollectiveAccess employs a relational model for organizing collections, which is particularly helpful for managing digital motion picture materials because several manifestations of a work may exist, each of which can be incomplete, but when taken together, approximate a single whole item. A relational database enables, as part of the cataloging process, the creation of logical relationships between bodies of work, assets, names, places and collections. This is important for the creation of a sensible catalog from a diverse and seemingly disparate set of elements within a collection.

Selected software tools for transformation and viewing of image file formats

Digital motion picture collections generally include highly specialized media file formats due to the extremely high quality requirements of commercial motion pictures. While CollectiveAccess supported many common media formats "out of the box," it did not initially support all of the file formats used in the StEM collection. To implement support for these formats, including DPX, OpenEXR and DCP 1.0, modular extensions were developed. These "plug-ins" wrap existing media processing libraries such as ImageMagick, CoreImage or FFmpeg in a way that allows them to be used by the CollectiveAccess media processing components. For OpenEXR support, ImageMagick with the OpenEXR open source libraries was selected. DPX support is included in ImageMagick as well.

Digital object repositories

Several repository options were selected since a system had never been built to the requirements listed and the project team felt there was much to learn by using a few

BUILDING A SOFTWARE SYSTEM TO PROCESS AND STORE THE StEM DIGITAL MATERIALS

different ones. Fedora Commons and iRODS were initially selected, in addition to the "files and folders" approach using the ZFS file system, which has its own set of preservation-oriented features. Fedora was selected for its archival features, mature implementation, acceptable performance, scalability, and a very active user community. iRODS was selected because it enables construction of a geographically distributed repository, and is well suited for very large data sets.

Some highlights of each of the selected repositories follow:

1. **Fedora Commons**[26]

   Key features:

   1. **Powerful digital object model:** In case of corruption or failure, Fedora has a rebuild utility that can completely rebuild the repository by crawling the digital object XML source files that are stored on disk – so if the Fedora database fails or is lost, it is possible to restore the entire repository simply by reading these files off the disk or a backup tape. Fedora-hosted digital objects are stored in a METS-like package called Fedora Object XML5. All of the metadata (descriptive, preservation and relationship to other objects) and managed datastreams that make up a digital object are serialized into a single XML file on a file system. By backing up those XML files, a preservation copy of the entire system is created.

   2. **Content versioning:** Fedora repositories offer implementers the option of versioning data objects. When a data object is versioned, the object's audit trail is updated to reflect the changes made to the object as well as when the change was made and by whom. A new version of the modified data is also added to the object's XML. This new datastream cascades from the original and is numbered to show the relationship between original and version. This allows users to retrieve older versions of a data object by performing a date/time search and retrieval, or the most current version if the date/time criteria are not included in the search.

   3. **Expressive inter-object relationships:** Relationships between objects can be stored via the metadata included in the objects. This allows implementers to link together related objects into parent/child relationships, and can be used to mirror relationships created in the CollectiveAccess catalog. This mirrored relationship may enhance the utility of applications interoperating with ACeSS data via the repository by providing a richer data model than would be possible with a simple file store.

BUILDING A SOFTWARE SYSTEM TO PROCESS AND STORE THE StEM DIGITAL MATERIALS

4. **Event history:** Every object in a Fedora repository contains an audit trail, which preserves a record of every change made to the object.
5. **Extensible metadata management:** Fedora's metadata management features enhance preservation by simplifying the process by which catalog data can be attached to media objects in the repository. This allows CollectiveAccess to export serialized catalog data as an intrinsic part of the archived media object.
6. **Audit trail of all modifications to objects**
7. **Provenance and history of content development over time recorded**
8. **Digital objects record extensive object properties:**
   - Includes created and modified dates, MIME type, format identifiers
   - Checksum (MD5, SHA1, etc.) preservation validation and integrity service
   - Datastreams validate the bytestream format
   - Digital objects validated based on content models
9. **Preservation monitoring and alerting service:**
   - Message broker for special events
   - Checksum failure alert
   - Email preservation manager
   - Kick off an automated process (e.g., migrate)

2. iRODS[27]

Key features:

1. **Data replication service**
2. **Periodic data integrity check**
3. **Distributed storages for disaster recovery**
4. **Metadata support for preservation description information**
5. **Administrative metadata is managed to ensure that authenticity and chain of custody are preserved**
6. **Audit trail management**
7. **Can track changes to the preservation environment**
8. **Implement policies through the use of rules**
9. **Scalable collections –** can manage both small and large amounts of information

BUILDING A SOFTWARE SYSTEM TO PROCESS AND STORE THE StEM DIGITAL MATERIALS

3. ZFS[28]

ZFS is a file system and logical volume manager designed by Sun Microsystems (now owned by Oracle) and made available as a part of their SOLARIS operating system. While ZFS is not designed to provide the functionality of a full repository system, its support for very large storage systems and high performance make it worthwhile to perform comparisons and evaluation.

While LTO-4 data tape does not meet the strict definition of a preservation repository, its status as a de facto preservation standard in many motion picture archives makes it worth mentioning in this section to complete the overall data-storage picture. Workflows and processes were developed to create backup copies of all repository contents – both media object and catalog information – to LTO-4 data tape, and this is discussed in Chapter 9.

The Academy Case Study System: ACeSS

With the individual system components selected, a simplified view of the overall system is now possible:

- **A cataloging application** (CollectiveAccess) providing a user interface to manage and describe archived objects, including digital motion picture materials and collateral materials.
- **A digital repository** (choice of Fedora, iRODS or ZFS with backup to LTO-4) capable of storing large volumes of digital motion picture data – individual frames, audio tracks, and full-motion "service" copies – as well as collateral materials.
- **A media ingestion system** (CollectiveAccess) capable of importing digital media in bulk into the cataloging system and repository.
- **A media transformation framework** (CollectiveAccess) capable of converting media in various digital motion picture formats (DPX, OpenEXR, Broadcast WAV, etc.) to formats viewable in mainstream software such as web browsers.

The components of ACeSS fit together in a linear fashion. The ingestion system imports data into the cataloging application (CollectiveAccess), which includes a media transformation framework that handles the many image, audio, video and document formats in the collections to be managed. CollectiveAccess then copies the imported media and derived preview versions to the active repository or repositories. Finally, a backup system transfers one or more copies of the repository content to LTO-4 data tape or other backup media for offsite storage.

8 CUSTOMIZING COLLECTIVEACCESS FOR ACeSS

As stated earlier, CollectiveAccess was deemed to be a satisfactory point of departure for ACeSS's cataloging and user interface functions. Several aspects of CollectiveAccess were then configured, customized and extended, with the expected result being an appropriate tool for managing digital motion picture materials, especially the StEM Collection. Broadly speaking, the modifications included:

1. An appropriate schema and configuration profile for cataloging digital motion picture materials were designed and implemented.

2. The CollectiveAccess database and application user interface were extended to support the specific requirements of digital motion picture materials. For example, efficiently supporting large quantities – tens or hundreds of thousands – of image files attached to a single cataloged object and efficiently supporting very large digital media files (>1 terabyte) required modifications to CollectiveAccess's media-handling modules.

3. An abstracted repository interface was developed that enabled CollectiveAccess to connect to a choice of external digital repositories.

4. File format support was extended to include the additional file formats used for digital motion pictures.

Implementing PBCore and PREMIS metadata schemas in CollectiveAccess

One of the largest challenges of this project was redesigning CollectiveAccess's metadata schema to incorporate PBCore and PREMIS Metadata. This required a change to the original PBCore nomenclature and additional metadata fields to accommodate information about digital motion picture materials (see appendix for the revised PBCore and PREMIS metadata schemas). For example, the PBCore term "Intellectual Content" was changed to "Work" and "Instantiation" to "Asset" to accommodate for existing workflows and terminology within the Academy. Additional digital motion picture terminology and descriptors were also added. Similarly, CollectiveAccess terminology, which originated in museums and libraries, was replaced with terminology more suitable for a motion picture archive. For example, the "entity" authority list (individual and corporate names) was changed to the "names" authority.

It was also determined that the PREMIS metadata schema in its entirety was too extensive for this project, so a subset of core PREMIS metadata fields was chosen. The core required elements were used, as well as those that would capture the necessary preservation metadata for digital motion pictures.

Exposing these schemas in a sensible way also presented user interface design challenges. An intuitive cataloging workflow process had to be developed for catalogers.

CUSTOMIZING COLLECTIVEACCESS FOR ACeSS

To accomplish this, a single data-entry form for all metadata was designed to accommodate cataloging workflows.

Implementing media handling in CollectiveAccess

The base CollectiveAccess media-handling system was built with the assumption that high-resolution "original" media are always hosted and controlled by CollectiveAccess. In an environment where individual media items are of a "reasonable" size and can be moved across the network within an acceptable amount of time, this assumption provides many benefits, the most important of which is a guarantee of media availability, i.e., no "broken links."

Requiring files to be copied into CollectiveAccess for cataloging and access becomes a problem when the files (or groups of files) in question are very large (hundreds or thousands of gigabytes). For these files, the cost in network bandwidth, transmission time and redundant storage is extremely high, and these resource requirements make the processing of data at the scale required for feature-length motion pictures impractical.

In a typical museum or physical-object archive setting, media files larger than 10 gigabytes are virtually unheard of; the vast majority of files tend to be well under 200 megabytes in size, and the ratio of media items to cataloging object is usually less than 10 to 1. For digital motion pictures, most individual files are in the same size range as their museum/physical-object archive counterparts, but certain files (notably Digital Cinema Packages) may be much larger, in the hundreds of gigabytes range. Furthermore, smaller files – which are often frame images – tend to come in much larger bunches than is typical in other applications. Where a typical museum object might have no more than 10 image files attached, a typical digital motion picture object will often have 100,000 or more image files. For modern high-fidelity digital motion pictures, each frame is similar in size to a high-quality still image scan, between 10 and 150 megabytes.

Clearly, routine copying of terabytes of motion picture data using commonly available computing and networking hardware is not practical for cataloging at the scale required by even a moderately sized motion picture archive. The solution to this problem is to, when necessary, simply avoid copying altogether. This was accomplished by relaxing CollectiveAccess's requirement that direct control be retained over original media. An "import reference" option was designed and implemented, allowing CollectiveAccess catalog records to link to externally hosted resources. When importing references, rather than providing the original media, a URL to the media in an external repository is provided, along with a lower-resolution proxy to be used for the generation of CollectiveAccess hosted derivatives for preview. The lower-resolution proxy need not be a complete version of the resource itself; it can be a frame image, an excerpt of a video stream, even an icon – whatever is appropriate for use as preview media.

CUSTOMIZING COLLECTIVEACCESS FOR ACeSS

Handling large numbers of files

Storage of digital media files in CollectiveAccess's initial design assumes relatively limited numbers of media attached to a single cataloging object. Each file may have arbitrarily complex metadata structures attached, including time-based cataloging. This is desirable for the typical museum/archive user and allows for flexible modeling of access control, rights management, crediting, etc.; however, it is not ideal for situations where there are thousands of frame images with limited metadata attached to a single object. In these situations the overhead of supporting these options quickly becomes undesirable and unsustainable.

As part of this project, an alternative storage mechanism for frame images was implemented. This "lightweight" media file storage system enables an unbounded number of frame images (or other files) to be efficiently attached to the existing CollectiveAccess media representation data structure. Thus, a set of 100,000+ frame images representing a motion picture object is attachable to the object as a single representation composed of the many frame images, rather than as 100,000+ "heavy" object representation records, as would ordinarily be required. The result is a media cataloging system that performs acceptably at the scale required by a motion picture archive that is managing digital motion pictures.

Linking CollectiveAccess to a choice of digital storage repositories

Support was added to CollectiveAccess for digital storage repositories. As described earlier, digital storage repositories are stand-alone storage systems that store and provide access to digital objects deposited by an application such as CollectiveAccess. Digital storage repositories may also provide additional features such as access auditing and preservation services.

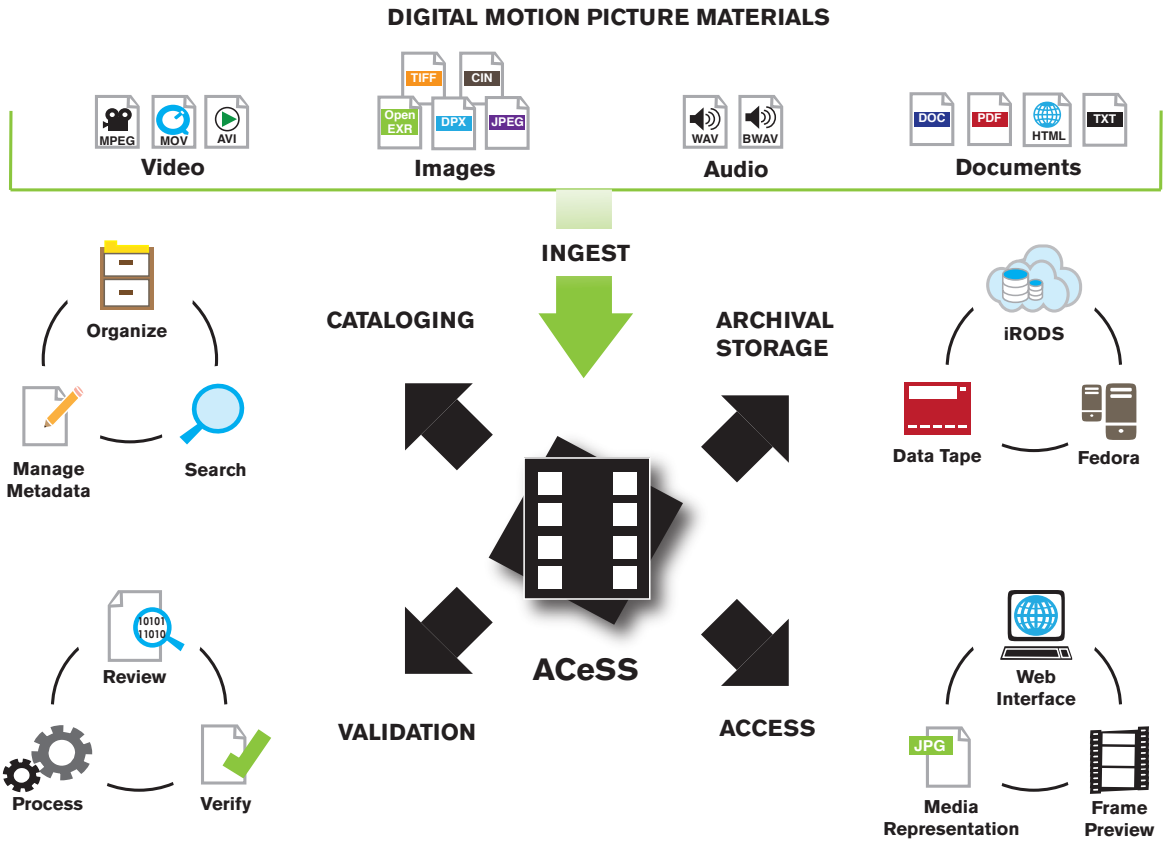
To better support use of repositories with CollectiveAccess, a storage abstraction layer was designed and implemented. This layer makes it possible for CollectiveAccess to deposit media and metadata into any supported repository, thereby leveraging the preservation functionality of various repository implementations. Support for repositories, as well as local file systems, is implemented as "plug-ins" – modular units of code that enable CollectiveAccess to interact with the target repository. Each plug-in translates generic actions ("insert object into repository," for example) into repository-specific actions as required. This approach enables ACeSS to support a wide range of storage repositories without requiring modifications to the CollectiveAccess core. For this project, plug-ins were implemented to support Fedora and iRODS, as well as local ZFS file systems.

CUSTOMIZING COLLECTIVEACCESS FOR ACeSS

Specific functionality defined within the abstraction layer includes replication of media and cataloging (administrative, technical and descriptive metadata) to external repositories.

CollectiveAccess is able to copy hosted media to external repositories for preservation, distribution or interoperation with other applications. The replication is automatic, configurable, and includes, along with media data, XML-serialized catalog data sufficient to reconstitute the media-associated catalog outside of CollectiveAccess. This feature is a critical component of CollectiveAccess's use in the ACeSS preservation workflow. Essentially, CollectiveAccess serves as cataloging user interface and framework. The replication features ensure that data (both metadata and media) cataloged and structured using CollectiveAccess's tools can be preserved over the long-term in a formal purpose-built repository. Figure 4 shows the overall arrangement of the ACeSS components and their interrelationships.

ACeSS Components and Process **FIGURE 4**



CUSTOMIZING COLLECTIVEACCESS FOR ACeSS

Monitoring of replication status

Replication of data can take significant time and may fail due to network issues, repository downtime, or some other unanticipated event. CollectiveAccess records the replication status – whether an item has been successfully replicated, if a replication attempt is in progress, or has failed – and makes status information available to CollectiveAccess users. Checksums are calculated by CollectiveAccess upon ingestion and prior to insertion in repositories. These checksums are stored in the CollectiveAccess database and can be compared to those computed by the target repositories.

Support for data backup devices

Backup of data is pervasive throughout the ACeSS architecture. At the most basic level, it is assumed that targeted storage repositories will implement their own backup methods. In addition, CollectiveAccess has been extended with an application plug-in that integrates directly with the Academy's "TapeOp Daemon." The TapeOp Daemon is a web service providing large-scale data backup using an LTO-4 tape device. With the TapeOp Daemon, a user can initiate a transfer of files from one server to the TapeOp server via FTP, and then initiate backup of those files onto LTO-4 tape. The Daemon provides feedback and drive status reports on demand, allowing service clients such as CollectiveAccess to provide real-time backup job status information to end-users.

9 PROCESSING THE StEM COLLECTION IN ACeSS

Content audit of the StEM

A key step in managing a digital motion picture collection is completing a content audit. The content audit for the StEM took place in multiple iterations. The initial content audit was both a physical and digital inventory of the StEM collection being held at the Academy Film Archive. The goal of this content audit was to sufficiently understand the contents of the collection so that ACeSS hardware and software requirements could be specified.

It was also determined in reviewing the physical inventory that the "Mini Movie" was recorded on Sony DTF2 (digital tape format) data tapes. In order to view and provide a data quality check and inventory of these digital materials, a DTF2 player would need to be purchased and the data would need to be offloaded onto another storage medium for direct access. Sony DTF2 players are obsolete. Occasionally they can be purchased by vendors through eBay for around $5,000. Another option was to outsource the data transfer job to a data service provider. The estimated cost for transferring the data from DTF2 to LTO data tapes was approximately $10,000 to $30,000, depending on the vendor.

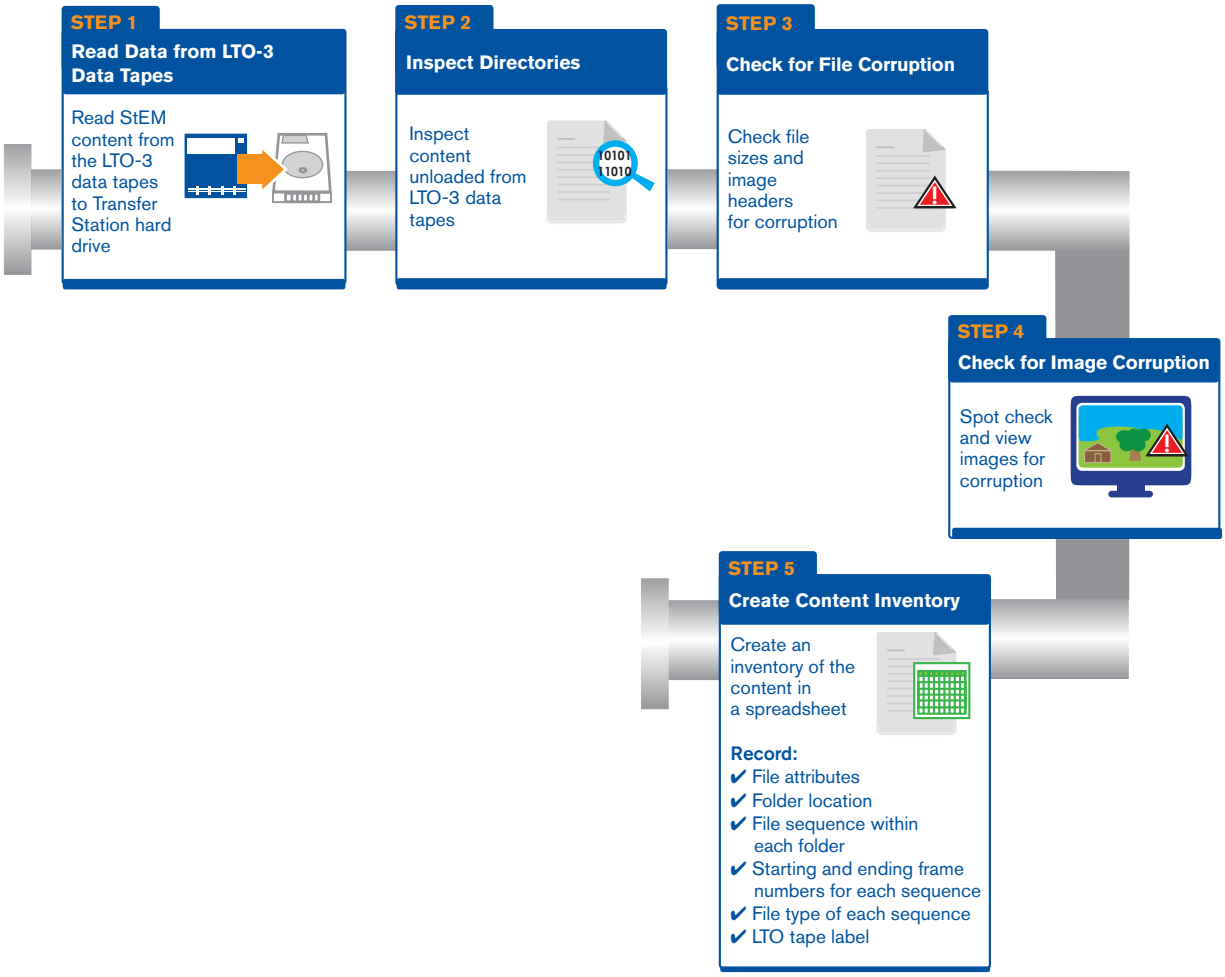
Before a decision was made on how to copy the DTF2 tapes, a full copy of StEM data on LTO-3 data tapes surfaced at Pacific Title Imaging, which subsequently made it available to the Academy.

The project team was able to copy the StEM data from these data tapes to its active network storage, at which time another content audit of the StEM digital materials took place.

The next content audit took place as the StEM collection was organized on CouncilNet prior to the initial ingests into ACeSS. During this process the project team copied, migrated and verified the StEM collection from LTO-3 tape to the CouncilNet Archive Server. Once the StEM collection was verified as not having been corrupted, the data was organized by the Mini Movie, Display Reel, and supporting production elements. The items were then grouped together by version (e.g., 2K, 4K, 6K). Preparing and organizing the assets ahead of time made it easier to ingest and catalog in ACeSS. Figure 5 shows the content audit and LTO transfer workflow for the StEM collection.

PROCESSING THE StEM COLLECTION IN ACeSS

StEM Content Audit & LTO Workflow **FIGURE 5**

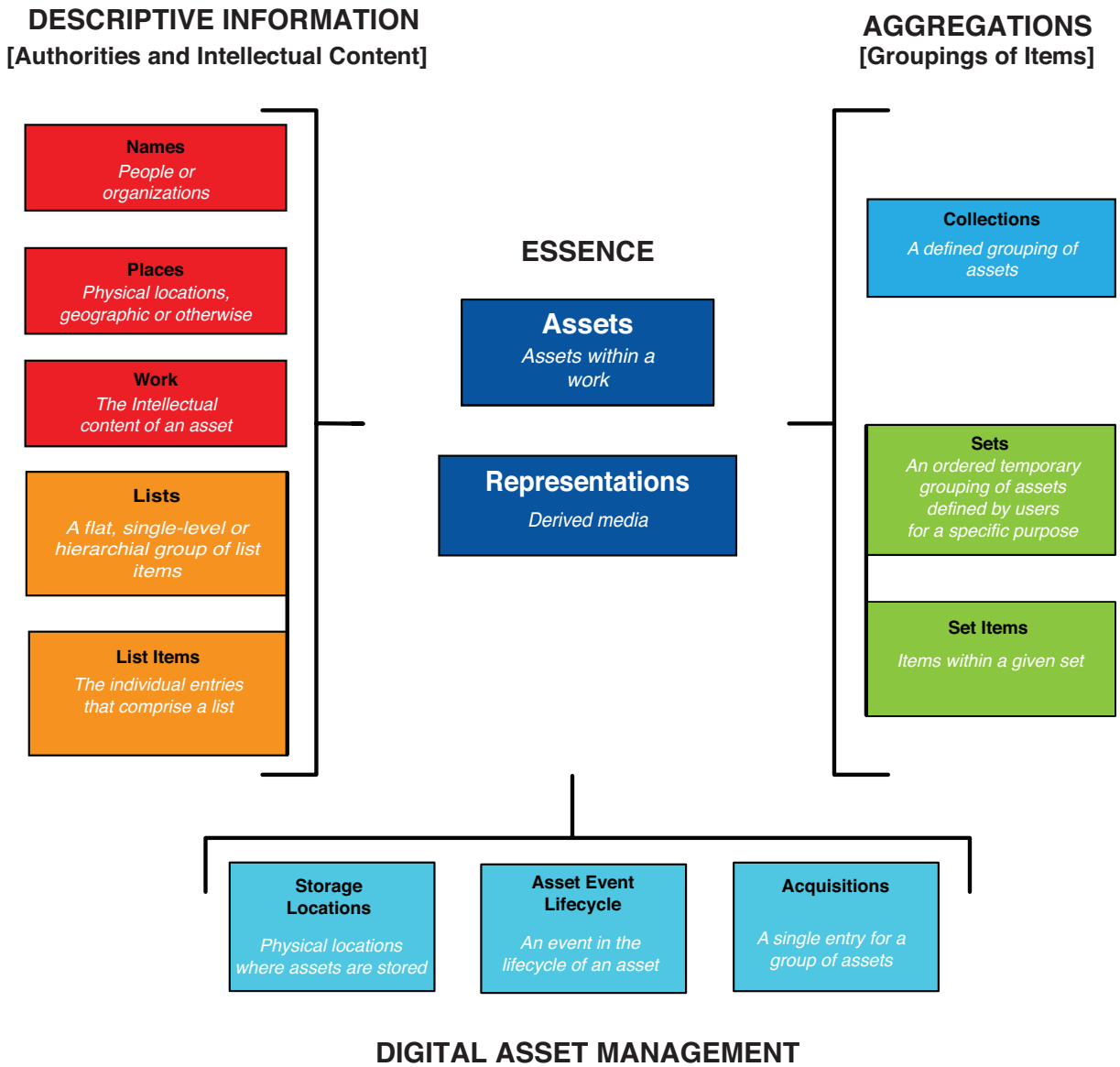


Cataloging the StEM collection in ACeSS

To work effectively with ACeSS it is critical to understand the fundamental structures in the CollectiveAccess database. While CollectiveAccess provides great flexibility in terms of the specifics of the data representing a given collection, it is up to the archive designers to determine the specific data fields in the CollectiveAccess catalog. CollectiveAccess's general data model provides a basic structure to describe the collection assets and their interrelationships. This model defines thirteen types of "items" that a collection may include.[29] Figure 6 displays the ACeSS data model.

PROCESSING THE StEM COLLECTION IN ACeSS

ACeSS Data Model **FIGURE 6**

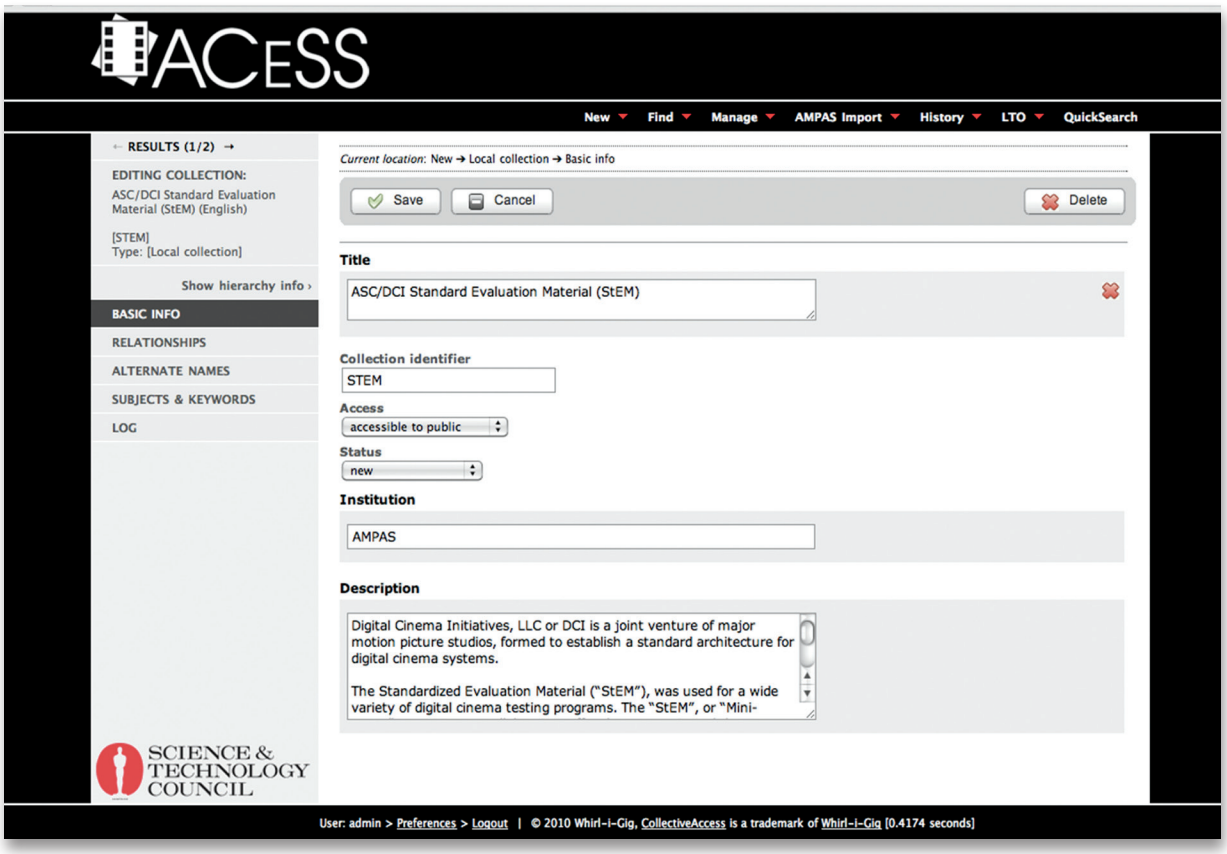


PROCESSING THE StEM COLLECTION IN ACeSS

With the data model tailored to properly describe digital motion picture materials, the StEM collection cataloging process could now begin.

ACeSS allows the user to navigate between forms for creating, editing and searching for records. The screenshot below is an example of a collection record in ACeSS.

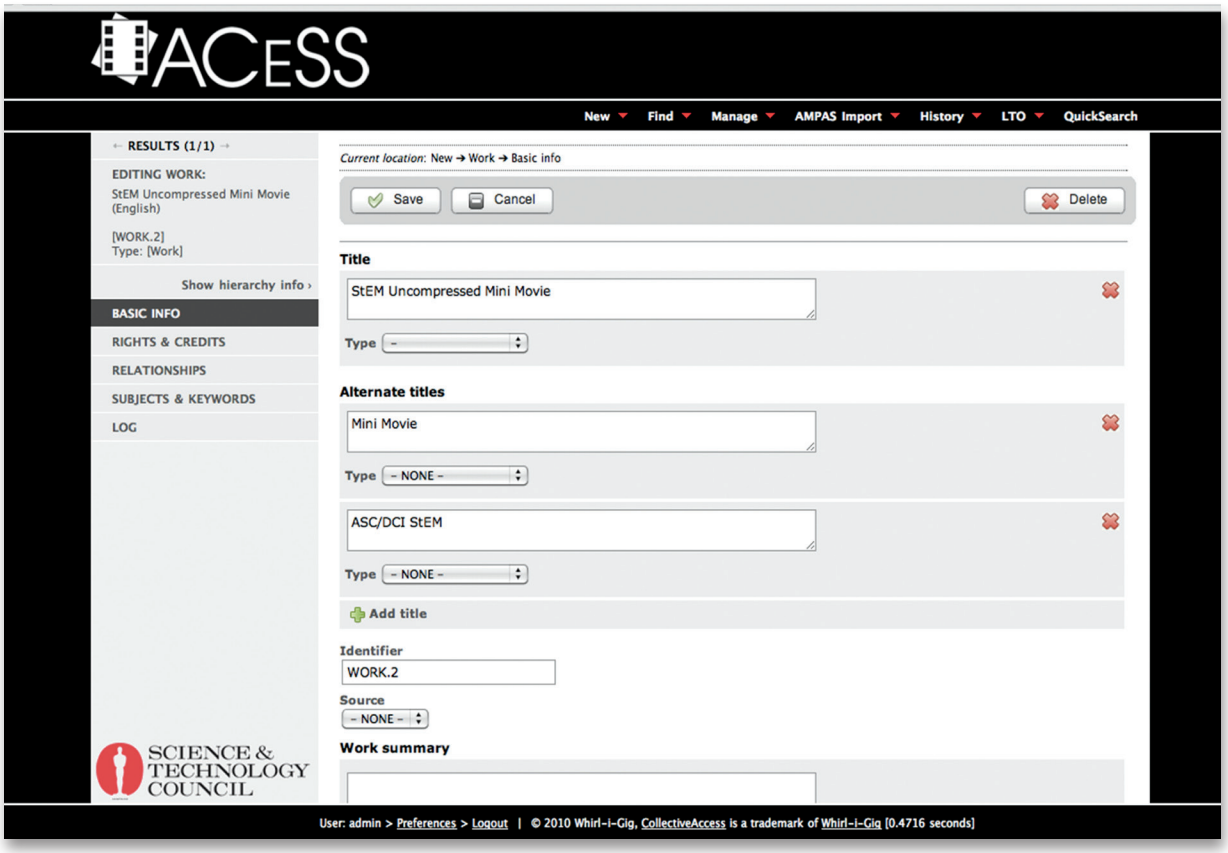
StEM Collection Record **FIGURE 7**



PROCESSING THE StEM COLLECTION IN ACeSS

This is an example of the Mini Movie work record. The information displayed in this work record consists of title, identifiers, rights and credits, relationships, subject, and keywords.

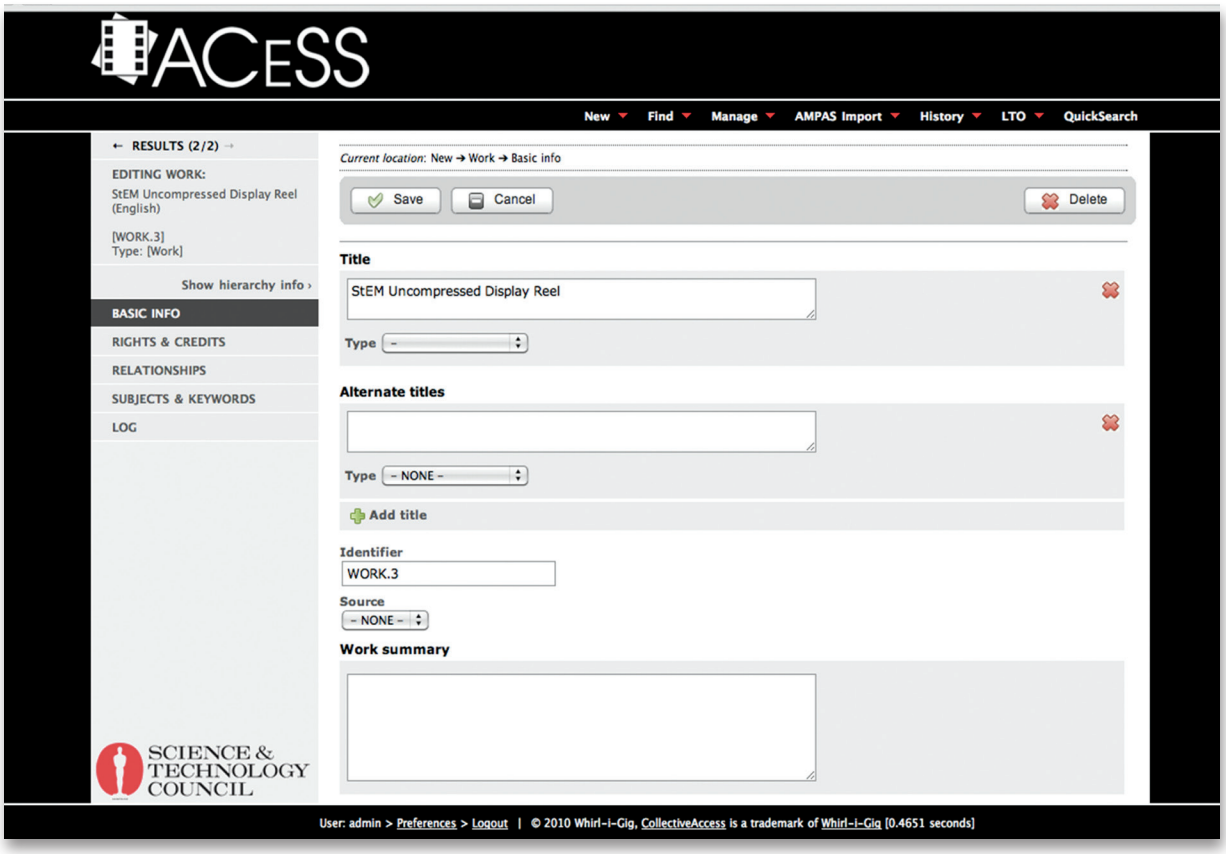
Mini Movie Work Record **FIGURE 8**



PROCESSING THE StEM COLLECTION IN ACeSS

This is an example of the Display Reel work record. The information displayed in this work record consists of title, identifiers, rights and credits, relationships, subject, and keywords.

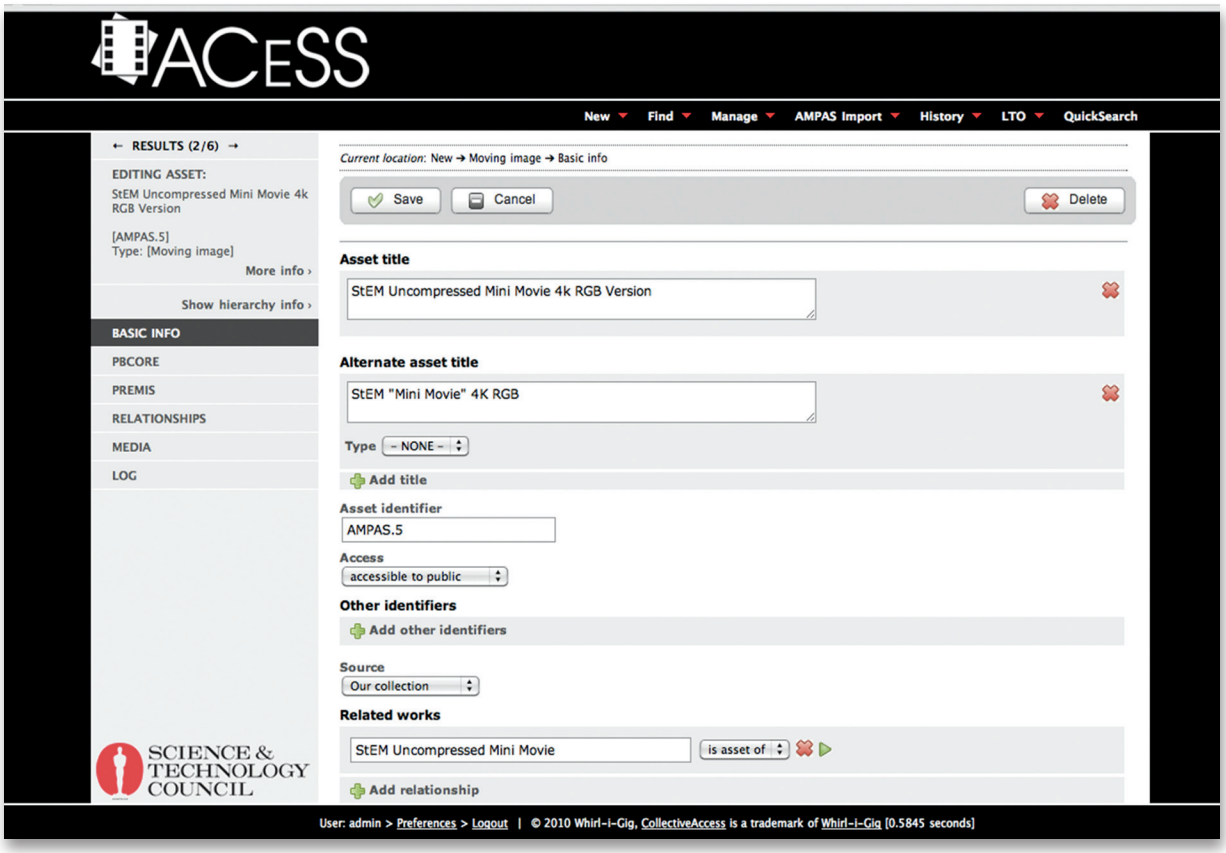
Display Reel Work Record **FIGURE 9**



PROCESSING THE StEM COLLECTION IN ACeSS

This is an example of the Mini Movie asset record. The information displayed in this asset record consists of basic info, PBCore metadata, PREMIS metadata, relationships and media.

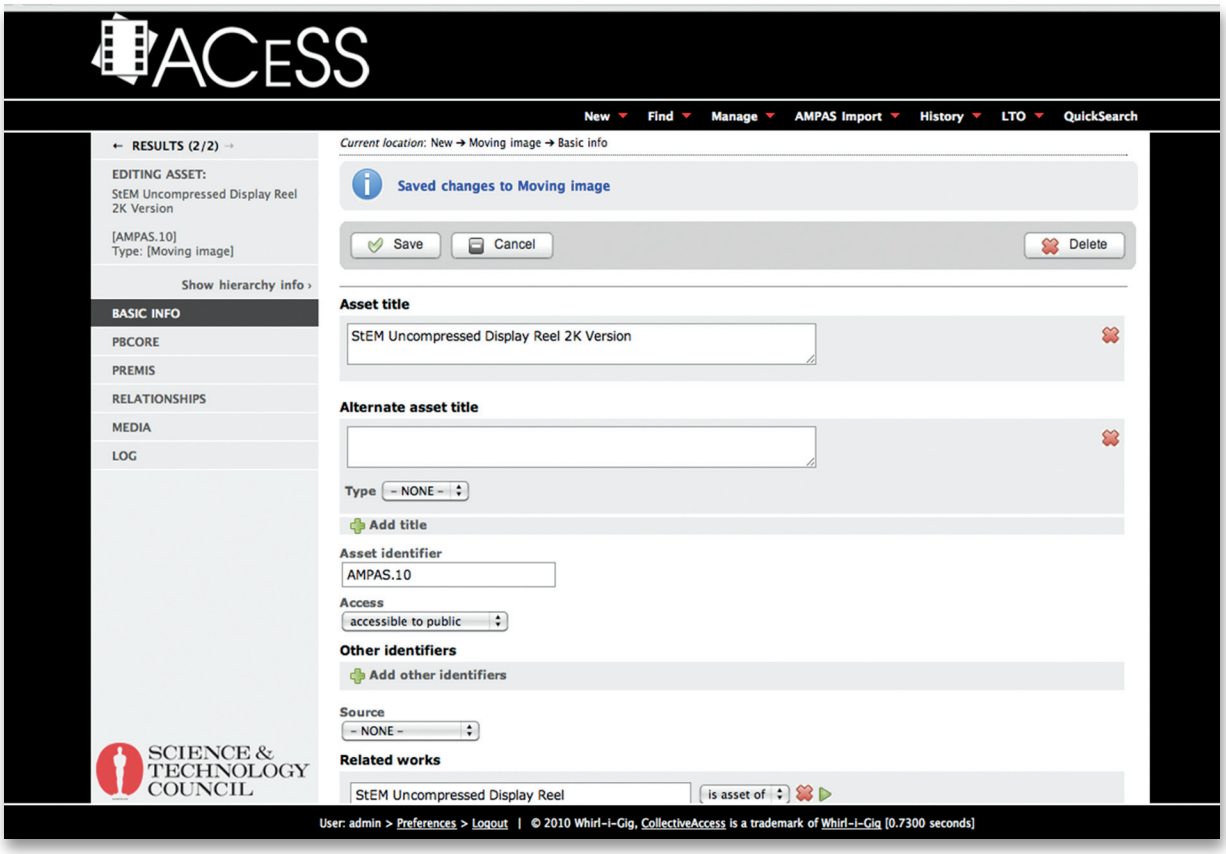
Mini Movie Asset Record **FIGURE 10**



PROCESSING THE StEM COLLECTION IN ACeSS

This is an example of the Display Reel asset record. The information displayed in this asset record consists of basic info, PBCore metadata, PREMIS metadata, relationships and media.

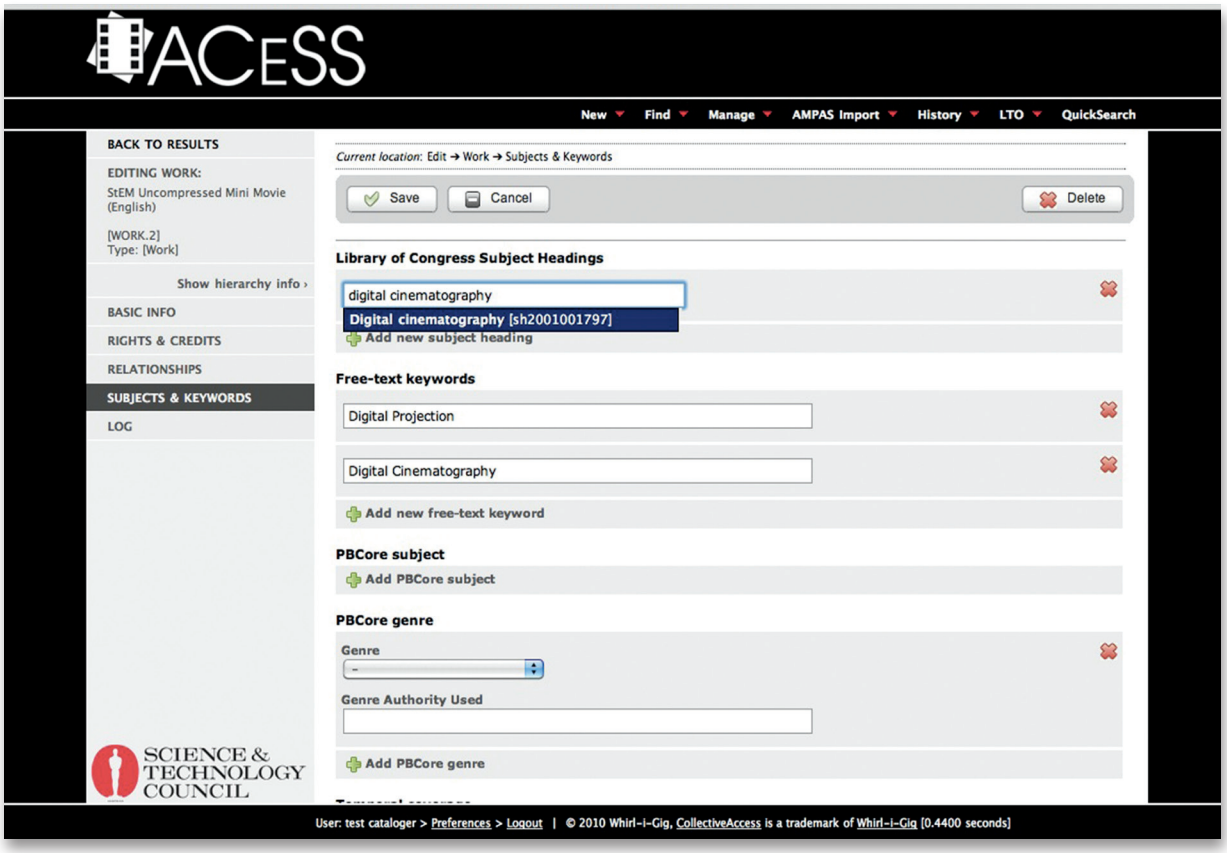
Display Reel Asset Record **FIGURE 11**



PROCESSING THE StEM COLLECTION IN ACeSS

The cataloger enters descriptive metadata for the corresponding work and asset records. The system also has the ability to utilize controlled vocabularies as well as to link to subject heading databases such as the *Library of Congress Authorities* and the Getty *Art & Architecture Thesaurus* (AAT®).

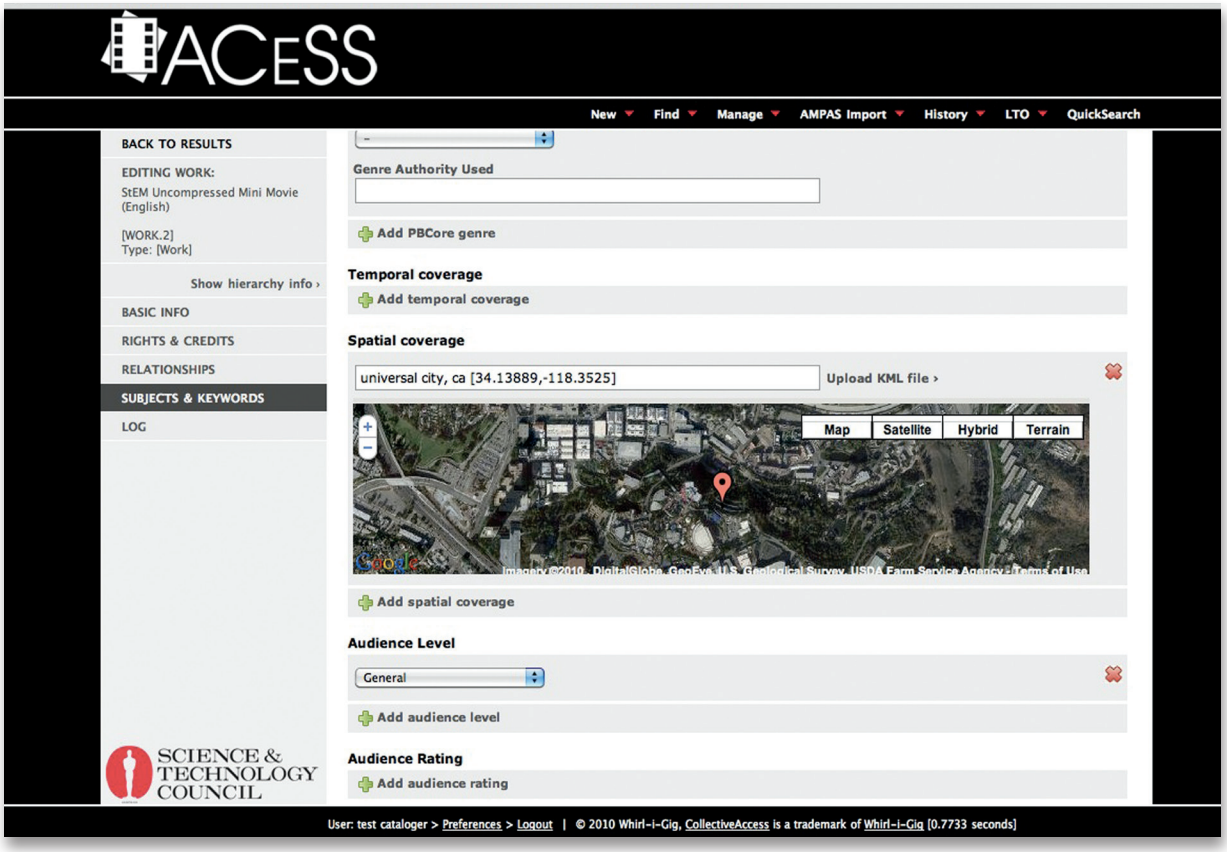
ACeSS Library of Congress Subject Heading **FIGURE 12**



PROCESSING THE StEM COLLECTION IN ACeSS

Georeferencing is another feature that has been implemented in the system. Georeferencing allows linking a physical location to an object. This can be done in two ways in CollectiveAccess: by using Google Earth or GeoNames. ACeSS uses Google Earth.

ACeSS Georeferencing **FIGURE 13**



PROCESSING THE StEM COLLECTION IN ACeSS

Cataloging is a core function of ACeSS. Figure 14 displays the cataloging workflow of ACeSS.

Cataloging Workflow **FIGURE 14**

PROCESSING THE StEM COLLECTION IN ACeSS

Ingesting the StEM collection in ACeSS

Once a portion of the StEM collection was cataloged, the actual StEM digital image files in this portion could be ingested into the system. In preparation for ingesting, a representation container for these frames was created, which is a multi-file structure subordinate to a standard CollectiveAccess object representation. The user then selected the frames to be ingested and initiated the ingestion process. Users can upload media either by manually attaching individual files or through a batch ingest process. Checksums are calculated by CollectiveAccess upon ingestion and prior to replication to repositories. These checksums are stored in the CollectiveAccess database and can be compared to those computed by the target repositories as a data integrity check.

Large numbers of large frames result in relatively long ingestion times, even with high-performance computing hardware and a high-performance network. For selected StEM collection portions, the measured ingest durations were:

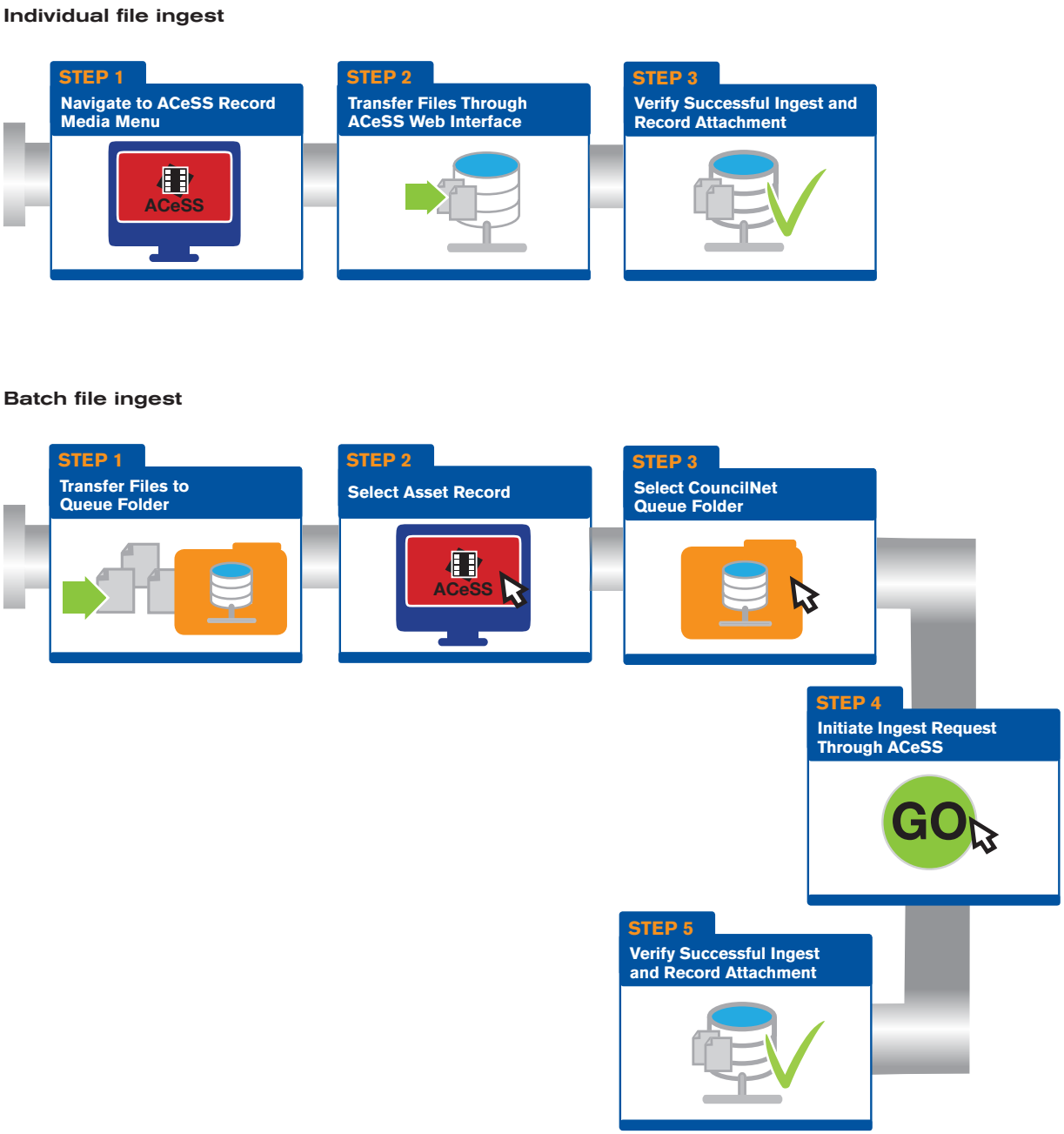
2K Mini Movie – 4.6 hours ingest processing time
4K Mini Movie – 12 hours ingest processing time
6K Mini Movie – 41.5 hours ingest processing time

ACeSS Batch Ingest **FIGURE 15**

PROCESSING THE StEM COLLECTION IN ACeSS

Figure 16 shows both the individual file and batch ingest processes of the StEM collection.

Ingest Workflows **FIGURE 16**



PROCESSING THE StEM COLLECTION IN ACeSS

Searching the StEM collection in ACeSS

While creating records is the primary operation in CollectiveAccess, having the ability to search and retrieve existing records within the system is useful for finding, modifying and updating metadata. CollectiveAccess features a modular search facility that allows one to choose from several low-level search engines such as Apache Solr, Zend _Search_ Lucene and Sphinx. ACeSS utilizes MySQL with built-in full text indexing. There are two ways an end-user can search for items in ACeSS: public search browser and within the cataloging module. CollectiveAccess also includes a faceted browse feature, which provides lists of terms, names and other structured metadata to facilitate discovery of items. There are four mechanisms to search for records in CollectiveAccess: *Find, Browse, AdvancedSearch and QuickSearch.* Figure 17 displays two of these search workflows in ACeSS.

Search Workflows **FIGURE 17**

Search using Find

Search using QuickSearch

PROCESSING THE StEM COLLECTION IN ACeSS

Storing the StEM collection in ACeSS

As stated earlier, ACeSS was designed to support various storage repositories through an abstracted repository interface called a "repository plug-in applications programming interface" (API). There are currently three options for storing the StEM collection in ACeSS:

- "Files and folders" using the ZFS file system
- iRODS distributed storage system
- Fedora Commons repository

Because some repository options have more features than others, ACeSS implements a very simple means for a repository plug-in to tell the system what its capabilities are in response to a query from the system. Of course, all plug-ins must be capable of storing media files – that is the bare minimum. In addition, a repository plug-in will support storage of serialized object metadata in one way or another, which thereby enables it to support the key cataloging features of objects, representations and relationships.

CollectiveAccess also uses the concept of "media volumes" to store files that were processed by the media processing system (e.g., motion pictures, images and documents). Each of these "virtual" volumes is mirrored to one or more storage subsystems. This allows configuration of CollectiveAccess to store media and metadata in multiple repositories. These can be multiple instances of a single repository type (e.g., three separate Fedora instances), or a mixed set (e.g., one local file system, one iRODS repository and one Fedora repository).

ZFS file system option

The default repository plug-in used by CollectiveAccess is the basic file system provided by the operating system, which in this case is ZFS. The files generated by the media processing plug-ins are put in a structure of nested directories. Files are named such that they can be easily linked to database content programmatically or by inspection.

iRODS option

iRODS support includes storage of objects, media, and relationships. As material is cataloged in CollectiveAccess, it is replicated out to a designated iRODS node or nodes. Metadata is output in a METS-like format as a file separate from media. Links between objects and representations are implemented using a flag in the iRODS metadata catalog.

PROCESSING THE StEM COLLECTION IN ACeSS

Once output to iRODS, the data is subject to all standard iRODS rules-based preservation and distribution features. All versions of media, including both original files and derivatives, are replicated to the iRODS repository. This makes it possible to serve media to web clients directly from iRODS, obviating the need for storage on a local file system, if desired.

Fedora Commons option

As with iRODS, the Fedora Commons repository option supports digital objects, media, and relationships. All media, including derivatives, is replicated to Fedora repositories. If desired, CollectiveAccess can be configured to serve media directly from Fedora via standard Fedora web-service URLs, making storage of media in a local file system unnecessary. Serialized metadata, in a METS-like format, is written to a Fedora repository as a datastream within the digital object for the CollectiveAccess collection object or media. Relationships are implemented using Fedora's native data model.

10 LESSONS LEARNED AND NOTEWORTHY OBSERVATIONS

When we started this project, we knew that we would face many challenges. As work on the project progressed, it became very apparent that preservation technologies and practices for digital motion picture materials are still in their infancy, and that commercial software offerings for this application are virtually nonexistent. Digital archiving in general is an emerging field, and even with the current level of research and engineering activity across all digital content specialties, this project raised questions that had never been asked previously. For now, film archives must start planning and preparing for digital motion picture deposits, if only as a first step toward evaluating various solutions and approaches as described in this report and elsewhere. The alternative is to risk the loss of vital digital motion picture materials that will surely arrive at their front door sooner or later.

What follows are, in the project team's opinion, some lessons learned and noteworthy observations that we hope will be useful for those film archivists prepared to engage a future that includes digital motion picture materials in their archives.

Organizing digital motion picture materials for archiving takes significant effort; the effort required grows exponentially if organization is deferred until after production completes

One of the most challenging aspects of this project was detailing the contents of the StEM collection. Although much effort was expended by DCI to document the collection prior to deposit in the Academy Film Archive, the collection's digital elements represent one of the industry's early digital mastering efforts, and there was not nearly the level of cataloging detail for the digital elements as there was for the film elements. The term "born archival" had not yet been introduced to the motion picture industry, and technical metadata that would have enabled automation efficiencies was not routinely captured at that time. An enormous number of still image files, unusual and unspecified data files, limited – if any – descriptive or technical metadata, and out-dated and obsolete content creation tools resulted in a huge effort just to understand what exactly was in the collection. The project team spent eight months analyzing and auditing the StEM digital materials, interviewing StEM production and post-production personnel, creating detailed inventories, organizing the digital materials by functional categories, building digital workflows, and inventing metadata frameworks to assist in the final archive-optimized organization.

Archive strategy development prior to production and accurate technical metadata collection during production are crucial for digital motion picture archives to preserve and maintain access to their digital deposits.

LESSONS LEARNED AND NOTEWORTHY OBSERVATIONS

Job descriptions and educational requirements for digital motion picture archive professionals do not exist

Nonprofit film archives are not equipped with the staff, technology or funding needed to manage digital motion picture materials in an archival setting. The case study project team included a technical project manager/librarian who specializes in content strategy, a film cataloger and collections specialist, a metadata librarian, engineers, software developers and information architects. With the exception of the film cataloger and collections specialist, these positions are not typically found in a film archive.

New job functions must be defined, and college degree and continuing-education programs need to be updated to adequately staff for the management of digital motion picture materials at an archive. Film archive professionals now need computer technology skills in addition to their photochemical skills to successfully manage and preserve access to digital motion pictures for an extended period of time. New and changing roles within the film archive such as technical project manager, digital curator, digital archivist, metadata cataloger, software developer, and computer network engineer must be defined and filled to meet the challenges of archiving digital motion picture materials.

These new professional roles at film archives are vital to the creation and support of the necessary tools, policies and practices for managing and preserving digital motion picture materials for the long term.

Open source software solutions may not be right for all archives

The base software platform selected for this project was open source and therefore free of any license fee or fixed support costs. The project team decided on open source rather than commercial products because open source software may be freely modified and customized without any practical restriction. This flexibility allowed the project team to rapidly adapt the selected cataloging and repository applications to best address the evolving project requirements at a relatively low cost. Commercial software vendors generally modify their products for a substantial up-front cost, and only if they see broader market opportunities for such modifications. Digital motion picture archiving is, at present, a very small and specialized market, and no commercial software vendors interviewed were willing to work within the project budget or to the project's exact specifications. An added benefit to open source software is that the features added as a result of this project are now freely available to the film archiving community.

While there are many benefits to using open source software, it is not without its own challenges, some significant. Open source software generally requires customization. This in turn requires additional consulting or in-house software developers, and technical support to successfully implement and maintain the software. Installing open source

LESSONS LEARNED AND NOTEWORTHY OBSERVATIONS

software can also present challenges not found in commercial solutions, such as developing enough familiarity with the software's functionality and internals to successfully install and operate it.

For smaller archives with limited technical resources, a contract technical support model that provides services such as installation, data migration, user interface and feature customization, staff training and post-implementation support would be required. Archives that are part of a larger organization sometimes depend on the organization's IT department for technical advice, support, and even system development. Long-term management of digital motion picture materials is fundamentally different from corporate IT functions such as email and desktop support, and therefore it should not be treated as a "standard" IT project.

Diverse file formats and lack of standards present additional technical challenges for archives

There are no industry standards for digital motion picture source and intermediate elements, which makes it very difficult to manage and store these materials on a long-term basis. The management of massive amounts of data, multiple and unique file formats (the StEM collection had seven known file types and several unknown file types), a lack of adoption of digital cataloging standards, and unclear and varying rights-management policies all add to the complexity of managing digital motion pictures in an archival setting.

Without accepted file format and metadata standards, there are no frameworks or models for hardware and software tool vendors to follow when designing digital preservation solutions. The creation and adoption of standards for technical and descriptive metadata are integral to long-term management and storage of digital motion picture materials: technical metadata enables the accurate interpretation of the content, and descriptive metadata enables organization and classification of the content, which makes it possible for users to successfully find their digital assets.

Long-term management of digital motion picture materials is expensive

Archival management and storage of digital motion picture materials requires sophisticated technologies and specially trained technical staff. Public and nonprofit film archives, in general, manage their film vaults with limited funding and resources, both technology-related and human. In such an environment, it is beyond challenging to build, staff and manage a comprehensive digital preservation program that is capable of reliably cataloging, ingesting and managing digital motion picture materials for any period of time, let alone meeting the definition of long-term preservation.

LESSONS LEARNED AND NOTEWORTHY OBSERVATIONS

The ACeSS project required over $600,000 in equipment and labor to design and build a suitable network infrastructure, adapt open source software, and develop workflows to effectively manage and store a relatively small number of digital motion picture materials. There will be ongoing costs to maintain ACeSS, and additional software development costs are expected as we gain operational experience with the system.

It takes a community

The ACeSS project presented a diverse set of challenges across a wide range of disciplines. The project would not have resulted in an actual system implementation without important contributions and input from not only the project team members, but also collaborators at the Library of Congress, CineGrid, DICE, the major studios and other organizations. Associations, consortiums and other collaborative structures that provide resource and knowledge sharing are crucial during these early days of digital motion picture preservation. The problems are complex, and no single individual or organization can be expected to solve them on its own.

APPENDIX

ACeSS Metadata Schema

WORK RECORD

- Title
- Title type
- Alternate title
- Identifier
- Source
- Work summary
- Work notes
- Work dates
- Description
- Creators & contributors
- Publisher
- Rights summary
- Related assets
- Related works
- Related names
- Related places
- Related collections
- Library of Congress Subject Headings
- Free-text keywords
- Subject
- Genre
- Temporal coverage
- Spatial coverage
- Audience Level
- Audience Rating

ASSET RECORD

- Asset title
- Alternate titles
- Identifier
- Other identifiers
- Source
- Related works
- Annotation
- Format identifier
- Creation date
- Issue date
- Specifications for digital assets
- Specifications for physical assets
- Time code type specification
- Location
- Media type
- Generation specification for assets
- File size
- Time start
- Duration
- Data rate specification
- Color specification
- Color encoding specifications
- Tracks
- Channel configuration
- Language
- Alternative modes
- Essence track
- Date available
- Object category
- Preservation level
- Object characteristics
- Object characteristics: Fixity
- Object characteristics: Format designation
- Object characteristics: Format registry
- Creating application
- Inhibitor type
- Storage
- Viewing environment
- Environment: Software
- Environment: Hardware
- Event
- Event Date Time
- Agent identifier
- Related assets
- Related works
- Related names
- Related places
- Related collections
- Media

NAMES RECORD

- Name
- Identifier
- Source
- Lifetime
- Description
- Notes
- Nationality
- Address
- Telephone/fax
- Email address
- Organization contact
- Website
- Alternate names
- Related assets
- Related works
- Related names
- Related places

APPENDIX

- Related collections
- Library of Congress Subject Headings
- Free-text keywords

PLACES RECORD
- Identifier
- Source
- Name
- Related places
- Relation
- Location in hierarchy
- Related assets
- Related works
- Related names
- Related places
- Related collections
- Alternate names
- Library of Congress Subject Headings
- Free-text keywords

COLLECTIONS RECORD
- Title
- Collection Identifier
- Access
- Status
- Institution
- Description
- Related assets
- Related works
- Related names
- Related places
- Related collections
- Alternate names
- Library of Congress Subject Headings
- Free-text keywords

STORAGE LOCATIONS RECORD
- Location in hierarchy
- Name
- Identifier
- Status
- Description
- Related assets
- Alternate names

ACQUISITIONS
- Title
- Identifier
- Assession status
- Description
- Extent
- Extent units
- Access
- Status
- Related names
- Related places
- Related collections
- Related assets
- Alternate titles

REPRESENTATIONS
- Title
- Access
- Status
- Caption
- Related names
- Related places
- Alternate names

REPRESENTATION ANNOTATIONS
- Title
- Annotation properties
- Access
- Status
- Description
- Keywords
- Related assets
- Related names
- Related places
- Alternate names

SETS
- Title
- Access
- Status
- Introduction
- Set items

SET ITEMS
- Caption
- Item description
- Is primary

END NOTES

[1] Academy of Motion Picture Arts and Sciences: Science and Technology Council, *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials* (Beverly Hills, CA: Academy of Motion Picture Arts and Sciences, 2007).

[2] *The Digital Dilemma*, 2.

[3] Digital Cinema Initiatives, LLC (DCI), *A Paper Archive of The StEM* (Digital Cinema Initiatives, 2004), 4.

[4] National Film Preservation Foundation, *The Film Preservation Guide: The Basics for Archives, Libraries, and Museums* (New York: National Film Preservation Foundation, 2004) http://www.filmpreservation.org/preservation-basics/the-film-preservation-guide, accessed July 2010.

[5] *The Film Preservation Guide,* 70.

[6] "History of the card catalog," *Library and Information Science,* wiki article, http://liswiki.org/wiki/History_of_the_card_catalog, accessed July 2010.

[7] Betty Furrie, *Understanding MARC Bibliographic: Machine-Readable Cataloging,* 8th ed. (Washington, DC : Cataloging Distribution Service, Library of Congress, 2009), http://www.loc.gov/marc/umb/um01to06.html, accessed July 2010.

[8] The term "film editor" is still used even when digital editing equipment is part of the production.

[9] TIFF – Tagged Image File Format; MOV – QuickTime Multimedia File Format; MXF – Material eXchange Format; DPX – Digital Picture Exchange; WAV – Waveform Audio File Format; TXT – File Extension for Text Files; PDF – Portable Document Format

[10] Allison Zhang and Don Gourley, *Creating Digital Collections: A Practical Guide* (Oxford: Chandos Publishing, 2009), 1-30.

[11] Richard Jantz and Michael J. Giarlo, "Digital Preservation: Architecture and Technology for Trusted Digital Repositories," *D-Lib Magazine 11,* no. 7 (June 2005), http://www.dlib.org/dlib/june05/jantz/06jantz.html, accessed July 2010.

[12] Andreas Aschenbrenner, *Long-Term Preservation of Digital Material – Building an Archive to Preserve Digital Cultural Heritage from the Internet,* (Master's Thesis, Technical University of Vienna, December 2001), http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/OAIS.html, accessed July 2010.

[13] Consultative Committee for Space Data Systems, *Reference model for an open archival information system (OAIS),* (Washington, D.C.: CCSDS, June 2002), http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html, accessed July 2010.

[14] American Library Association, "Definitions of Digital Preservation," (ALA Annual Conference, Washington, D.C., June 24, 2007), http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm, accessed July 2010.

END NOTES

[15] Marsh, Hillary, "How to Do a Content Audit," January 2005, Content Company website, http://www.contentcompany.biz/articles/content_audit.html, accessed July 2010.

[16] Zhang, *Creating Digital Collections,* 31.

[17] American Library Association, *Task Force on Metadata Summary Report,* June 1999, http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html, accessed July 2010.

[18] EC Directorate General Information Society and Media, "CEN.BT Technical Committee 372: Cinematographic Works," Cinematographic Works Standard, rev. April 13, 2010, http://www.filmstandards.org/index.php?p=cen-tf, accessed July 2010.

[19] The term "archiving" is in quotation marks because commercial systems claiming this characteristic do not meet the definition of long-term archiving for motion pictures.

[20] DICE, "IRODS:Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems," iRODS website, https://www.irods.org/index.php/IRODS:Data_Grids,_Digital_Libraries,_Persistent_Archives,_and_Real-time_Data_Systems, accessed July 2010.

[21] Fedora Commons, "Fedora Commons," http://www.fedora-commons.org, accessed July 2010.

[22] Oracle Corporation, "Oracle Solaris ZFS," http://www.oracle.com/us/products/servers-storage/storage/storage-software/031857.htm, accessed July 2010.

[23] Academy of Motion Picture Arts and Sciences, "Image Interchange Framework," Academy website, http://www.oscars.org/science-technology/council/projects/iif.html, accessed July 2010.

[24] CineGrid, "CineGrid," http://www.cinegrid.org, accessed July 2010.

[25] Seth Kaufman, "CollectiveAccess," Whirl-i-gig website, http://www.collectiveaccess.org, accessed July 2010.

[26] Fedora Commons, "Fedora Commons," http://www.fedora-commons.org, accessed July 2010.

[27] DICE, "IRODS:Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems," https://www.irods.org/index.php/IRODS:Data_Grids,_Digital_Libraries,_Persistent_Archives,_and_Real-time_Data_Systems, accessed July 2010.

[28] Oracle Corporation, "Oracle Solaris ZFS," http://www.oracle.com/us/products/servers-storage/storage/storage-software/031857.htm, accessed July 2010.

[29] Seth Kaufman, "Collective Access," Whirl-i-gig website, http://www.collectiveaccess.org, accessed July 2010.

BIBLIOGRAPHY

Academy of Motion Picture Arts and Sciences: Science and Technology Council. *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials*. Beverly Hills, California: Academy of Motion Picture Arts and Sciences, 2007.

American Library Association. "Definitions of Digital Preservation." ALA Annual Conference, Washington, D.C., June 24, 2007. http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm, accessed July 2010.

American Library Association. *Task Force on Metadata Summary Report*. American Library Association, June 1999. http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html, accessed July 2010.

Aschenbrenner, Andreas. *Long-Term Preservation of Digital Material – Building an Archive to Preserve Digital Cultural Heritage from the Internet*. Master's Thesis, Technical University of Vienna, December 2001. http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/OAIS.html, accessed July 2010.

CineGrid. "CineGrid." http://www.cinegrid.org, accessed July 2010.

Consultative Committee for Space Data Systems. *Reference model for an open archival information system (OAIS)*. Washington, D.C.: CCSDS, 2002.

DICE. "IRODS:Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems." iRODS website. https://www.irods.org/index.php/IRODS:Data_Grids,_Digital_Libraries,_Persistent_Archives,_and_Real-time_Data_Systems, accessed July 2010.

Digital Cinema Initiatives, LLC (DCI). *A Paper Archive of The StEM*. Digital Cinema Initiatives, 2004.

EC Directorate General Information Society and Media. "CEN.BT Technical Comittee 372: Cinematographic Works." Cinematographic Works Standard, rev. April 13, 2010. http://www.filmstandards.org/index.php?p=cen-tf, accessed July 2010.

Fedora Commons. "Fedora Commons." http://www.fedora-commons.org, accessed July 2010.

Furrie, Betty. *Understanding MARC Biliographic: Machine-Readable Cataloging*, 8th ed. Washington, D.C.: Cataloging and Distribution Service, Library of Congress, 2009. http://www.loc.gov/marc/umb/um01to06.html, accessed July 15, 2010.

"History of the card catalog." *Library and Information Science*, wiki article. http://liswiki.org/wiki/History_of_the_card_catalog, accessed July 15, 2010.

BIBLIOGRAPHY

Jantz, Ronald and Michael J. Giarlo. "Digital Preservation: Architecture and Technology for Trusted Digital Repositories," D-Lib Magazine 11, no. 7 (June 2005). http://www.dlib.org/dlib/june05/jantz/06jantz.html, accessed July 2010.

Kaufman, Seth. "Collective Access." Whirl-i-gig website. http://www.collectiveaccess.org, accessed July 2010.

Marsh, Hilary. "How to Do a Content Audit." January 2005. Content Company website http://www.contentcompany.biz/articles/content_audit.html, accessed July 2010.

National Film Preservation Foundation. *The Film Preservation Guide: The Basics for Archives, Libraries, and Museums*. New York: National Film Preservation Foundation, 2004. http://www.filmpreservation.org/preservation-basics/the-film-preservation-guide, accessed July 2010.

Oracle Corporation. "Oracle Solaris ZFS." http://www.oracle.com/us/products/servers-storage/storage/storage-software/031857.htm, accessed July 2010.

Yee, Martha M. *Moving Image Cataloging: How to Create and How to Use a Moving Image Catalog*. Westport, CT: Libraries Unlimited, 2007.

Zhang, Allison, and Don Gourley. *Creating Digital Collections: A Practical Guide*. Oxford: Chandos Publishing, 2009.

ACKNOWLEDGMENTS

This report would not have been possible without the support of the members of the Science and Technology Council and the support and funding provided by the Library of Congress.

The ACeSS Team

Project Director
Andrew Maltz, Director, Science and Technology Council
Academy of Motion Picture Arts and Sciences

Project Lead and Principal Investigator
Nancy Lydon Silver, Digital Archival Program Manager, Science and Technology Council
Academy of Motion Picture Arts and Sciences

Software and Hardware Design and Development
Seth Kaufman, Whirl-i-gig, Inc. – Lead Developer

Armen Filipetyan, Data Continuum
Mark Girard, Catalyst Technologies, Inc.
Catherine Lillie, Whirl-i-gig, Inc.
Maria Passarotti, Whirl-i-gig, Inc.

Content Strategy, Cataloging and Information Architecture
Karen Barcellona, Academy of Motion Picture Arts and Sciences,
Academy Film Archive – Lead Cataloger

Amber Billey, Whirl-i-gig, Inc.
Kara van Malssen, Whirl-i-gig, Inc.
Henry Wang, Digital Preservation Graduate Intern, Science and Technology Council
Academy of Motion Picture Arts and Sciences

Library of Congress Project Administration
Martha Anderson
Laura Campbell
Carl Fleishhauer

ACKNOWLEDGMENTS

We would like to thank the many individuals and organizations who participated in the ACeSS project:

Academy of Motion Picture Arts and Sciences
Andrew Bradburn
Jeffrey Brown
Brian Drischell
Alex Forsythe
Jane Glicksman
May Haduong
Fritz Herzog
Carol Krumbach
Michael Pogorzelski
Norma Vega

DCI - Digital Cinema Initiatives
Wade Hanniball
Glenn Kennel
Howard Lukk
Walt Ordway
Evans Wetmore

Jim Houston

CineGrid and the CXP Team
Many Ayromlou
Nathan Brock
Don Brutzman
Fred Davis
Tom Defanti
Jim DeRoest
Paola Grosso
Paul Hearty
Laurin Herr
Kunitake Kaneko
Joe Keefe
Ralph Koning
Michal Krsek
Shaofeng Liu
Louise Ledeen
Dana Plepys
Luc Renambot
Jurgen Schulze
Daisuke Shirai
Natalie Van Osdol
Jeff Weekley
Bing Zhu

Pacific Title Imaging
Marc Ross

NOTES

NOTES

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 activites 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