Challenges in delivering digital content in archival and production environments

Brittan Dunham

is Archive Director for a private archive housed in an active production environment, where she manages the digital archive, on-site and off-site physical media archives, and oversees media management for the production team. She is a graduate of New York University's Moving Image Archiving and Preservation programme, during which time she worked as a projectionist and archivist at Tisch School of the Arts. She interned at the Museum of Modern Art, contributing research to their Modern Women book and exhibition series, and at the University of Hawaii, where she preserved a critical collection of media relating to Hawaii's labour history.

Tel: +1 323 305 5033;
E-mail: brittanclaire@gmail.com

Candace Ming

is the Project Manager and Archivist of the South Side Home Movie Project (SSHMP). She is a graduate of New York University's Moving Image Archiving and Preservation programme and interned at the Museum of Modern Art, doing crucial research on the newly discovered 1913 footage of Bert Williams. She has a wealth of experience as an archivist and as a film projectionist. Most recently, she managed the moving image archive of the New York Police Department's Video Production Unit. She is responsible for digitising and preserving all films donated to SSHMP and also conducts community outreach and preservation workshops.

The South Side Home Movie Project, The University of Chicago, 5733 S. University Ave, Rm. 207, Chicago, IL 60637, USA
Tel: +1 773 795 3329;
E-mail: cming@uchicago.edu

Abstract   Archives both large and small face similar challenges. Indeed, digital asset management can be difficult whether managing 300 assets or 300,000. This paper examines the creation and management of digital asset management systems by two wildly different organisations, showing the need for accurate metadata, standardised workflows and automated processes no matter the size of the archive or how the content is delivered.

KEYWORDS: open source, automated workflow, production, community archives

INTRODUCTION

Archives both large and small face similar challenges. Indeed, digital asset management can be difficult whether managing 300 assets or 300,000. This paper discusses the creation and management of digital asset management systems by two wildly different organisations, showing the need for accurate metadata, standardised workflows and automated processes no matter the size of the archive or how the content is delivered.

THE SOUTH SIDE HOME MOVIE PROJECT

The South Side Home Movie Project (SSHMP) was founded in 2005 by Dr. Jacqueline Stewart, Professor of Cinema and Media Studies at the University of Chicago. The aim of the project is to collect, preserve, digitise and exhibit amateur small–gauge home movies from residents of the South Side of Chicago. The archive includes material from several sets of neighbourhoods surrounding the University of Chicago, ranging from the 1930s to the 1980s. Housed as an independent archive under the Cinema and Media Studies Department, SSHMP operates separately from the Special Collections Department and has created its own unique database and workflow system. Under the direction of the Project Manager and Archivist, SSHMP regularly engages with the community to solicit film donations in the form of screenings, workshops and other lecture-based series to both activate and illuminate the films in the archives. SSHMP deals with a high volume of material both from in-house transfers and files received from its outside digitisation vendor, the Smithsonian National Museum of African American History and Culture (NMAAHC) in Washington, DC. The workflow includes ingesting both the master files and creating several derivative versions.

Uniqueness

Home movies and other amateur films are not rigorously studied and often remain unseen in the attics, basements and homes of their creators. Exposing the public to these private family films from the South Side of Chicago often disrupts the canon of both film history and the history of events and movements. Footage of the Bud Billiken Parade, the 1964 World's Fair in New York and Resurrection City in Washington, DC provide a personal and intimate view of these events by the people that lived them. These home movies also disrupt the ideas of Black leisure in Chicago by showing African Americans both at home and abroad during times of struggle and incredible racial divide. These films have power beyond the images they show and expose how little the country has changed in the last 70 years. SSHMP's mission is to transfer and screen these unique films both for the families and also for the world at large. The format of these films (analogue film in 16-mm, 8-mm and Super-8 formats) also makes them a huge preservation priority due to the obsolescence of the equipment needed to play them and the often poor storage of the films.

Infrastructure and collaboration

At the inception of the project in 2005, the focus was more on screening and connecting residents of the South Side to one another. Several events were held where different films from different families and neighbourhoods showed their films and remarked on the similarities and met other residents with whom they might not previously have interacted. Oral histories were also conducted, and four collections were digitised and kept in the SSHMP archive. In 2015, the archive hired a Project Manager and Archivist to create and manage the digital and physical infrastructure needed to operate SSHMP as professional archive for the materials and continue collecting and digitising film.

Several problems presented themselves in the creation of both a workflow for digitisation of films and the digital storage. The digitised films were stored on two mirrored hard drives, but without standard naming conventions or organisation. Additionally, the previous catalogue was no longer accessible, so both description and physical metadata for the materials needed to be recreated. The digital files were also not in a preservation master format so the decision was made to retransfer these materials.

Capitalising on the resources of the University of Chicago, several key partners were identified to help create the infrastructure necessary to continue collecting and digitising film.

For server storage and digital backups, a crucial need, SSHMP partners with the Research Computing Center to use LTO-7 tape backups. The cost is cheap, and having this resource on campus, as opposed to using an off-site vendor or purchasing and maintaining a server, is an incredible benefit.

For physical inspection and storage of materials, SSHMP partners with the Film Studies Center, which maintains a climate-controlled vault for the collection of 16-mm and 35-mm film it uses to support classes and screenings for the Cinema and Media Studies Department. Again, this is a crucial partnership as it allows SSHMP not just to inspect films in a clean and professional environment, but also collect and oversee the long-term preservation of donor films in a professional vault space.

Finally, CollectiveAccess — an open source, web-based tool for managing digital and analogue assets — was identified as the digital asset management system most appropriate for this collection of unique and diverse materials. CollectiveAccess was chosen because of its ease of use, ability to create multiple users with varying access levels, web-based structure, native ability to annotate video and its ability to upload proxy copies directly to their corresponding records (Figure 1). CollectiveAccess also offers customisability. As home movies are more difficult to catalogue than other forms of narrative or industrial films, the customisable nature of CollectiveAccess makes it much more effective than an out-of-the-box solution. Using CollectiveAccess, it was possible to tailor the system to SSHMP's specific needs (Figure 2).

Additionally, the installation profile of CollectiveAccess features built-in metadata schemas (PREMIS and PBCORE) that may be used to standardise the data entered into the catalogue.

Working with the Humanities Division Information Technology Department, CollectiveAccess was configured to catalogue and publish digital assets. CollectiveAccess also has the ability to push content to a front-end site seamlessly. This site can be updated easily and shows only the fields required, in this case description metadata and subject/keyword tags, to the end user. Once a digital version is uploaded, it automatically appears on the site with its associated metadata.
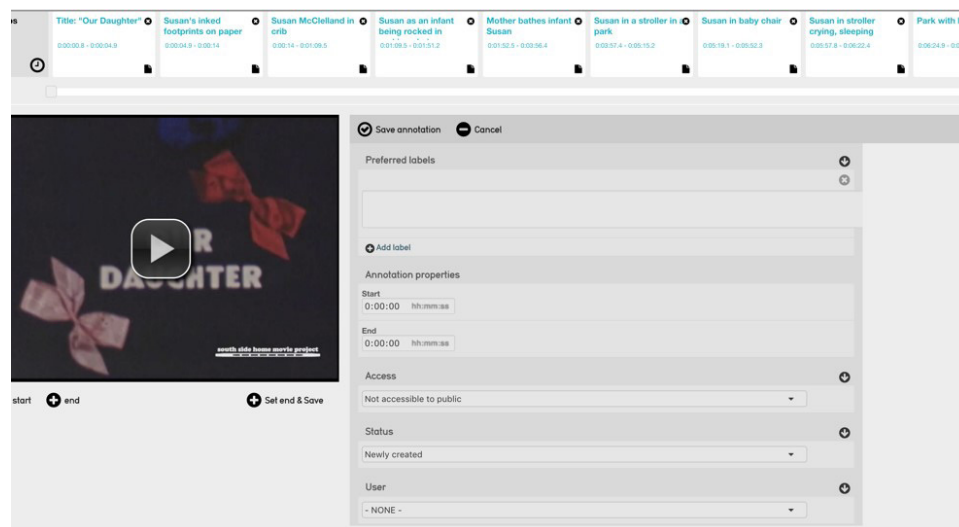

Figure 1: Screenshot from the CollectiveAccess Video Annotation page

Figure 2: Screenshot of a customised film inspection report in CollectiveAccess

In short, the fact that it was not necessary to create a separate site, combined with how easy the software was to use, made CollectiveAccess the obvious candidate for digital asset management.

Workflow

SSHMP has several workflows for digitising its moving image material. Most material can be inspected and transferred in-house using an 8-mm and Super-8 telecine transfer machine. However, a small portion of the collections include Super-8 with magnetic striped sound and 16-mm film. For in-house materials, SSHMP digitises to a lossless uncompressed file. This is later encoded into two proxy files: a clean, high-quality h264 for the film donor and a low-quality h264 to be uploaded to CollectiveAccess. For films digitised by NMAAHC, a 2K Digital Picture Exchange (DPX) is received and then encoded into three different proxies. A ProRes 444 'master' is made to produce the other two proxies listed earlier. All of these files are stored on a server partition where there is currently 8 TB of data from over 300 digitised films.

For further digital preservation purposes, master copies are uploaded to the Digital Library Development Center (DLDC), a new initiative at the university to collect and store digital content from multiple constituents across the university. Prior to this workflow, SSHMP outsourced digital preservation services to a company called Digital Bedrock. Digital Bedrock provided a high level of fixity and obsolescence checking as well as geographic dispersal of digital masters. Following the creation of the DLDC, however, SSHMP is exiting Digital Bedrock in order to keep materials in-house and allow other units within the university to explore its collections. Use of the DLDC also allows the collections to be accessed on an aggregate site called explorechicagocollections.org, which collates Chicago-based collections from across Illinois and provides finding aids so that users can access them. This seemed like an incredible opportunity for the archive to enjoy more widespread use by the community and the DLDC is able to provide fixity and obsolescence checks for materials free of charge.

Metadata management

SSHMP's version of CollectiveAccess is configured with PREMIS and PBCORE

metadata schemas to support the capture of physical item description and preservation activity metadata. PREMIS also supports location-based metadata, allowing for records to point to their digital counterparts on the server. Uploading proxy copies of media also allows for easier cataloguing and easier association between physical items and digitised copies. Although CollectiveAccess is integrated with Library of Congress Subject Headings, these headings are not sufficiently granular to accurately describe the diverse varieties of subjects included in home movies. Thus, for subject headings, SSHMP uses a combination of bespoke headings as well as keywords from Chicago Film Archives. The Chicago Film Archives houses a small sample of home movies and its subject headings also include Chicago-specific events and organisations in a standardised format, which is very useful for content. Describing and keyword tagging home movies is fraught with complications related to the accurate depiction of a non-narrative film. One film could have multiple locations, subjects and even years. The ability to customise subject headings is vital, and CollectiveAccess manages this with ease. Additionally, CollectiveAccess allows for time-based annotation of moving images, allowing users to jump to points they find interesting in a film. This is especially useful for home movies as subjects can change rapidly. Time-based annotation often provides a more accurate description of what occurs in a film as opposed to a simple descriptive summary.

Currently, several research assistants input metadata across all related fields: physical, descriptive and preservation. CollectiveAccess allows an administrative user to define specific roles based on access so research assistants cannot delete fields, add to lists and vocabularies or move or add metadata elements. However, they can create and edit collections and objects, so the need for administrative intervention in their cataloguing process is minimal. All metadata are contained within the CollectiveAccess database, but records can be easily exported by collection, object or user set into a PDF or Excel spreadsheet. CollectiveAccess also allows for the creation of finding aids which can be exported or viewed on the public site.

PRIVATE PRODUCTION ARCHIVE

In 2008, a private production company was formed to create video content for and relating to an artist and musician. In two years, the company had evolved into a full-scale management and entertainment company, with a large collection of digital and physical media that was growing constantly. It soon became necessary to preserve the collection and to access and use it for new projects. In 2013, an archivist was brought on to manage the collection and construct a digital asset management system that would serve both the Archive and Production Departments. The Archive now contains several collections relating to artists managed by the company, with digitised and born-digital media well into the petabytes. The collections are private and only used internally. Finished projects, which may include archival footage or photos, are released to the public in a tightly controlled, curated manner. However, the Archive understands the historic, cultural and political significance of its larger collections and all long-term planning for them takes into consideration their potential impact.

Uniqueness

The bulk of these collections are made up of a decade worth of behind-the-scenes footage, including scenic B-roll shot in hundreds of locations around the world. This has essentially become an in-house stock footage library for the Production Department to access and repurpose. Additionally, the company's primary collection contains the history and life's work of an influential artist and cultural icon. The practical day-to-day use of these collections, as well as their cultural

significance, provide the Archive Department with a unique challenge to meet the needs of a fast-paced production environment while keeping preservation and museum-readiness as the top priority. The Production Department also frequently uses mixed-media techniques to create a final aesthetic that blurs time and format. This adds to the complexity of archiving and preserving the film, video (and sometimes other physical assets), and born-digital components of a single project.

Infrastructure and collaboration

When the Archivist was hired in 2013, the company's assets consisted of physical media, most of which had been digitised prior to use, and born-digital assets stored on replicated external hard drives or on a storage area network (SAN) that was shared with the Production Department and backed up to Linear-Tape-Open (LTO) tape. All of these assets were in high-security cold storage but they lacked organisation, categorisation, documentation and a plan for long-term preservation. The assets were not easily accessible to staff, and all understanding of the collection's contents was left to institutional memory. At the time, the company had licensed but not begun using a commercial media asset management (MAM) system.

Following an assessment of the collection and the environment, the Archivist began to work through the individual assets and projects to organise and make sense of the collection, and to migrate as much of it as possible from drives to the SAN. A cataloguer was hired to begin collecting metadata in the existing MAM system. It was already clear that the scope of the collection had outgrown its IT infrastructure, as well as the capabilities of the existing MAM system, but because of the unpredictability of external hard drive storage, the work could not wait for these infrastructure updates to take place. As a temporary measure, any assets that would not fit on the SAN were migrated to newer, more reliable hard drives during the

organisation process, and checksums were run in order to monitor fixity. Larger digital assets were moved from the SAN to LTO-7, and an LTO-7 backup of the entire SAN was made. In addition to working through the assets themselves, the Archivist also began creating and documenting standards and naming conventions, as well as consulting data storage vendors to plan for the growth of the collection.

Ultimately, a vendor was selected and the company partnered and worked closely with them to design a system that would meet its unique needs. Extremely high-security, triplicated, geographically separated backups that meet the most current digital preservation standards were the first priority. Beyond that, it was necessary to have separate Production and Archive SANs, the latter of which could scale in size with the collection, and the ability to securely migrate assets and catalogue records from the existing infrastructure. Finally, the company needed a highly secure, scalable MAM system for internal use. While many solutions, both commercial and open-source, both off-the-shelf and custom-built, were considered and tested, no solution met every need perfectly and a compromise had to be made. The needs of the Production, Creative, Digital and Archive Departments were all considered, and the company chose a system that is proven to perform at a high level in a fast-paced media production environment, while allowing for customisation through collaboration with the developers and workflows driven by application programming interface (API).

Once the infrastructure was in place, the Archive's previous SAN was changed to the dedicated Production SAN and a much larger, scalable SAN became the central location for the Archive. All assets were finally consolidated to this new location from LTO, external hard drives and the previous SAN, and multiple offsite backups were created. This was followed by several months of testing and development

of the new MAM system by the Archive Department and the vendor before it went into production. The result of this complicated, multi-year, cross-departmental project to refresh the Archive and Production infrastructure was to finally have stability of the assets and environment, and to be at a place where the staff could begin ingesting assets into the MAM system for the company's use.

As all production archivists and digital asset managers know, the scalability of this type of infrastructure can be costly and difficult to manage. It requires open, honest communication with management, and the understanding that storage costs must be considered and built into production budgets from the beginning of the project.

Workflow

New projects may be shot on one or several different types of media, using any combination of extremely hi-res digital cameras, 8/S8-mm, 16-mm and 35-mm film cameras, and occasionally even legacy-model video and digital cameras. For the most part, all film and video footage is digitised by outside vendors before use and is overseen by the Production Department. All post-production is digital, and much of the footage within the collection is born-digital. Some smaller projects have required digitisation by the Archive. Video Home System (VHS) and Digibeta transfer and photo scanning are done in-house, while film and other video formats are outsourced to a trusted vendor. All digitised and born-digital footage is transcoded for editing.

Raw digital/digitised footage is immediately delivered to the Archive for safe-keeping, then ingested and catalogued in the MAM system. Automated workflows handle backups and fixity checks. Descriptive metadata is added manually by a cataloguer, while administrative and technical metadata are pulled from the file. Careful planning and decision-making on the part of the

Archive Department in the development stages of the new infrastructure allow much of this workflow to be automated, freeing up time and resources for Archive staff. Completed projects are media managed by editors, following the Archive's standards and conventions, and delivered to the Archive with their final output (masters, and any other final versions created for web, broadcast, international broadcast, etc). This is the general workflow for new projects being delivered from Production to Archive.

For 'legacy projects', that is, projects created before the Archive was fully-functioning and able to handle this workflow, the Archive Department has had to go back and media manage these projects to the best of its ability, in many cases lacking the institutional memory to completely understand how a particular editor or producer organised their project before standards were in place. It has been up to the Archivist to manually organise, ingest and catalogue these projects and all of their related assets in a way that makes the most sense for the individual project, and the collection as a whole. The goal is always to be able to relink the project if needed, and to have all raw footage stored separately according to the Archive's standards and conventions. Born-digital footage is stored in its camera native format, as well as Production's chosen transcode format (usually ProRes 422 HQ). Film scans are stored as DPX sequences with transcode copies, and all other analogue footage is transcoded to lossless formats, which can differ based on the source media. The Archive Department follows new advancements in technology and trends in the field to plan for any future file migration, or any new tools that could improve workflow.

Metadata management

The Archive uses a private, proprietary set of naming conventions that is based on

industry best practices, but that has proven through trial-and-error to work best for the internal needs of the company. Technical and administrative metadata are captured by the MAM, and extremely specific descriptive metadata is added manually by the Archive Department using fixed fields and values created internally. Customisation of the MAM system through collaboration with developers, and very clear, intentional creation of fields and values is crucial to the company's ability to use the collections. It is important to note that in a production environment, the knowledge an archivist has of the assets they manage and how their company or organisation uses those assets must be vocalised and advocated for when working with vendors and developers. If a schema were to be adopted simply because it worked for someone else, the company may soon find that its investment is of little practical use to its staff. Understanding what information is important to the company/ artist and meticulously documenting it, while time-consuming, has proven crucial to staff adopting the tools and embracing the Archive.

Through the process of organisation, ingest and cataloguing, it became clear that the company's assets needed to be categorised into different collections in order to grow with the company and its artists. The primary collection that existed when the Archive began in 2013 still exists and continues to grow as it was originally understood, but additional collections for other artists, physical (non–media) assets, and other company enterprises have since been created. These are distinguished through metadata fields and security/access privileges within the MAM system, and their distinction is important to the larger understanding of the Archive and planning for the future of each collection. Because the groundwork has been laid, it will be easy to add additional collections or make changes to existing collections in the future.

While the cataloguing and tagging of legacy and new assets is still an ongoing

project for the entire Archive Department, there is now a full-time staff member dedicated to making the collections fully searchable in the MAM system. Without an archivist dedicated to metadata management, the Archive Department might never get through the backload of assets, and the collections would essentially be useless to staff.

COMMON GROUND BETWEEN TWO VERY DIFFERENT ENVIRONMENTS

One might well ask what a fast-paced production archive and a smaller community-based archive have in common. The answer is that regardless of the different tools used, the primary goal of both organisations is to provide access to digital assets, whether to in-house staff or the general public. Additionally, the organisations share similar workflows, relying on automation and scripts to transcode, rename and transfer media to separate digital repositories. Simply put, many of the challenges remain the same, regardless of whether the archive uses open source or proprietary tools.

Unique materials

Both archives deal with unique materials in need of preservation. African American home movies carry with them incredible cultural and historical significance. Often overlooked and ignored in film history, African American home movies offer an incredible wealth of information on the lifestyles and movements of African Americans from the mid to late twentieth century. However, as the SSHMP archive covers the entirety of the South Side of Chicago, its home movies from white families and other minorities also show the diversity and uniqueness of this historic area, as well as how it has changed (or in some ways not changed) in the last 60–70 years. They also show the daily life of families such as the still shown in Figure 3 from the Gustina Steele Collection, showing a trip to 63rd Street Beach.

Figure 3:   Still from the Gustina Steele Collection courtesy of the South Side Home Movie Project

The private archive represents the work of an extremely influential artist and cultural figure, and the output of the company they own. While the holdings and their contents are mostly private, their impact will surely be felt in the long term. In particular, the impact that the artist and company has had on African American culture and the dominance of African American artists in mainstream culture cannot be overstated. Very few artists of this calibre are so committed to preserving their legacy in real time, or even aware of the possibilities available to them if they take control of their assets at this level. That fact alone makes this archive unique and extremely forward-thinking.

Automation

As stated earlier, both archives rely heavily on automation. SSHMP uses various scripts and watch folders to transcode media and perform batch renaming of content to standard naming conventions. CollectiveAccess also has a tool for batch uploads, making it possible to load content into a folder and have CollectiveAccess match the filename to the relevant record. The same can be done using set items, allowing for cataloguers to change fields *en masse* for information that is consistent across a collection. SSHMP uses the bag-it utility to transfer files securely from external drives to the server and also from the server to the DLDC cloud storage. The DLDC performs automated fixity checks of files and runs reports that are sent to the archive. All of these automated processes free up time to perform community engagement activities, perform inspections and digitise film and supervise various research assistants.

The private archive employs a combination of manual processes and automated scripts in order to assure quality control while working quickly and efficiently. While the option exists (and was even recommended by the IT vendor) to automate MAM ingest, the archivists felt that every project requires quality control and individual decisions about what is ingested. This is slightly more time-consuming on the front end, but allows the team to avoid errors that would ultimately consume more time and energy to fix after the fact. Once the ingest process begins, proxy creation and basic metadata capture are automated within the MAM system. Descriptive metadata is then added manually, and assets can be catalogued and tagged individually or batch updated. Archive, download and

restore workflows are all automated within the MAM system. As mentioned previously, incremental offsite backups are automated. The MAM software is highly customisable with API workflows that allow it to work with other software regularly employed by the Archive and Production departments. Future plans include integration with editing software, bar-coding software (for physical assets) and watch folders for file delivery.

Access

Both archives' access strategies rely on access for in-house users. SSHMP delivers high-quality access files to film donors via a proprietary cloud storage service utilised by the University of Chicago. The cloud storage service has different permissions levels, effectively locking folders so that they can only be accessed by the film donors and their family members as well as the Archivist. The cloud storage system allows for download, sharing and commenting based on permissions levels.

Lower-quality files are delivered to a digital archive via a batch upload and automatically pushed to the public. All files are listed as accessible unless a film donor specifically requests a home movie be removed. However, because of the personal nature of these films, the Archive does not allow direct downloads of films or the exporting of metadata from the public site. Users who wish to use films for documentary or research purposes must contact the archivist and provide a detailed description of their project and the intended use of the films.

Without going into too much detail, access to the private archive is essentially limited to the archivists and IT Director through a number of security protocols. Staff may only access assets through the MAM system, which is heavily protected through permissions placed on staff credentials, and a second layer of permissions set by the archivists through user roles within the MAM system. A user's

role in the MAM system determines what collections they are able to see, what files within those collections, or which versions of a project, and if they are able to add metadata, download proxies, or perform any other actions relating to a file. At this time, only the archivists are able to make permanent changes to an asset or record. There are no plans for public-facing access. However, as more assets are ingested into the MAM system, allowing staff to see the entire breadth of the collections, the collections are expected to be used in new creative ways. Making the contents of the archive as accessible as possible to staff has been the top priority since beginning the project, because the ultimate goal is to allow them to share more of the incredible holdings with the public.

Metadata

Although the archives collect different degrees of metadata, the level of granularity is the same. As mentioned previously, SSHMP requires a high level specificity due to the non-narrative structure of home movie films. Scenes change rapidly in home movies, and identifying people, locations, events and year ranges makes discovery and access of these films easier for the end user. Keyword and subject tagging are essential for the description of these films and the vocabulary list is constantly being updated based on both user requests and input from research assistants. For example, a user can search the archive for 'trees' and find everything on that term. Another vocabulary list employed is titled 'Camera Techniques'. This list features specific shots or scenes common in home movies. The Archive receives regular requests for different shots like 'filming from a car', and prior to the creation of this list had no way to track that type of scene. Like the subject/keyword list, the Camera Technique list was populated based on input from users and research assistants cataloguing materials and is constantly updated with new terms.

Similarly, the private archive captures very granular, descriptive metadata and depends on extensive keyword tagging. Descriptions of the contents of a shot are used because Production staff often search in much the same they would through a stock footage library. At the organisational level, even before ingest, naming conventions are used to assign each project its chronological and contextual place in the larger collection. Required metadata fields within the MAM system further describe the file itself, and any information available about how it was shot, transcoded and edited, its creator, and its role in a given project. Finally, its contents are described and it is tagged with any keywords that might help a producer looking for a specific type of shot, a person who appears in the shot, a location, or even a specific outfit, song or phrase. In both cases, the specific ways the collections will be searched and used is as important a consideration as having controlled standards and conventions.

CONCLUSION
The differing strategies used by the archives often culminate in the same goal. The accessibility of digital assets is the top priority of both archives, and whether employing open source or proprietary tools, it is imperative to accurately describe, retrieve and transfer materials across platforms. The creation of a digital/media asset management system is essential to support archives. The archives use a similar level of granularity in their metadata and both employ a heavy set of automated workflows to support the cataloguing and transfer of digital assets. Although on the surface it would seem that the archives have very little in common, this overview of policies and procedures points to the opposite.

Among archives small and large, from private archives to university libraries and even government institutions, the adoption and enforcement of procedures to control and move digital assets is vital to success in long-term preservation and access.