IJDC | *General Article*

It's How Many Terabytes?! A Case Study on Managing Large Born Digital Audio-Visual Acquisitions

Laura Uglean Jackson
University of California Irvine

Matthew McKinley
California Digital Library

Abstract

In October 2014, the University of California Irvine (UCI) Special Collections and Archives acquired a born digital collection of 2.5 terabytes – the largest born digital collection acquired by the department to date. This case study describes the challenges we encountered when applying existing archival procedures to appraise, store, and provide access to a large born digital collection. It discusses solutions when they could be found and ideas for solutions when they could not, lessons learned from the experience, and the impact on born-digital policy and procedure at UCI Libraries. Working with a team of archivists, librarians, IT, and California Digital Library (CDL) staff, we discovered issues and determined solutions that will guide our procedures for future acquisitions of large and unwieldy born digital collections.

Received 31 March 2016 ~ *Accepted* 9 November 2016

Correspondence should be addressed to Laura Uglean Jackson, Special Collections and Archives, University of California, Irvine. Email: lugleanj@uci.edu and Matthew McKinley, California Digital Library, University of California, 415 20th Street, 4th Floor, Oakland, CA 94612-2901. Email: matthew.mckinley@ucop.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v11i2.391

Introduction

In October 2014, the University of California Irvine (UCI) Special Collections and Archives acquired a born digital collection of 2.5 terabytes – the largest born digital collection acquired by the department to date. Special Collections and Archives has actively acquired born digital collections since approximately 2006, and has streamlined procedures in place for acquisition, accession, description, and preservation. However, we quickly found that these procedures, designed for comparatively smaller collections, were not scalable to meet the acquisition, access, and preservation needs for a 2.5 terabyte moving image collection.

This case study describes the acquisition of the UCI 2014 Commencement collection, and the challenges met when applying existing procedures to appraise, store, and provide access to a large born digital collection comprised mostly of uncompressed files of video content. It discusses solutions when they could be found and ideas for solutions when they could not, lessons learned from the experience, and the impact on born-digital policy and procedure at UCI Libraries. Working with a team of archivists, librarians, IT, and California Digital Library (CDL) staff, we discovered issues and determined solutions that will guide our procedures for future acquisitions of large and unwieldy born digital collections.

About the Acquisition

June 2014 marked the beginning of UCI's 50th anniversary. In June 1964, President Lyndon B. Johnson dedicated the site upon which UCI was built. To mark this occasion and jump start the 50th anniversary celebrations, President Barack Obama gave the keynote address at the 2014 commencement ceremony held at the Anaheim Angels Baseball Stadium. The significance of this event cannot be understated. The 50th anniversary would last two years (June 2014 – June 2016), and celebrate UCI's past while looking ahead toward the future. The Strategic Communications Office was at the center of the celebrations and commencement – branding the campaign, designing an interactive website, and creating an inspiring video to show during commencement. Consequently they had most of the administrative and visual material documenting the creation, management, and footage of the 2014 commencement.

The Office of Strategic Communications' responsibilities include marketing, branding, media relations, and the creation of several publications at the university level. They are the source of the university archives' largest collection of campus photographs and imagery.[1] In October 2014, Strategic Communications contacted university archives to transfer several boxes of physical material and an estimated four terabytes of digital material related to the 2014 commencement. They provided a detailed list of the material that would be transferred and if it existed in physical, digital, or both formats. This list was divided into 'Video and Film,' 'Photography,' 'Printed Collateral,' and other logical categories. Despite the large volume of digital materials, the assistant university archivist assumed that the well-organized inventory represented the entire collection. The archives lent a four terabyte hard drive, write-blocker and

1 The UCI Strategic Communications Photographs collection can be viewed at: http://www.oac.cdlib.org/findaid/ark:/13030/c8tt4pp0

USB 3.0 cable to Strategic Communications whose staff administered the transfer. Once the transfer was completed, it was determined that there was approximately 2.5 terabytes of data.

Using a write-blocker, the archivist examined the main directory. There were two folders titled 'Commencement Archive June14' and 'Vox Pop.' The 'Commencement' folder contained 55 gigabytes of material, and the 'Vox Pop' folder contained 2.4 terabytes of material. Nearly all of the files in the 'Commencement' folder matched the categories listed in the spreadsheet. This material included images and footage of commencement, planning documents, and audio and video used during the commencement. File types included PDF, MP4, WAV, JPG, and Microsoft Office documents – all widely used and readily accessible formats. However, it was quickly determined that the inventory provided by Strategic Communications did not include the material found in the 'Vox Pop' folder, and the transfer of these materials was not explicitly communicated to the Libraries. This material is the focus of this case study.

The Vox Pop folder contained a complex directory structure, file types with which the archivist was completely unfamiliar, and unintuitive file names. The directory structure ran as much as 14 levels deep (for example: G:\Vox Pop\ELEMENTS\ARCHIVE\ARCHIVE FROM UCI\UCI HOVERCRAFT VIDEO\CAMERA RAW\022614 Day 1\Day 1 - Tuesday\AVCHD (MTS)\Card2\PRIVATE 2\AVCHD\BDMV\STREAM).

File types included CPI, MTS, MXF, BIM, CR2, and over 10,000 JPG files that appeared to be duplicate images. Some files were several gigabytes in size and most could not be read by existing software on library computers. The archivist had no idea what this material was and the main contact in Strategic Communications was of no help, stating that it came directly from the videographer's computer. The archivist was under the impression that the files contained raw video of the commencement itself, yet video of the commencement existed in viewable, MP4 and MOV files in the 'Commencement' folder. Believing this was the case, she was hopeful that the material could simply be weeded. However, after discussing the situation with the University Archivist, she learned it was possible that the footage was not of the 2014 commencement ceremony, but of important footage of campus life. Thus the archivist contacted the Director of Creative and Digital Strategies within Strategic Communications, a key player in the planning of the commencement. She provided a wealth of information to understand the content and context of this folder of material.

The archivist learned that the files were not of the commencement itself but of a video shown during the commencement.[2] Vox Pop was the name of the company that produced this video. Many of the unfamiliar file types were raw camera files containing all of the footage that Vox Pop shot over several days in spring 2014. The footage included student interviews, shots of campus, and research interviews. Although some of the footage was used for the final video, much of it was not. The files were b-roll in a raw camera format, viewable only with advanced video software. While uncompressed video files tend to be large in file size, these were especially large because footage was recorded at an extremely high resolution (2000 dpi) for display on the Jumbotron at Angel Stadium. Except for the material that made it into the film, the b-roll did not exist in another format. These files also supposedly included metadata created by a member of the production crew during filming, although we have yet to identify these files. The 10,000 JPEG files were generated from a time lapse camera, which produces thousands of still images over time.

2 To view the video shown during commencement, see: https://www.youtube.com/watch?v=NFsDsKvAzj0

 doi:10.2218/ijdc.v11i2.391

About UCI's Born Digital Accessioning, Preservation and Access Procedures

At the time of acquiring this material, the process for accessioning born digital material was as follows:

Special Collections and Archives Staff

1. Describe the material in a collection level finding aid,
2. Create record for material in collection management database,
3. Assign a unique identifier and naming convention to born digital material.

Library IT Staff

1. Photograph original digital media carrier;
2. Capture disk image or ZIP package of digital media content using dedicated forensic workstation and write blocking hardware;
3. Scan for viruses;
4. Extract/generate technical metadata about digital media carrier/container as well as file-level content metadata, including MD5 checksums;
5. Ingest files and metadata into Merritt[3], a preservation repository administered by the CDL.

This process ensures that born digital material is rescued from carriers that are obsolete or failure-prone and placed in a trusted digital repository for preservation. No appraisal or file migration is performed because the immediate goal is preservation of the original material. Access to the Merritt collection containing the material remains restricted to Special Collections and Library IT staff. However, selected materials have since been downloaded, normalized to access formats, and made publicly available on UCISpace @ the Libraries[4], the UCI Libraries' institutional digital repository.

Problems and Issues

Appraising, accessioning, and providing access to the material proved difficult due to the enormous size of the collection, amount of files, and uncompressed specialized file types. The large size also complicated the work of file arrangement, metadata generation and ingest into the Merritt preservation repository.

Appraisal

Meaningful appraisal was difficult due to the lack of knowledge about the collection's contents combined with its size and confusing structure. Although the literature emphasizes the importance of a donor interview prior to acquiring born digital material,

3 Merritt repository service: http://merritt.cdlib.org
4 UCI Space @ the UCI Libraries: http://ucispace.lib.uci.edu/

this was rarely done with previous university archives materials and thus not part of standard procedures. Our ability to appraise, make available, and preserve the information was never greatly affected by the absence of a pre-acquisition donor interview. In this scenario, a full donor interview would have ameliorated many of the appraisal issues.

The collection was acquired on the assumption that it contained student interviews, shots of campus, and other imagery usually valued and used by patrons. However, we never actually saw the footage. Viewing the raw video formats was impossible without specialized video software, so the archivist could not appraise the content. Even if it was possible, there were far too many video files to effectively appraise at the item level. In addition, the files were not organized or named intuitively, preventing the archivist from assessing the footage by way of titles or descriptions.

Uncompressed Files

While the archives have traditionally asked for uncompressed files, the sheer size of this acquisition made us question: are raw video files really what we want? For born digital moving image files, Library of Congress recommends the "final production version of content rather than pre-production version" be captured; that the "original production resolution and frame rate" be captured; and the "version and file-based format that was delivered to the content distributor" be acquired (Library of Congress, 2015). In 2014 the Federal Agencies Digitization Guidelines Initiative (FADGI) born digital subgroup of the audio-visual working group also crafted recommendations for creators and archivists of born digital video based on related case histories from federal agencies. While this group recommended creating uncompressed files, they acknowledge that "none of us live in a utopian world", and since "uncompressed video files can be huge and a burden to manage and maintain […] lossy compression can be appropriate for certain projects" (Murray, 2014). In the proceedings from the colloquium 'New Skills for a Digital Era', Richard Pearce-Moses and Susan E. Davis (2006) assert that "information professionals should know when to recommend one format over another for long-term preservation".

So what is better for long-term storage: a familiar and widely-used file format, or the original, uncompressed file format? What will be lost if the file is converted, and what might be gained? According to Pearce-Moses and Davis:

> 'Ultimately, the choice of format is a calculated risk, and it's impossible to predict the future. The more one understands the ultimate goals of storage and the more techniques one is familiar with, the more options to evaluate as the best' (Pearce-Moses and Davis, 2006).

Preservation Repository Ingest and Storage

Now committed to storing the original collection in its entirety, we needed to devise the best way to actually put the content in to the Merritt repository. Storage limitations on local hardware made following existing digital preservation procedures and ingesting to Merritt via a networked drive impossible. An alternate approach was simply mailing the hard drive to CDL for Merritt ingest processing, but this came with risks that needed to be addressed, including potential loss/damage to the drive and maintaining the integrity of the content files.

 doi:10.2218/ijdc.v11i2.391

We were not certain whether Merritt's digital object model could accept and intelligibly display over 50,000 files nested within a complex hierarchy of directories. Content in Merritt is organized into collections and objects: each collection may contain one or more objects, and each object may contain one or more digital content files. The Merritt object page thus lists all content files stored within the object, along with object level metadata and Merritt-generated technical metadata files (see Figure 1). To conform to the organization of other Special Collections content in Merritt and avoid disturbing the 'original order' of the digital material by arbitrarily separating it into multiple objects, we planned to store all the digital content files within a single Merritt object. It was our understanding that Merritt did not preserve the file directory structure of ingested material, which meant that all 50,000 files would exist at the same 'level' within the object. This raised several potential issues:

1. Although we planned to capture the full path of each file as preservation metadata, reconstructing the collection directory from this metadata for future access would be both time and labor intensive.
2. After backing up the collection, we ran the free DupeGuru[5] tool to identify duplicate filenames, and found 20,000 instances of identical filenames within the digital material. Since identical filenames at the same directory level are not allowed in either our local OS environment or in Merritt, these duplicate files would need to be renamed, significantly altering an important element of the original archival content.
3. With all 50,000 files listed on a single web page, and with many of the files renamed due to Issue 2, it might be extremely difficult to locate, verify and retrieve a single file when needed from within the Merritt interface.

Access

Access was another challenge. Not only would the files need to be rendered in a different format, but the archives department was also in the middle of changing its digital repository system software from DSpace to the digital asset management platform, Nuxeo. There was a delay in putting materials into the current system, because it would need to be done again for the next system. Even if we could have put files online, we weren't sure how the new system would operate for providing access to video files. The current DSpace-based system was not ideal for displaying videos. Streaming or downloading a video of several hundred megabytes could be slow and problematic; how long would a video several gigabytes in size take to access?

5 DupeGuru: http://www.hardcoded.net/dupeguru

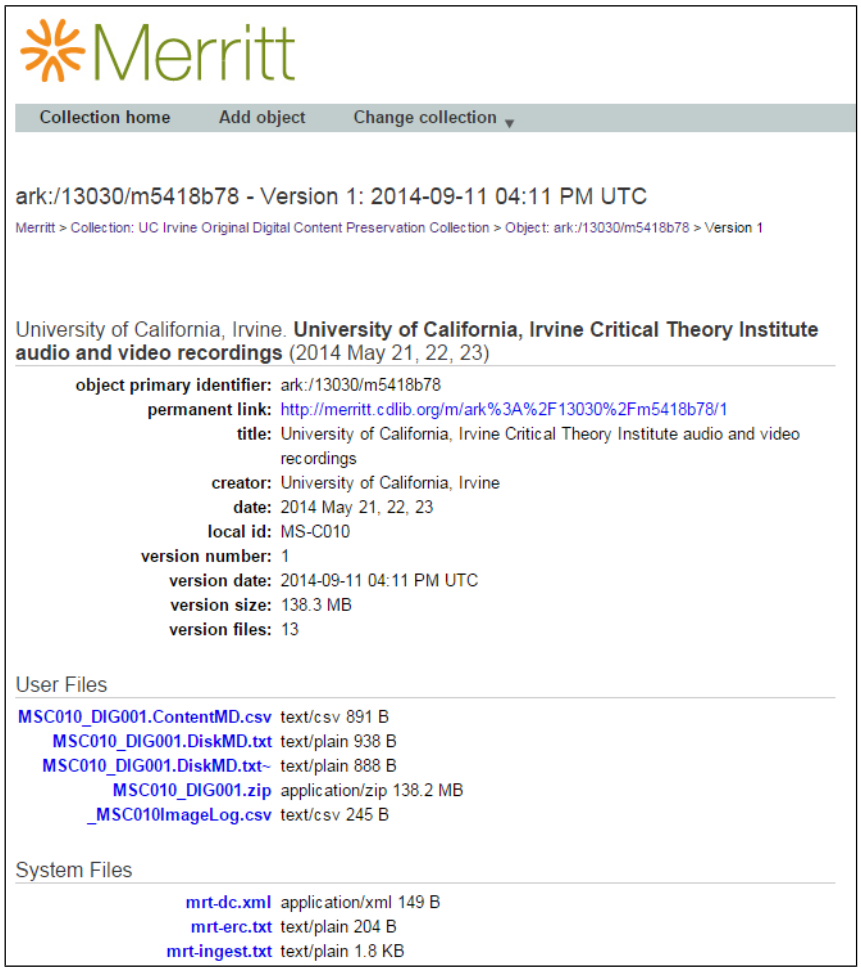

Figure 1. Merritt object page with content and metadata files.

A final consideration was that this was clearly the first of many multi-terabyte collections to come. At the time of this writing, Special Collections have several large digital collections in queue for processing, and many more identified for potential acquisition. Therefore we needed to both address the above issues and derive a workable procedure from our results so that incoming large digital collections may be processed in an efficient and coordinated way.

Solutions

As is often the case, attempting to solve these issues involved collaboration and communication. The authors shared information constantly as they explored different approaches, with each contributing specialized knowledge from their respective departments of IT and Special Collections and Archives. Reaching out to others on the UCI campus and within the UC system also helped the authors to clarify problems and arrive at solutions.

 doi:10.2218/ijdc.v11i2.391

Appraisal

We took several approaches to solving the issue of content identification and appraisal. Talking to the creator of the material was crucial to understand the organization of the folders and files. The Director of Creative and Digital Strategies spent every day with the film crew, and knew a lot about the organization and content of the files. After reviewing the directory structure with the archivist, she explained the directory's logical organization: files were first sorted by day of filming, then by camera. This brought much needed clarity for the archivist, who was unfamiliar with how a video production team might organize recordings.

Still, the sheer number of files made appraisal difficult. We found a useful tool in Karen's Directory Printer[6], freeware software that allows users to generate a text file or 'print' containing file and folder metadata, such as size, date last modified and attributes. In order to examine and sort the content by these fields, we imported the print into a Microsoft Excel spreadsheet. This way we were able to easily sort and identify the largest file size (25 GB), and note which files could be weeded. We 'weeded' duplicate versions of the final video, such as edited versions without graphics or with louder audio and many files that had originally come from Special Collections and Archives, such as images and video that was already digitized. In terms of 'weeding' we did not actually delete files but instead simply did not transfer them to the preservation repository. The print also allowed us to document the structure of the folders and files in case anything happened to original organization upon transfer into Merritt.

Uncompressed Files

The directory print was also useful for examining the most common file formats. We used PRONOM and Library of Congress format descriptions to learn more about the formats. Unfortunately, these reliable format directories, which contain information pertinent to archivists, did not include most of the file extensions for which we needed more information (such as CPI, MTS, BIM, and CR2). We discovered various websites via web search that contained some helpful information, although not nearly as full and useful as the registries created for the purposes of preservation. Even after learning more about the formats, we did not discover ways to easily view the content of the footage.

Preservation Repository Ingest and Storage

After discussion with CDL, we determined the most efficient way to get the content in to Merritt would be to mail the USB hard drive directly to CDL and have their team perform ingest procedures. In order to guard against loss or damage while in transit and to protect the integrity of the original files during initial processing, we needed to create a verifiable 'base' copy of the collection as soon as possible. With this in mind, our first step was to connect the drive via USB write blocker to a workstation and use the Bagger GUI software to back up all content files on to a second 8Tb USB hard drive, storing them in a 'Bag' conforming to the BagIt file packaging format.[7] Bagger generates a file

6 Karen's Directory Printer: http://www.majorgeeks.com/files/details/karens_directory_printer.html

7 'BagIt is a hierarchical file packaging format for the exchange of generalized digital content. A 'bag' has just enough structure to safely enclose descriptive 'tags' and a 'payload' but does not require any knowledge of the payload's internal semantics. This BagIt format should be suitable for disk-based or network-based storage and transfer.' (Source: https://wiki.ucop.edu/display/Curation/BagIt) Access the Bagger software here: http://sourceforge.net/projects/loc-xferutils/files/loc-bagger/2.1.2/

containing MD5 checksum values for all content files upon Bag creation, values which we used going forward to confirm that all content files remain unchanged from their original state.

Next, to augment the minimal technical metadata generated by Merritt, we used the File Information Tool Set (FITS)[8] to gather more information about the content files. FITS is a software package that runs an array of open source file identification tools and consolidates their output in order to identify, validate and extract technical metadata from a wide range of file formats. While FITS is a powerful and relatively easy to use tool, its default operation for every content file is to output an accompanying XML file containing FITS metadata into a specified directory. As this would both double the number of total files to preserve and cause filename collision issues due to the many identically named files, we thought it would be better to write all FITS output to a single XML file, which we achieved from the command line by writing a batch script to iterate through every file and run FITS on each, removing extraneous XML tags from the output before appending to a single FITS.xml file. This file was ingested to Merritt along with the content files, so that future users could learn more about the organization and file makeup of the collection.

Finally, after this local processing, we used Bagger to 'validate' the fixity of the content by verifying the current file checksum values against the initial back-up checksum values immediately before shipping the drive to CDL. Upon receiving the drive, CDL staff again validated the fixity of the content before ingesting into Merritt.

An email conversation with the Merritt service manager at CDL cleared up some confusion about how Merritt stores digital content at the object level: as long as the full path to each content file is included in the accompanying ingest manifest text file, Merritt will preserve the files in a multi-layered directory structure for future retrieval. This immediately solved the first two Merritt storage issues: if the entire directory structure is preserved within Merritt, we would not need to rename identical filenames or worry about reconstructing the file directory for future access. However, it's important to note that the full path is not displayed in the Merritt user interface, only the filename (this was the source of our confusion). Since the full path cannot be ascertained in the interface, retrieving a particular file would involve downloading all files sharing the filename and investigating each in the local OS file explorer until the correct file path is found. Though not an ideal retrieval process, this does not hinder the basic preservation of the content within Merritt.

Access

Providing access to this collection remains to be completed. The raw format files will need to be converted to access formats (such as MP4) both because their size prevents them from being quickly streamed or downloaded and because the raw formats are unreadable to most end users. One of the only scalable parts about the existing process was creating an accession record in our collections management database (Archivists Toolkit) and creating a collection-level finding aid describing the material.[9] However, the finer details about the acquisitions process, the meaning of the strange folder titles, and the organization of the directory did not easily fit into Archivists Toolkit. Thus we fell back on the more open-ended memo to file (both physical and electronic). Here, we documented which folders, subfolders, and files were separated; an overview detailing

8 File Information Tool Set (FITS): http://projects.iq.harvard.edu/fits

9 The finding aid for the collection can be accessed at: http://www.oac.cdlib.org/findaid/ark:/13030/c80g3nf5/

 doi:10.2218/ijdc.v11i2.391

the individuals involved with the transfer and acquisition; a general content note for each folder and some subfolders; and notes about the organization of the directory. When the material is ready for access the finding aid may need to be updated to describe the material at a deeper level. We still need to figure out the best way to accomplish this given the amount of collection material.

As mentioned earlier, at the time of writing, we were changing our online digital collections system from DSpace to UC Libraries Digital Collections (UCLDC), a browser-based system from CDL based on a modified version of Nuxeo content management software[10]. UCLDC will eventually allow for on-the-fly access format conversion, integrated metadata enhancement and 'one-click' publishing to the Calisphere access interface. Although these improvements will make preparing collection materials for access more streamlined and efficient, the large size and complex structure of the collection will still leave us with plenty of hard work and hard questions.

Lessons Learned and Impact

Appraising and processing unorganized born digital material raises unique challenges and issues not found in appraising mass amounts of physical material. It requires software, hardware, and additional subject and technical expertise just to view the contents and assess them. Processing a collection so large in size and number of files also requires creativity and improvisation – technical limitations can render best practices and recommended procedures irrelevant, forcing a new approach while still attempting to adhere to archival principles.

Appraisal

No matter how 'deep' you choose to go, assessing a collection of this size for acquisition takes time and effort. If you want this done quickly it will probably not be done at all; it is much easier to take it all than spend the time figuring out how to assess it, let alone actually assessing it. One way to assess a portion of it would be to prioritize larger files for analyses, while acquiring smaller files sight unseen – for example, focusing analysis on files 5Gb or larger while acquiring 1-2Gb files without prior review.

Also, knowing what to expect is key. Donor interviews may not be essential for smaller acquisitions, but are essential for larger digital acquisitions. Making sense of opaque filenames, complex directory structures and strange file types all require some knowledge of the creation context of the files. While this seems obvious now, blindly accepting born digital accretions set a bad precedent. Talking with the donor upfront about their digital environment and work habits will now be a priority for significant born digital collections.

Finally, document everything. Whether in a collection management database, note to file, or printed directory, explain what the material is and what you did to it as far as acquiring and preserving. This documentation may prove immensely helpful to a future archivist or researcher attempting to determine a collection's full provenance. The FADGI guidelines for file archivists contains recommended practices for what exactly to document (FADGI, 2014).

10 Calisphere: http://calisphere.cdlib.org/

Uncompressed Files

Uncompressed files aren't always necessarily the best files for preservation, despite published best practices suggesting otherwise. They can be a headache to appraise, preserve, and access. Since UCI Strategic Communications could not produce smaller/more accessible files, we accepted the uncompressed files as a matter of course. However, our experience altered our approach to subsequent born digital video acquisition. Soon after acquiring the commencement collection, Special Collections and Archives embarked on a video-oral history project dubbed the 50th Anniversary UCI Stories project[11]. A video-production unit on campus filmed the interviews and offered both the raw camera files and MP4s to the archives. The archives thus accepted only the MP4s.

Preservation Repository Ingest and Storage

We learned about the advantages and limitations of the preservation storage repository, Merritt. Limits in local storage capacity and Merritt's network ingest functionality meant we had to mail a hard drive to CDL. Additionally, it was apparent that more preservation metadata was needed for these files than Merritt generates automatically through the ingest process. We used FITS to generate preservation metadata, which we now plan to use on all collections of born digital material. You need to strike a balance between the technical limitations of your preservation and access solutions vs. maintaining the 'integrity' of a collection. An in-depth knowledge of your systems should inform how you organize and process born digital collections, but not at the expense of basic archival principles. Thus, you make compromises based on your institutional policies and resources, as well as the nature of the acquired collection, which means any resulting procedures are always a work in progress as you 'appraise' in a technical sense more born digital collections. For example, Merritt was able to preserve the commencement collection's file structure, but left much to be desired in terms of searching for and retrieving individual files. Nevertheless, we decided to sacrifice easy file retrieval in order to continue using a reliable system that was already integrated into our preservation infrastructure.

Access

Access will need to be determined after everything is in Merritt. When the accessioning process is finally finished, we can focus on how to get the content to the users. Fortunately, we should be able to use our existing access mechanisms for this, such as UCISpace and Calisphere. However, this will necessitate the creation of access files from the raw video, and possibly some curation to explain the relationship of the videos and their metadata files. It may be that even the metadata files will need to be migrated to a user-friendly format, but we won't know this until we completely understand what types of information they hold.

11 As part of the UCI's 50th Anniversary Historical Documentation project, the Special Collections and Archives implemented a video-oral history project modelled after StoryCorps.

 doi:10.2218/ijdc.v11i2.391

Conclusion

This situation has embodied the theme of UCI's 50th anniversary, 'Bright Past, Brilliant Future,' as it concerns the preservation of UCI's history, yet marks a turning point for future born digital acquisitions. This collection allowed us a glimpse into future accessions of a similar, if not larger, size. We learned that appraisal is an easy thing not to do, but a partial appraisal of formats and general content can be achieved by talking with the donor and creating simple documentation of the files.

Perhaps, like iterative processing, iterative appraisal will be the norm with born digital files. A standard practice is to migrate content from disk media without appraising the contents because required software and/or hardware is unavailable. However, the end goal is always access. Perhaps further appraisal (reappraisal) will be done when the access part of the process is ready.

Large born digital collections are inevitable, and scalable procedures are necessary if we are going to continue to preserve the history of institutions, regions, and individuals. Despite some misunderstandings, blunders, and ill-conceived assumptions, if anyone wants to project this footage on a stadium-size jumbotron in the future, they should be able to do so!

References

Federal Agencies Digitization Guidelines Initiative (FADGI) Audio-Visual Working Group. (2014). *Creating and archiving born digital video, part III: High level recommended practices.* Retrieved from http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p3_20141202.pdf?loclr=blogsig

Library of Congress. (2015). *Recommended formats statement, 2015-2016.* Retrieved from https://www.loc.gov/preservation/resources/rfs/

Murray, K. (2014). *Let's start at the very beginning: Guiding principles for creating born digital video.* The Signal Digital Preservation Blog. Retrieved from http://blogs.loc.gov/digitalpreservation/2014/02/lets-start-at-the-very-beginning-guiding-principles-for-creating-born-digital-video/?loclr=blogsig

Pearce-Moses, R., & Davis, S.E. (2006). *Knowledge and skills inventory.* New Skills for a Digital Era. Retrieved from http://www2.archivists.org/news/2008/new-skills-for-a-digital-era