
Jeffery Warda and Briana Feston-Brunet, Managing Editors
Edited by Helen Bailey, Briana Feston-Brunet, Karen Pavelka, and Jeffrey Warda

Volume Three Copyright © 2015
Electronic Media Group
American Institute for Conservation of Historic and Artistic Works
All rights reserved by the individual authors

Layout by Amber Hares
(Original design by Jon Rosenthal, JonRosenthalDesign.com)
Typeset in Trade Gothic LT and Myriad Pro

American Institute for Conservation of Historic and Artistic Works
Washington DC

The Electronic Media Review was published once every two years in print format by the Electronic Media Group (EMG), a specialty group of the American Institute for Conservation of Historic and Artistic Works (AIC), until 2013 and published online only thereafter. The Electronic Media Review is distributed as a benefit to members of EMG who held membership during the year of the issue. Additional copies or back issues are available from AIC. All correspondence concerning subscriptions, membership, back issues, and address changes should be addressed to:

American Institute for Conservation of Historic and Artistic Works
727 15th Street NW, Ste. 500
Washington, DC 20005
info@conservation-us.org
http://www.conservation-us.org

The Electronic Media Review is a non-juried publication. Papers presented at the EMG session of the AIC Annual Meeting are selected by committee based on abstracts. After presentation, authors have the opportunity to revise their papers before submitting them for publication in The Electronic Media Review. There is no further selection review of these papers. Independent submissions are published at the discretion of the EMG Publications Committee. Authors are responsible for the content and accuracy of their submissions and for the methods and materials they present. Publication in The Electronic Media Review does not constitute official statements or endorsement by the EMG or by the AIC.
A ROADMAP TO PRESERVING DIGITAL OBJECTS

DAWN AVELINE, GLORIA GONZALEZ, AND SIOBHAN HAGAN

ABSTRACT
The University of California, Los Angeles Library produces and collects a steadily growing amount and widening variety of digital objects and collections. Libraries act increasingly as stewards of archival audio and video files digitized from analog magnetic tape. Video files produced with cell phone cameras, email correspondence of professors and literary authors, social media accounts, and computer-aided design architectural files are among many other types of reborn digital and born-digital files. Along with the varied nature of these digital collections comes a diverse set of preservation risks and needs. Here we present the forces that compel an institution along the path toward increasingly robust digital preservation practices. With reference to Anne R. Kenney and Nancy Y. McGovern’s five organizational stages of digital preservation, the discussion illustrates how the varieties of digital collections, from reformatted analog videotape to activist cell phone videos to digital archival collections, can drive organizational change. Effective uses of conceptual frameworks and tools aid workflow efficiencies.

INTRODUCTION
At what point in the course of a cultural heritage institution’s development does a digital preservation program become imperative? When and how does an institution determine that typical backup and replication of digital assets is insufficient?
The following discussion illuminates the main forces coalescing to bring a formally articulated and implemented digital preservation plan into the foreground of a research library's planning activities. Each institution must design its own model for a digital preservation framework, based on its multifaceted needs—needs which arise from the types of materials in its collections, modes of access, human resources and funding. Our article examines these forces and some of the processes and tools we have used to begin to address the needs of the University of California, Los Angeles Library. The effort is ongoing and will continue to evolve. Digital preservation policies and workflows comprise something of a journey for institutions. By demonstrating the dynamic interplay between specializations and departments, this story highlights the complexity of the journey—where there is no “package tour” but where institutions can find traversable routes and expedient tools to preserving digital content.

ROADMAPS
UCLA’s path towards digital preservation can be described, with small modifications, according to Kenney and McGovern’s 2003 article, “Five Organizational Stages of Digital Preservation,” which identified five stages that an organization moves through on their way to establishing digital preservation workflows:

1. **Acknowledge**: Understanding that digital preservation is a local concern;
2. **Act**: Initiating digital preservation projects;
3. **Consolidate**: Seguing from projects to programs;
4. **Institutionalize**: Incorporating the larger environment, and
5. **Externalize**: Embracing inter-institutional collaboration and dependency.

At UCLA these five stages of developing digital preservation are interlayered with three important driving elements: collections, staff, and tools. Within this context, stage one, the acknowledgement that digital preservation is a local (i.e., intra- rather than extra-mural) concern, means that the institution, across several departments, began to recognize the increasing scale of digital content for which it was responsible, as well as the heterogeneity of the digital collections themselves. The UCLA Library produces and collects a growing variety of digital materials. We are the stewards of archival audio and video files digitized from magnetic tape; video files produced with cell phone cameras by activists; email correspondence; social media; digital architectural files; and many other types of “reborn-” and born-digital files. The heterogeneity of these digital collections brings with it a diverse set of preservation risks and needs that can only be addressed within a local context.

The next phase, or action, can only be accomplished with the help of staff armed with the expertise and bandwidth to pursue digital preservation objectives (Atkins et al. 2013; Bermès and Fauduet 2011; Nadal 2007). An examination of the growth within the UCLA Library of its overall preservation program demonstrates that appropriate staffing remains essential to establishing a digital preservation effort. The instar of preservation at the UCLA Library may be regarded as the establishment of the Library Conservation Center in 2004, launched by the hiring of a full-time collections conservator. Four years later, in 2008, UCLA hired its first full-time Preservation Officer. With the help of externally sourced funding (an indicator of stage two activities in Kenney and McGovern’s framework) UCLA Library was able to hire its first audiovisual preservation specialist (co-author Siobhan Hagan) to work within the preservation department in 2011. The audiovisual preservation specialist assembled an audiovisual lab with facilities for the care and inspection of film and the reformatting of video and audio. The arrival of this key staff member brought new capacity for preserving and reformatting audiovisual materials. Reformatting activities then introduced new digital content to the collections, and this added pressure and a sense of urgency to digital preservation concerns.
Another significant type of collection drives digital preservation efforts, albeit from a slightly different angle. Recently, the Library’s Special Collections Department has started to acquire archival materials that include digital manuscripts materials. Along with such acquisitions come new exigencies for accessioning and preservation. A prominent example is the Susan Sontag (1933–2004) papers. This collection includes a variety of media, including a hard drive containing Ms. Sontag’s emails as well as 16 mm black and white home movies. Responsible ingest and preservation planning for such digital objects in the archival setting falls under the purview of co-author Gloria Gonzalez, who joined the Special Collections Department in 2012.

**COLLECTION DRIVEN ACTION—AUDIOVISUAL MATERIALS**

To demonstrate how collections themselves drive us toward digital preservation, we take as examples two very different types of audiovisual materials, the Garry South collection and the Iranian Green Movement collection. Any moves toward the preservation of audiovisual content necessarily engender a subsequent digital preservation component. The total cost of transfers for most content, including the staff time to generate the transfers, perform quality assurance of resultant files, and create metadata as well as provide access via the digital library, means that audiovisual content draws a great deal of attention within an institution to digital preservation. Thus, collections themselves spur the institution into Kenney and McGovern’s first two phases: acknowledge and act.

Garry South was a political consultant who managed Gray Davis’s campaigns for Lt. Governor of California in 1994, and Governor in 1998 and 2002; among his papers are included many video recordings of broadcast media, news coverage and campaign ads. The Iranian Green Movement collection encompasses thousands of cell-phone videos providing coverage of the Green Movement in Iran during the contested 2009 elections, produced by activists on site and brought to the UCLA Library by activist Ali Jamshidi.1

Through the generosity of the Arcadia Fund, the Library was fortunate enough to have the resources to reformat the unique analog videocassettes from the Garry South collection. With that initial hurdle surpassed the inevitable digital preservation dilemma presented itself.

The impetus for a managed digital preservation program very often arises as a consequence of the sheer size of the audiovisual files. Roughly calculated, one hour of analog video reformatted to a digital file takes up 100 GB of space; replication of files for basic security quickly doubles or triples the necessary disk space. Reformating the Garry South collection left the library with approximately 11 TB of data and 120 digital video files to preserve.

With some types of audiovisual formats, though, these rough calculations diverge significantly. The 2,000 or so born-digital videos recorded on mobile phones during the 2009 Iranian Green Movement provide an instructive contrast (fig. 1). Resulting from the need to smuggle them out of Iran and protect the anonym
ity of their creators, the videos had been substantially compressed from their original state. In any case, a one-minute MOV file recorded on an iPhone takes up approximately 140 MB and, typically, cell phone videos run only a few minutes long.

Thus, these thousands of born-digital cell phone videos require only 10 GB of storage space leaving the library with two vastly different collections to preserve (fig. 1). The production workflows that created the videos were completely different along with their technical specifications. Therefore the preservation master format standards had to be different. It doesn’t make sense to use the exact same standards for analog audio and visual (AV) items reformatted to digital files: an extremely large preservation master file will be created that has no additional audiovisual information than the much smaller original. The UCLA Library decided to treat the ingested file as the original master, and generate a separate preservation master file utilizing the standard wrapper, but using as close to the same technical specifications in all other aspects as the original video file as possible. The library will conduct regular obsolescence monitoring and planned migration of the preservation master file specifications of this cell phone video collection.

The plan for the digital preservation of analog-to-digital reformatted AV files and born-digital AV files is in place, but resources are still limited. With already too much to do, too few people, and restricted time and money, the decided workflow can be technically complicated for many, and time consuming for all. These files and their accompanying metadata require a detailed preservation plan workflow that is as automated and efficient as possible.

Choosing your preservation master file standard is only one part of the enormous whole of digital preservation strategy. As discussed, based on the original specifications, UCLA Library creates a digital preservation master file. Then, utilizing the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation (table 1), the library assigns a level to a collection which then also helps to inform what tools and services are needed to preserve it digitally (table 1).

If the library assigns an NDSA level 4 to the Garry South digitized video collection, it needs to address everything in the previous levels as well. These various activities require various tools to complete: how should the library check fixity; how will it virus-check content; how can it batch migrate files when that time comes? There is a plethora of tools available to aid in the digital preservation of complicated AV files that are documented in several blog entries, presentations and articles online. Ideally, these tools need to be as automated as possible and need not require much translation from an IT professional, and preferably they could all be accessed in one place. The multifaceted needs of organizations undertaking digital preservation must be addressed by equally changing models that are timeless yet trendy, while accessible and practical to the already overcommitted archivist, librarian or conservator.

**COLLECTION DRIVEN ACTIONS—DIGITAL OBJECTS**

Over the past three decades, the UCLA Library Special Collections Department has seen a gradual increase in what we call “hybrid acquisitions,” collections containing both analog and digital media. More recently, new accessions now regularly include digital materials in the form of images, documents, email, and many other digital formats.

In 2012 Special Collections engaged co-author Gloria Gonzalez as the digital archivist in a contract position. Her responsibilities include in-house digital archives training and coordination with our Library information technology group to fulfill technical requirements for processing, storing, preserving, and providing access to digital files.

Until recently, Library Special Collections managed the digital media in its collections by treating it the same way as analog materials, by placing the media in boxes and placing them on shelves. The first step to address-
This survey gave us a better understanding of our digital backlogs while, at the same time, attention was focused on addressing new acquisitions, including a digital addition to the Susan Sontag papers, which proved to be a major catalyst for our progress. To accession the Sontag files, an archival ingest computer workstation was set up. The workstation’s key features include a write-blocker, which prevents the computer used to ingest digital materials from inadvertently altering any of the source files in any way. This first iteration of the workstation served as a type of “sandbox” allowing the digital archivist to explore a wide variety of tools for acquisition.
One of the main tools used is an open source software application called BitCurator (www.bitcurator.net), the result of a project led by University of North Carolina and the Maryland Institute for Humanities at the University of Maryland. This suite provides robust forensics applicable to archival practice. With BitCurator in place, perfect copies of digital media can be generated without altering any metadata. The software also allows archivists to identify the file formats present in collections, such as audio files, software, geospatial files, images, presentations, spreadsheets, databases, text documents, and videos.

In identifying relevant digital media in the collections and establishing a secure ingest workstation, the Special Collections Department was able to implement NDSA’s second level of digital preservation, at least in certain domains, in less than a year (Owens 2012). However, our preliminary workflows were time intensive when it came to migrating files into access and preservation formats. The Sontag files alone measured at only 6 GB but included over 25,000 text documents and 17,000 emails. It took a few hours to migrate about ten email databases one at a time; however, this kind of time requirement is not scalable.

In order to move forward efficiently, we needed to find a comprehensive, automated approach. One approach that is under consideration is the utilization of Archivematica, a free, open-source suite of software tools designed to maintain standards-based, long-term access to collections of digital objects (www.archivematica.org). It processes digital objects from ingest to access in compliance with the ISO standard functional model for Open Archival Information Systems, by integrating of a number of open-source tools with a “micro-services” approach. A system built using micro-services combines many small, lightweight service modules. The services are arranged in independently deployable groups and communicate with each other via a well-defined interface. Together, these micro-services multiply the system’s value (Abrams 2010). Archivematica automatically performs many processing tasks including virus checking, checksum verification and file format conversions. File migration and preservation planning are controlled using a file format policy registry. The format policy sets rules or commands for each file format type. Archivematica comes with a solid format policy that can be adjusted to meet particular institutional needs. The strengths provided by Archivematica micro-service architecture are key (Van Garderen 2010a; Van Garderen 2010b).

Loosely-coupled is an attribute of systems that refers to a modular design approach. By combining simple, self-sufficient commands, Archivematica reduces interdependencies across modules or components. This reduces the risk that changes within one module will create unanticipated changes within other modules. This approach specifically seeks to increase flexibility in adding modules, replacing modules and changing operations within individual modules. The opposite would be a tightly coupled system, which often require difficult and taxing upgrades. With Archivematica, you can replace one service at a time, instead of the entire application. This supports healthy evolution to meet user needs. Additionally, the micro-services Archivematica allows for highly scalable configurations (Van Garderen 2010a).

NEXT STAGES
To move into Kenney and McGovern’s third stage—consolidation—an institution begins to create policies, identify ongoing funding streams, and generally consolidate efforts and resources. In the UCLA context, this is evidenced by the transition to permanent the roles of AV preservation specialist and digital archivist. At the time of this writing, recruitments for these positions are underway. In addition, we are reaching across several library departments, including Preservation, the Digital Library and Information Technology, Special Collections, and Administration to establish a digital preservation working group. This group meets regularly to build inter-unit relationships, expose new digital preservation...
efforts, discuss strategies and bring new collections or developments in technology to the foreground.

We are also leveraging a web-based collaboration and project management tool, Confluence, to improve communication about projects and workflows among staff members who have a stake in the process. Reformatting activities require input from cataloging and metadata teams, conservators, curators, imaging specialists, vendors, and digital library mavens. Improving communication and building teams is helping us move from one-off projects to programs, where decisions and processes can happen automatically.

Stages four and five—institutionalize and externalize—respectively, have begun to occur as we start to establish formal relationships with campus partners for storage, such as the UCLA Institute for Digital Research and Education Data Center, or Chronopolis at the University of California, San Diego Supercomputing Center. We are also looking at partnering with external organizations to develop new models, such as the Digital Preservation Network (DPN).

Resources, not surprisingly, remain limited. At the time of this writing, the Library supports one full-time staff member with audiovisual preservation expertise, yet we somehow accomplish much more with far fewer staff members and less money. Still, our digital collections and their accompanying metadata require a detailed preservation plan workflow that is as automated and efficient as possible.

The lifecycle of managing digital collections remains familiar to librarians and archivists, in as much as the acquisition, ingest, processing, metadata, preservation, and delivery of materials correlate to analog materials. The specialized technical skills surrounding hardware and software selection, along with the myriad preservation needs of the new varieties of digital media and evolving standards in audiovisual preservation can present overwhelming challenges to any organization. While they may seem initially daunting, by employing new frameworks and software, organizations can advance along the path towards a digital preservation program. Together, librarians, archivists, curators, conservators, and other information professionals, can work together to find our way to digital sustainability.

ACKNOWLEDGEMENTS
The authors wish to acknowledge our colleagues and friends at the UCLA Library: Sharon E. Farb, Todd Grappone, Tom Hyry, and Stephen Davison, and Howard Besser at NYU, for their support and encouragement.

NOTES
1. A detailed account of the Green Movement Collection and an examination of issues surrounding its collection and dissemination is to be found in Besser et al. 2014.

REFERENCES


Dawn Aveline
Preservation Officer
UCLA Library
11020 Kinross Avenue
Los Angeles, CA 90095
(310) 794-9352
avelined@library.ucla.edu

Gloria Gonzalez
Library Strategist
Zepheira
(310) 295-8599
gloria@zepheira.com

Siobhan Hagan
AV Archivist
University of Baltimore
Langsdale Library
1415 Maryland Ave
Baltimore, MD 21202
(410) 837-4268
shagan@ubalt.edu