

A Practical Approach to Digital Preservation Planning at a Mid-Sized Academic Library

Christine S. Wiseman

Atlanta University Center Robert W. Woodruff Library, Atlanta, GA, United States
cwiseman@aucr.edu



Copyright © 2016 by Christine S. Wiseman. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Like many mid-sized academic institutions, the Atlanta University Center Robert W. Woodruff Library finds itself in need of assessment and consolidation of existing digital content management platforms as well as exploring options for long term preservation of digital assets. The AUC Woodruff Library has nearly ten years of experience developing digital services, programs, and collections that expand access to hidden primary-resource collections. Depending on the size, format, and complexity of the project, the library engages in both in-house digital conversion as well as outsourcing to vendors. In 2015 the library was the recipient of two grants that will result in significant expansion of its holdings of digitized audio and visual resources that require access and long term preservation.

As digital projects evolve into digital programs focusing on the mass digitization of entire collections, institutions are faced with ensuring long term accessibility to vast quantities of digital assets. Without increased staffing, institutions must engage in collaborative solutions and develop workflows that maximize efficiency. This presentation and paper addresses the practical approach one institution undertook to evaluate the digital content management and preservation landscape in order to inform rapid growth and expansion of its digital programs. The library formed three internal working groups; one to address content management system consolidation, one to plan for digital preservation, and a third to train staff on coding. The presentation will outline the process of beginning digital preservation planning including conducting an inventory of digital assets, evaluating digital content management and preservation systems, developing in-house technical skills, and formulating a digital preservation framework for the library that is realistic for a mid-sized institution and consistent with the Library's overall strategic plan.

Keywords: digital preservation, content management systems, digital libraries, digitization, academic libraries.

Introduction

For libraries and archives, the digital content management and preservation landscape is rapidly evolving. As digital projects evolve into digital programs focusing on the mass digitization of entire collections, institutions are faced with ensuring long term accessibility to vast quantities of digital

assets. The last decade has seen digital collections grow exponentially and institutions must meet the challenge of providing continued access as well as long term preservation to collections both physical and digital. Without increased staffing, institutions must make hard choices and develop workflows that maximize efficiency.

Established in 1982, the Atlanta University Center Robert W. Woodruff Library is unique on a number of fronts. It is an independent, non-profit academic library and research center providing information services to the world's largest consortium of Historically Black Colleges and Universities (HBCUs): Clark Atlanta University, the Interdenominational Theological Center, Morehouse College and Spelman College. The Archives Research Center's (ARC) history dates back to the establishment of the Collection in 1925 under the auspices of Atlanta University's Trevor Arnett Library "Negro Collection." The AUC Woodruff Library has nearly ten years of experience developing digital services, programs, and collections that expand access to hidden primary-resource collections. Depending on the size, format, and complexity of the project, the library engages in both in-house digital conversion as well as outsourcing to vendors. Like many mid-sized academic institutions, the Atlanta University Center Robert W. Woodruff Library (AUC Woodruff Library) finds itself in need of assessment and consolidation of existing digital content management platforms as well as exploring options for long term preservation of digital assets. In 2015 the library was the recipient of two grants that will result in significant expansion of its holdings of digitized audio and visual resources that require access and long term preservation.

This presentation addresses the practical approach one institution undertook to evaluate the digital content management and preservation landscape in order to inform rapid growth and expansion of its digital programs. The library formed two internal working groups, one to address content management system consolidation and one to address digital preservation. More recently, a library wide coding group formed to help develop in-house technical skills so staff can develop and maintain platforms and systems for preservation and display of online resources. The presentation will outline the process of beginning digital preservation planning including conducting an inventory of digital assets, evaluating digital content management and preservation systems, developing born digital workflows, and formulating a digital preservation framework for the library that is realistic for a mid-sized institution and consistent with the Library's overall strategic plan. Finally, strategies to maintain staff technical skills and suggestions for collaboration with colleagues are provided to help supplement internal efforts and resources.

Digital Initiatives

Digital initiatives at the AUC Woodruff Library date back to 2005 beginning with a five year partnership with Cornell University, SOLINET (now LYRASIS), and nine historically black colleges and universities to expand access to the founding documents of HBCUs. As a result of this partnership, the AUC Woodruff Library gained expertise in digitization standards, metadata creation, and digitization methodologies. The AUC Woodruff Library serves as the technical administrator of the HBCU Library Alliance Digital Collection (<http://hbcudigitallibrary.auctr.edu/>), a collection that has grown from approximately 9,000 images to more than 16,000 images representing the founding materials of twenty-two HBCUs. Images of materials dating from the early 1800s to the present document the role of HBCUs in the history of African-American higher education. In 2006, the Library became the custodian for the Morehouse College Martin Luther King Jr. Collection, a collection of 10,000 personal items of Dr. Martin Luther King, Jr. As a result of this custodianship, the Library has engaged in archival processing of the collection, creating item level, web-based publicly accessible inventories, and digitizing the collection. Dissemination of the collection has been achieved through the web-based inventories, scholarly forums, inclusion in curriculum and instruction, and presentations and articles to professional communities.

Since 2011, the HBCU Library Alliance has been preserving digital content in the MetaArchive Cooperative, a distributed LOCKSS (*lots of copies keeps stuff safe*) digital preservation network for the content held in the HBCU Library Alliance (HBCU LA) Digital Collection. On behalf of the

HBCU Library Alliance, The AUC Woodruff Library provides technical support and hosts the server as a preservation node on the network. In addition, the AUC Woodruff Library utilizes both CONTENTdm and Digital Commons, a hosted institutional repository for discovery and access to digital content. The *WorldCat Local* discovery tool provides access to content in both of these repositories through a “Google-like” search box on the Library’s website. Discovery of content within these repositories is also possible using search engines such as Yahoo and Google. Alternatively, a user can navigate directly to either CONTENTdm or Digital Commons to conduct individual searches, or link to digitized content in through searching the library’s ILS (OCLC’S World Share Management Services) or online archival finding aids. To further expand access, metadata from the repositories is harvested and pulled into several statewide repositories including the Digital Library of Georgia, the Georgia Knowledge Repository, the Digital Public Library of America, and the NDLTD Global ETD Search. Currently there are approximately 77,000 images available in CONTENTdm and over 4,400 publications in Digital Commons. Online digital content includes still images, manuscripts, publications, theses and dissertations, and audiovisual materials.

Within the AUC Woodruff Library, the Digital Services Department (DSD) is responsible for managing and implementing digital conversion projects, providing access to digital content, as well as library systems administration. DSD members work collaboratively with staff of the Archives Research Center since the bulk of the digital content originates from their collections; a premier archives of primary source materials on African American and African Diaspora history and culture. In 2013 the library formed the Content Management Evaluation Working Group (CMEWG) to issue recommendations regarding digital collections storage, management, and preservation technology. The overarching goal of the CMEWG was to implement a holistic approach to planning for the conversion, storage, preservation, and access of digital collections in all formats. In the process of evaluating digital content management systems, the working group considered a myriad of features. Also taken into account were the following overarching goals to provide context and focus throughout the process:

- Consolidation of digital collections
- Interoperability with existing library systems
- Digital asset management functionality for master and access files
- Long term preservation of digital assets

After conducting an extensive literature review, the working group assembled a lengthy list of functional requirements in order to compare the offerings of various systems. The list of requirements was developed from internal brainstorming coupled with examples found in the literature. Requirements were grouped by areas of functionality such as general considerations, formats supported, metadata, access and privacy and preservation features. The CMEWG looked at both vendor and open source products. Vendor products are appealing because they offer out-of-the box functionality and built in technical support. Open source solutions, while offering flexibility and customizations require substantial internal technical expertise to manage, customize and perform upgrades. Moreover, it was quickly realized that not one product or platform would satisfy all needs. In the course of evaluation, the working group noticed that the functionality of content management systems could be grouped into three broad categories: 1) presentation of digital surrogates, 2) repository management, and 3) asset preservation. The landscape is further complicated by the fact that products are rapidly changing and that an institution’s current systems architecture is critical to the decision. The final recommendations of the working group consisted of a combination of short and long term goals. Recommendations did not call for the immediate migration to a new content management system but instead directed focus on the following:

- Continue to monitor and evaluate the content management landscape
- Advance in-house expertise in open source platform development and maintenance
- Develop a digital preservation policy

- Consider future consolidation and migration of digital collections to a Fedora-Hydra repository.

Because digital collections have grown significantly since its first foray into digital project development, coupled with receiving several grants for digitization, the library's digital assets storage needs have expanded exponentially. The following criteria developed for determining the priority in which collections are earmarked for digitization.

- Evidence of heavy use collections/series. Collections for which there is well-defined anticipated uses for the materials, and evidence of past use of the materials in their original formats. Heavy use is defined by ARC staff based on usage statistics.
- Free and clear of copyright restrictions. Collections deemed as priority for scanning are either in the public domain (75 years) or have minimal copyright issues based on donor agreement or other legal restrictions that may impede access. In cases where materials are protected under copyright laws, rights and permissions for electronic use have been secured.
- Value for research. Priority is given to materials that have significant intrinsic value because they represent the unique holdings and subject strengths of the RWL library. Other high priority collections support the long term scholarly interests of the AUC community and document the history of the AUC institutions.
- Collections featured in exhibits or programs. Collections featured in physical exhibits may also be digitized in anticipation of expanded use, or for inclusion in a digital exhibit. Collection items highlighted in public programs such as lectures and presentations, or used by faculty to support instruction, are often digitized to expand access and limit handling of originals.
- Visual materials. The digitization of visual materials such as photographs, negatives, artwork, maps, and other items with illustrations is desirable. Not only does it reduce handling of the originals, but it produces digital surrogates of materials with aesthetic appeal that can garner wide attention and be used to market and promote the institution.
- Items difficult to access in their current format. Digitization of items that are rare and/or fragile can protect them from repeated handling. However, fragile items will be evaluated prior to digitization and will not be put at risk in order to digitize. Other items that are difficult to access include obscure formats and audio and visual formats that may not be accessible without digitization.
- Completion of collections that are partially digitized as part of an exhibit, project, or from user requests. It is advantageous to seek to represent complete collections/series of digitized content to provide context and present information as a coherent collection.

Digital Preservation Planning

Similar criteria will need to be applied to establishing priorities for long term preservation of digital content. In 2015 the library formed a Digital Preservation Working Group (DPWG) to address growing concern over the need to store and provide access to digital content over the long term. Moreover, the 2015-2020 Library Strategic Plan identified the development and implementation of a digital preservation, access and collection storage plan as an objective under the goal to support Technology Capacity. The DPWG includes staff from archives, digital services, information technology, and records management and is charged with working cross-departmentally to investigate, test and formulate a framework for ensuring the long term access and preservation of locally created and hosted digital content. Preliminary objectives set forth for the working group include:

- Identify key resources, best practices, and training opportunities for digital curation, preservation, and born digital management.
- Establish workflows for identifying, prioritizing and capturing born digital content and research.
- Evaluate and test tools that permit independent access to digital resources and support born digital acquisition and workflows.
- Conduct a digital assets inventory for both digitized content, born digital and digital institutional records.
- Recommend strategies and draft a digital preservation and collection storage plan for the AUC Woodruff Library.

Activities during year one included identifying key resources, best practices and training opportunities on digital preservation. DPWG members participated in relevant webinars, hands on workshops, and two members attended a week long digital preservation management workshop. One staff member is working toward earning an online certificate in XML and RDF based systems that are used to create, manage and transform digital content and metadata; skills that will undoubtedly be critical to managing a digital preservation repository.

DPWG activities are informed by following the structure set forth in the literature including the Five Organizational Stages of Digital Preservationⁱ and the NDSA Levels of Digital Preservationⁱⁱ. Key to establishing the organizational framework necessary to sustain digital preservation is developing a policy. The DPWG drafted a framework for a digital preservation policy as seeking administrative approval and buy-in is critical to implementation. Formulating a digital preservation policy framework should not be an onerous task, nor should it be lengthy. A policy is meant to be a high level document that outlines an institution's commitment and approach to digital preservation. One necessary component is to clearly articulate the policy's alignment with the Open Archival Information System (OAIS) Reference Modelⁱⁱⁱ. The OAIS Reference Model, developed by the Consultative Committee for Space Data Systems, provides a functional framework for what is required in a repository to preserve and provide access to digital content over the long term. There are five functional entities; ingest, archival storage, data management, administration, and access. An ISO standard (ISO 14721:2003), the OAIS Reference Model has been widely adopted as the standard model for developing digital preservation systems. Other key components of a digital preservation policy are sections outlining administrative responsibility, organizational viability, financial stability, technological and procedural suitability, systems security and procedural accountability. The policy draft developed by the DPWG adheres to this structure and will be evaluated and assessed as part of a digital preservation consultation to be held later on this year.

In keeping with all major resource allocation decisions, not all content can or should be preserved at the same level. Before digital preservation priorities can be established, however; one must identify and locate digital assets. In addition to digitized content, digital assets may include born digital institutional records and born digital archival collections. Like many institutions the AUC Woodruff Library has a data center with numerous servers. Digital content resides in a myriad of server shares and in some instances on external hard drives and RAID arrays. DPWG members are presently conducting a survey of digital assets to include information such as file location, file format, file size, metadata and date of creation. Fortunately, early on the library adopted standard practices for creating and storing digital content including a file storage system, a consistent file naming scheme, and metadata standards for description.

Digital preservation planning represents a continuum that necessitates ongoing institutional commitment. At the AUC Woodruff Library, we have just started our journey to ensure long term access to our digital content. Once we obtain administrative approval of the policy framework, other near term goals include:

- Develop plan for handling permanent institutional records
- Test digital preservation and web archiving tools
- Establish born digital workflows that include fixity checking upon ingest
- Set up a digital preservation workstation
- Develop a five year storage plan to include criteria for establishing priorities.

Coding Interest Group

One rapidly emerging challenge that is likely shared by small and mid-sized academic libraries is recruiting and retaining technical staff. Many of the aforementioned tasks in digital preservation and curation require some degree of technical expertise. The trend towards using open source tools necessitates that staff can operate in a Linux environment, and is comfortable with command line operations. Traditional IT departments are often reluctant to adapt and support open source applications. Moreover, the digital preservation and curation landscape is changing rapidly and staff expertise must adapt accordingly. The AUC Woodruff Library is committed to support training staff in the skills necessary to provide 21st century library services. In February 2016 librarians from different departments working on shared projects --including updating the library's website-- realized there was a shared interest among staff in expanding their skills by learning computer languages. This led to the formation of an interdepartmental Coding Interest Group focused on providing peer instruction and learning to support aspiring librarian coders.

The Coding Interest Group began by discussing what languages or skills would be best to pursue. Group leaders presented options and provided information about the demand for specific skills in the library field; suggestions included interest in HTML (along with CSS) and JavaScript. Further discussion revealed a general interest in web development—reference librarians build and maintain LibGuides, and archivists build and maintain online finding aids, content management systems, and digital exhibits. Additionally, the frequent use by all types of librarians of LibGuides, the library website's Content Management System, and other public-facing online resources suggested strongly that HTML and related technologies would be the best starting point for the group. Accordingly, the group decided to begin coding with HTML5 and CSS since the most immediate need for most of the members lay in that direction. XML and its related programming languages, XSLT and XQuery, may be particularly good candidates for the future of the group. Understanding how to create, manipulate, and transform XML documents would be highly beneficial to group members who work in the Digital Services Department and the Archives Research Center. Other types of programming skills may assist with performing office tasks such as processing data sets with formulas in a spreadsheet program like Microsoft Excel or using a dynamic programming language like Python could save hours of effort. The coding group continues to work together through online courses; a number of group members have earned a certification in HTML. Plans are forming to put newly acquired skills to practical use in solving issues with LibGuides, web based digital exhibits, and customizing style sheets for online display of finding aids.

Conclusion

The management and preservation of digital assets is far too complex and daunting to be handled in isolation. It truly takes a village to ensure long term access to our collective cultural heritage collections in digital form. A mid-sized institution may not have the resources to designate a full time staff member to oversee digital preservation activities. Therefore, institutions must look toward collaboration both internally and externally to succeed in providing sustained online access to unique digital resources. Look for opportunities to work with and learn from colleagues. In the Atlanta area two local groups meet regularly to share best practices in archival and digital preservation, the Atlanta Area Digital Archivists and the Georgia Archives Space Users Group. As a result of local collaborations, the AUC Woodruff Library staff have participated in valuable training and formed partnerships that have led to joint grant proposals. In addition, memberships in professional organizations aligned with digital preservation such as the Digital Library Federation are critical to

advancing knowledge and identifying collaborative opportunities. As the AUC Woodruff Library's digital initiatives evolve from a series of projects to a full-fledged program with policies, procedures and strategic objectives, partnerships both internal and external are critical to our success. Moving forward along the continuum of management and preservation of digital collections, it's reassuring to know that the challenges we encounter are most likely not unique, and we have a vast network of colleagues and resources from which to draw.

References

"NDSA Levels of Digital Preservation," Library of Congress, accessed May 19, 2016, <http://www.digitalpreservation.gov:8081/ndsactivities/levels.html>

Kenney, Anne R. & Nancy Y. McGovern." The Five Organizational Stages of Digital Preservation," in *Digital Libraries: A Vision for the 21st Century*, ed. Patricia Hodges et al. (Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2003). <http://dx.doi.org/10.3998/spobooks.bbv9812.0001.001>

Reference Model for an Open Archival Information System (OAIS); Consultative Committee for Space Data System, Washington, DC, 2001; 4:1. <http://public.ccsds.org/publications/archive/650x0b1s.pdf>

ⁱ Kenney, Anne R. & Nancy Y. McGovern." The Five Organizational Stages of Digital Preservation," in *Digital Libraries: A Vision for the 21st Century*, ed. Patricia Hodges et al. (Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2003). <http://dx.doi.org/10.3998/spobooks.bbv9812.0001.001>

ⁱⁱ "NDSA Levels of Digital Preservation," Library of Congress, accessed May 19, 2016, <http://www.digitalpreservation.gov:8081/ndsactivities/levels.html>

ⁱⁱⁱ Reference Model for an Open Archival Information System (OAIS); Consultative Committee for Space Data System, Washington, DC, 2001; 4:1. <http://public.ccsds.org/publications/archive/650x0b1s.pdf>