

2010

## Digitizing the Morehouse College Martin Luther King, Jr. Collection

Courtney Chartier

*Atlanta University Center Robert W. Woodruff Library*

Follow this and additional works at: <https://aquila.usm.edu/theprimarysource>



Part of the [Archival Science Commons](#)

---

### Recommended Citation

Chartier, Courtney (2010) "Digitizing the Morehouse College Martin Luther King, Jr. Collection," *The Primary Source*: Vol. 29 : Iss. 1 , Article 3.

DOI: 10.18785/ps.2901.03

Available at: <https://aquila.usm.edu/theprimarysource/vol29/iss1/3>

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in The Primary Source by an authorized editor of The Aquila Digital Community. For more information, please contact [Joshua.Cromwell@usm.edu](mailto:Joshua.Cromwell@usm.edu).

---

# Digitizing the Morehouse College Martin Luther King, Jr. Collection

Courtney Chartier, Atlanta University Center Robert W. Woodruff Library

---

In June 2006, the papers of Martin Luther King, Jr. were being prepared for auction at Sotheby's.[1] A group of prominent Atlantans learned of the sale, and pooled their resources in an effort to buy the documents as one collection, and return them to Atlanta. The collection was then transferred to Dr. King's alma mater, Morehouse College. As the official library of Morehouse, the Robert W. Woodruff Library was granted custodianship of the collection.[2]

The Morehouse College Martin Luther King Jr. Collection (MCMLK) is comprised of over 10,000 documents (approximately 43 linear feet) and 1,100 books from Dr. King's personal library. Materials span Dr. King's lifetime, but focus primarily on his work as a minister, President of the Southern Christian Leadership Conference, and a Civil Rights leader. There is very little family or personal material. The research value of the Collection is vast, ranging from such treasures as the handwritten drafts of Dr. King's Nobel Peace Prize acceptance speech and lecture, to everyday items such as register tape from unidentified businesses. The Collection also contains a large number of handwritten sermons by Dr. King, which are previously unpublished outside of an editorial context.

Early in the processing of the MCMLK Collection executive staff decided to reformat the entirety of the manuscript collection to ensure preservation and the security of the materials. Reformatting would allow the Archives to not circulate originals, protecting them from theft and unintentional damage. Digitization also would allow staff to recreate and manipulate iconic pieces for temporary display or reproduction requests with no further handling of the manuscripts.

Due to the volume of the collection, the Woodruff Library contracted with NA Publishing, Inc., to scan the documents on site. An overhead scanner was employed by NA Publishing staff, who visited Woodruff Library several times over nearly two years, completing the scanning in batches of approximately 4,000 images. Images created onsite were saved to an external hard drive, and returned to NA Publishing headquarters with the technician. There, the images were cropped and oriented using automated software.

The MCMLK Collection was processed on the item level and described in a traditional Encoded Archival Description (EAD) finding aid, with each series represented in its own XML file.[3] The scanning was concurrent to the processing. As the two archivists completed a series, or the majority of a series, each item was numbered to correspond to the series, subseries and folder number. For example, Series 1, Subseries 1, Item 1 was numbered 1.1.0.10, with the final number in increments of ten. These numbers were used by the scanning technician to name the raw TIFF files as they were created and inserted into the item level description in the EAD by archivists.

The file numbers were created not only as a method of tracking the scans, but also as a unifying field for matching the images to appropriate metadata. With the technical assistance of staff at the Digital Library of Georgia (DLG)[4], metadata was automatically harvested from each item description into a spreadsheet. The tab delimited format allowed for automated matching between

metadata and images in a content management system.

Before the MCMLK project began, Woodruff Library had already implemented use of CONTENTdm as a partner in the Historically Black Colleges and Universities Library Alliance's digital collection collaborative.[5] CONTENTdm is collection management software produced by OCLC that handles the storage, search and display of digital collections. Based on positive experiences with this previous implementation, a second CONTENTdm instance was purchased for the management, display and search/browse functions of the MCMLK Digital Collection.

Once the TIFF images were returned and the metadata had been harvested, a Woodruff Library Digital Services Technician converted the images to JPEGs, then simultaneously ingested the metadata file and the scans. CONTENTdm was set to automatically match the numbered scan files (compound objects) with the numbers embedded into each digital object unit in the metadata.

Once the images and metadata were in CONTENTdm, the archivists would start a lengthy quality control process, matching the images to the physical items, and the metadata to the finding aid. If a compound object had technical problems, it was either reported to the Digital Services Technician for editing, or marked for rescanning at a later date. Common technical problems included illegibility, incorrect orientation, missing images, and corrupt JPEGs. If the metadata was incorrect, it was edited by the archivist.

After quality control was complete, the archivists created links to the objects in CONTENTdm, then pasted the links into <dao> units in the EAD. Each series page was then converted to HTML, using a stylesheet written and customized by DLG staff. The archivists then conducted a second quality control process, to ensure that the links were unbroken and connected to the correct images.

In January 2009, approximately one half of the Collection was opened for research. This initial release included portions of the collection that were considered to be of most interest to the public, including Dr. King's writings. Over the next year, the processing, scanning and quality control process continued on the remaining series in the Collection. The Collection was completed and the remaining series opened to the public in January 2010.

Once the MCMLK Collection was open for research, the finding aid was made available online, using a simple Google-style search box. Once a researcher was in the reading room of the ARC, they could browse or search a version of the finding aid with links to digital images within the item description. Users could also browse or search the digital collection, using the simple and advanced search options built into CONTENTdm.

The next goal for the public web portal was to institute a more refined search mechanism. The archivists, in consultation with Woodruff Library IT and programming staff, decided to implement eXtensible Text Framework (XTF)[6] as the search architecture for the MCMLK Collection. Particularly attractive were the simple and advanced search functions for structured data, as well as XTF's ability to display browse pages and search results in an easily customizable format.[7]

Since the MCMLK Collection was opened for research, it has become the most used collection in the ARC holdings. Due to the unique nature of the collection's ownership[8], the digital surrogates are accessible only from workstations in the reading room of ARC. The book collection, as well as portions of the manuscripts that were not reformatted, are available for researcher use in the reading room.

Despite these limitations, the collection is frequently accessed, particularly by students and faculty

of the AUC schools. Morehouse College staffs an office associated with programming for the collection, with whom the staff of ARC have collaborated on class curriculum, possible research projects for independent study and fielding research requests from visiting scholars.

Another challenge in providing research access is in handling patrons who want to “see” the Collection. When the papers first came to Atlanta, they were put on display at the Atlanta History Center for several months. There is still a perception that the collection is a museum piece, and not a research collection, though selections are not on permanent display.[9]

In retrospect, there were several decisions made regarding digitization that hampered the process. As the physical processing progressed, the archivists continually found items that had been sorted into the wrong series. Frequently those series had already been scanned. The number naming system for the scans allowed new numbers to be inserted, but archivists had to consistently track additions, create their metadata separately, and pull them separately for each scanning visit.

Equally problematic were items that had to be rescanned due to technical problems. Frequently these items could not be fixed by the Digital Services Technician at Woodruff Library, and the original items had to be pulled and rescanned on the next scanning trip. After the first delivery of additions and rescans to already scanned series, the staff learned that the process of correcting compound objects was complex and time consuming, requiring staff to completely reload and index a series within the content management system every time new images arrived. Loading rescans also required new links, as the original link disappeared when the compound object was changed.

These issues led the staff to streamline the process of quality control. In the first few rounds of scanning, archivists conducted quality control checking of the scans after they were ingested into CONTENTdm. In order for Woodruff Library staff to make changes to any images, the Collection then had to be reloaded into the system. Once this was realized, the staff shifted the quality control to fall after the JPEG images had been created, but before they were ingested into CONTENTdm, so that editing changes could be made. The Collection then was not loaded until after all rescans and additions were created, added and checked. This cut down significantly on the time needed to create a single series within the digital collection.

The Morehouse College Martin Luther King, Jr. Collection was the first whole scale digitization effort of the Archives Research Center. Despite the missteps along the way, it was ultimately an important teaching case, not only due to the problems, but also to the implementation of new technology. Digitization will allow for the continued preservation of an essential resource to Civil Rights Movement, and American, history.

---

[1] For more information on the sale and transfer of the Collection, <http://www.npr.org/templates/story/story.php?storyId=5618143>

[2] The Woodruff Library serves four historically black colleges and universities: Clark-Atlanta University, Interdenominational Theological Center, Morehouse College and Spelman College.

[3] Archivists used XMetal editing software and an EAD template from the EAD Official Site. Initially, archivists were going to use the Archivist’s Toolkit (AT) open source software to generate the EAD encoded finding aid, the first such finding aid published by the Archives Research Center. However, the instance of AT used was not local, but maintained at Boston University. The connection would consistently drop, and all work would be lost. Based on this experience, as well

as other quirks of the AT, staff made the decision to move to a traditional EAD creation method. ARC continues to use AT.

[4] Specifically, Toby Graham, the Head of the Digital Library of Georgia, shared a script that automatically generated tab delimited metadata from EAD for ingest into CONTENTdm using standard Dublin Core metadata fields. The steps in the process included evaluating the EAD files to see at what <container> level the digital objects existed. EAD files with digital objects at multiple <container> levels had to be processed sequentially. Depending on those levels, an appropriate XSLT file created by the DLG was used to extract the metadata into an XML file using simple MSXSL commands. The file could then be edited in a text editor, then imported into MS Excel as an XML file, saved as text. The file was then ready for ingest into CONTENTdm. These types of services fall into the mission of the Digital Library of Georgia as a collaborative center for support of digital collections and library resources in Georgia.<http://dlg.galileo.usg.edu/?Welcome>

[5] The product of this collaborative is the digital exhibition “A Digital Collection Celebrating the Founding of the Historically Black College and University.”<http://hbcudigitallibrary.auctr.edu/>

[6] XTF is an open source product of the California Digital Library, specifically designed for searching across archival collections, particularly encoded documents.<http://www.cdlib.org/services/publishing/tools/xtf/>

[7] XTF is now implemented for all of ARC’s EAD encoded finding aids.

[8] Morehouse College is the owner of the collection, and Woodruff Library is the guardian and provider of research access. However, copyright is still maintained by the King Estate.

[9] The Center for Civil and Human Rights in Atlanta holds exhibition rights to the collection. The Center is not yet constructed. <http://www.cchrpartnership.org/> The Archives does plan to rotate image reproductions in the near future.

---

**Courtney Chartier** is currently the Assistant Head of the Archives Research Center of the Atlanta University Center Robert W. Woodruff Library. She has worked previously as the Processing Archivist for the Morehouse College Martin Luther King, Jr. Collection and the Voter Education Project Organizational Records. As a student she worked as an Archives Assistant at the Dolph Briscoe Center for American History (UT-Austin) and at the Archives and Special Collections of the University of Mississippi. She holds a BA in American Studies (UT-Austin), a MA in Southern Studies (Mississippi) and a MS in Information Studies (UT-Austin). She is currently the Senior Co-Chair of SAA’s Archivists and Archives of Color Roundtable, and the Chair of SGA’s Outreach Committee.

---