Digitization in the Real World

Lessons Learned from Small and Medium-Sized Digitization Projects

Edited by
Kwong Bor Ng & Jason Kucsma

Metropolitan New York Library Council

Published in the United States of America by
Metropolitan New York Library Council
57 East 11th Street, 4th floor
New York, NY 10003-4605
p: (212) 228-2320 f: (212) 228-2598

Web site: http://www.metro.org

ISBN: 978-0-615-379998-2

Cover Design: Jason Kucsma *(illustration by Smartone Design, licensed via iStockphoto.com)*

Reviewers Committee: Mark F. Anderson, Jill Annitto, Anna Craft, Jody DeRidder, Renate Evers, Wei Fang, Maureen M. Knapp, Sue Kunda, Mandy Mastrovita, Ken Middleton, Emily Pfotenhauer, Mark Phillipson, Alice Platt, Mary Z. Rose, Stacy Schiff, Jennifer Weintraub, Andrew Weiss.

From Confusion and Chaos to Clarity and Hope: Reorganization of Work Flows, Processes, and Delivery for Digital Libraries

Jody L. DeRidder (The University of Alabama Libraries)

Abstract

Digitization support within an institution may be fractured across several departments, only partially funded, and may suffer restraints imposed by delivery software which seriously hamper progress. Most digitization is undertaken with little thought for the future; the result is digital file chaos and confusion. Without clarification of file identities and relationships, preservation and migration to new systems are seriously hampered. Additionally, low funding for archival staff may preclude the creation of valuable item-level metadata. The University of Alabama Libraries leveraged the expertise available across the library to build a cross-departmental collaboration with which to face our challenges, recognizing that obstacles become opportunities for creative solutions. We are involved in a series of pilot projects to explore how to address the gap in archivist staffing to create item-level metadata. This chapter shares our discoveries and solutions.

Keywords: Cross-departmental collaboration, Digitization, Digital file organization, Metadata creation, Open source software.

The first few years of most digital library initiatives are marked by 'boutique' collection development, in which the standards,

organization, methodology, metadata, file names, and consistency vary considerably. At the time of my arrival at the University of Alabama in mid-2008 as head of the new Digital Services department, over thirty such digitization projects had been completed. Each collection had its own file-naming system and metadata fields, with inconsistencies throughout; nothing was standardized. Metadata in the delivery software did not retain in any predictable fashion a reference to the related archival files, and could not be exported in full. Digital Services staffing was minimal, requiring time from the Cataloging and Metadata Services for subject headings and upload, from archival staff for preparation of content and descriptions, and from Web Services to manage the interface and software support.

The scope of the task ahead was to expand heavily on the scanning staff and equipment, develop a feasible set of systematic work flows for supporting a large increase in scanning, build a cross-departmental team capable of supporting digital library development, and to create an organized and reusable set of digital content that is not dependent upon resident knowledge for continuation or restoration. Challenges included a simultaneous reduction in archivist work hours, minimal space for expansion, difficult relations between some departments, and insufficient time available from Web Services.

As in many smaller organizations, our digitization effort is tremendously dependent upon cross-departmental collaboration. Programming assistance and web delivery, metadata services, archivist expertise and a regular influx of well-chosen content are all critical to the development of our online research collections. A previous gift from EBSCO Industries to the libraries supports digitization and the development of technical infrastructure, but it does not support the processing, arranging, and description of archival collections. Our need for content creates a demand on the archivists that they simply do not have the resources to meet.

Recognizing the need for improved cross-departmental communications and teamwork, our dean (L. A. Pitschmann, personal communication, August 25, 2008) called together lead representatives (including two associate deans) from Library Technology, Web

Services, Collection Development, Cataloging and Metadata Services, Special Collections and Archives, and Digital Services, to form an ongoing Digital Programs group which would meet regularly to hash out problems, develop alternatives, research opportunities and assign priorities. The creation of this group was a stroke of brilliance. By forming this framework for participation, setting forth a strategic goal and providing clear administrative support, our dean laid the groundwork for success. Given our multiple operational and relational challenges, we could only succeed by seeking solutions with the assistance of all impacted parties.

Against this backdrop we are working through four major problems: digital file chaos, the inability to reunite metadata with the archival content, software restrictions on the number of collections, and a lack of archivist time to create item-level metadata.

Problems and Solutions

Digital File Chaos

Managed, efficient production and expanded growth of collections requires standardization, not only of work flows and procedures, but also of storage and file naming conventions. Delivery systems become outmoded; migration into replacement software requires consistency of legacy content. Consistency of file names and storage patterns can also support cost-saving automation.

As mentioned, we already had over thirty digitized collections, each with different file naming systems and metadata in various states of disarray across a completely disorganized file system. What little documentation existed was scattered. Collections had been digitized for years with no road map, and with no concern beyond getting the content onto the web. We needed a clear methodology for file organization.

I consulted with the archivists to gain a greater awareness of the scope of current and future digital content. After much debate, we determined that it was most important for us to store content in such a way that we could retrace the material to the archival analog

collection. Digital collections are transitory and overlapping by their very nature; we decided that the perception of a digital collection must be determined by metadata, not by origin (searching on a shared value can retrieve all components of a digital collection). Together we developed a file naming scheme to encompass all our holdings and projected digitization plans for the next few years. We created hierarchical levels of organization: first by holder, then by collection within those holdings, then by item within that collection, and finally by sequence for delivery (Figure ALAB-1). Each hierarchical level is concatenated with an underscore in the file name, so that provenance and location, as well as sequence for delivery, are automatable and clear.

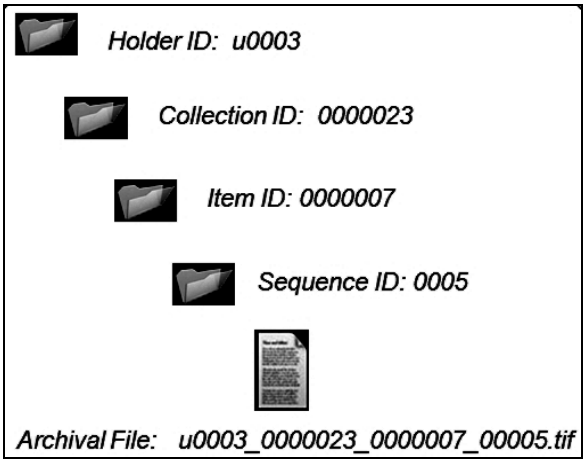


Figure ALAB-1. [University of Alabama Libraries Digital File Naming Scheme. (©2009, University of Alabama Libraries. Used with permission.)]

A "holder" is construed loosely, as we could not incorporate all the variant hierarchies of our organization feasibly into a file name. Supporting metadata clarifies identity and relationships. It was important for us to prefix each identifier with a letter so it can be used as an ID attribute within XML files (W3C, 2005). The n groupings are non-profit agencies, the p is for patron holdings, and u is for content from holdings areas within our university. We grouped content by format, as each format requires different handling and metadata description. For example, u0001 is the still image collections from the

Hoole Special Collections; u0002 is rare books, u0003 is manuscript collections, and so forth. Any of these holdings areas may have multiple collections.

For the collection number, which composes the second segment of an identifier (concatenated with an underscore separator), we echoed the existing collection numbering system whenever possible. Thus, MS 1980 will be the collection number ending in 1980, for example.

As an example, u0003_0000252 is the identifier for the digitized manuscript content of the collection known to our archivists as MS 252. Items and their subsidiary pages, however, are numbered sequentially. Item numbers follow the collection number, again, concatenated after an underscore. For example, the fifth digitized letter in the MS 252 collection would be identified by u0003_0000252_0000005. If it's a multi-page item, there will be a fourth set of numbers here, one per page, to identify the sequence for delivery. The image for page 4 of letter 5 would be identified by u0003_0000252_0000005_0004.

Thus, every part of every object has a defining and unique identifier which relates it to every other file in some fashion. We struggled with all the different anomalies we currently had and could foresee, simplifying this identification system as much as possible while still retaining the generic flexibility to apply it to all of our digitized content, regardless of the form or treatment. We then expended a great deal of effort to bring all legacy content into the new file organization scheme, gathering what information was still available to make sense of the chaos of files that remained from previous projects. Because of this system of file naming and organization, we are now able to automate much of our quality control. We have scripts that locate missing sequences, incorrectly named files, or files in the wrong place. We even have scripts to repair file names when large quantities of file names are in error. This has already saved us hundreds of hours.

Given that each level of the organization scheme contains or relates information which applies to each level below it, our associate

dean (Thomas C. Wilson) decided it made sense to echo this file naming scheme in the directory organization for storing files. Everything from a particular holder should be together in a directory named for that holder. Everything from a particular collection should be in a directory named for that collection. Thus, in our storage directories, all of the Hoole manuscript collections will be within the same directory (u0003). This is an intuitive use of the file directories to provide clarity and simplicity of organization.

For example, within the u0003 directory is a subdirectory 0000159, which contains information about all digitized content from MS 159. Within this collection directory exists a subdirectory for each item, named for the item number. The storage directories thus echo the file naming scheme, providing clear, simple, automatable organization. Drilling down through the file system to the logical depth locates the digitized archival file (Figure ALBA-2).

Metadata and documentation are stored at the levels to which they apply. Thus, metadata about the collection is in the Metadata folder at the collection level, metadata about an item at the item level, and metadata about a specific page is at the page level. Each sub-file inherits the information available at the levels above it. Thus, provenance documentation added at the collection or holder level clearly applies to all files in the directories below it. If some information only applies to page 4 of a letter, it is stored in that file's directory.

An organizational patterning such as this (Figure ALAB-3) retains the item structure, both physically and nominally, through the file identifiers. The simplicity, systematic numbering, sequencing, and clear documentation stored at the level applicable makes the digital content resurrectable for future delivery systems, without complex metadata schemes or database dependence. The organizational scheme is built to be scalable and extensible enough to manage digital content into the foreseeable future. In addition, because the directory structure echoes the file names, we were able to automate the storage of content and the creation of attendant LOCKSS (LOCKSS, 2008) manifests.

Figure ALAB-2. University of Alabama Libraries LOCKSS Content Organization. (© 2009, University of Alabama Libraries. Used with permission.)

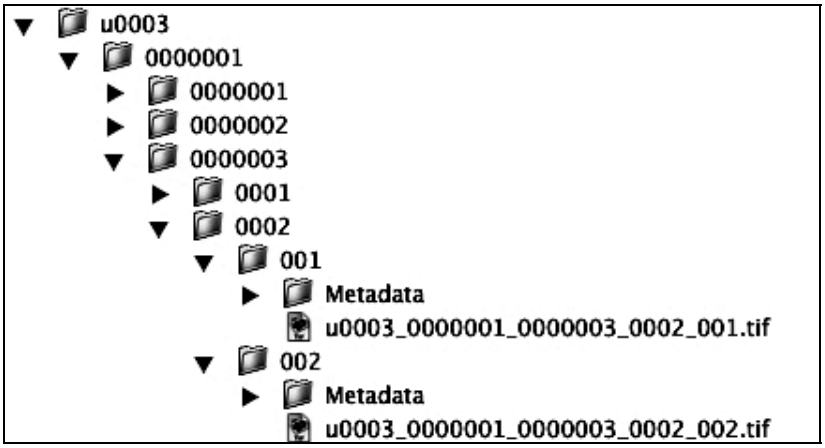


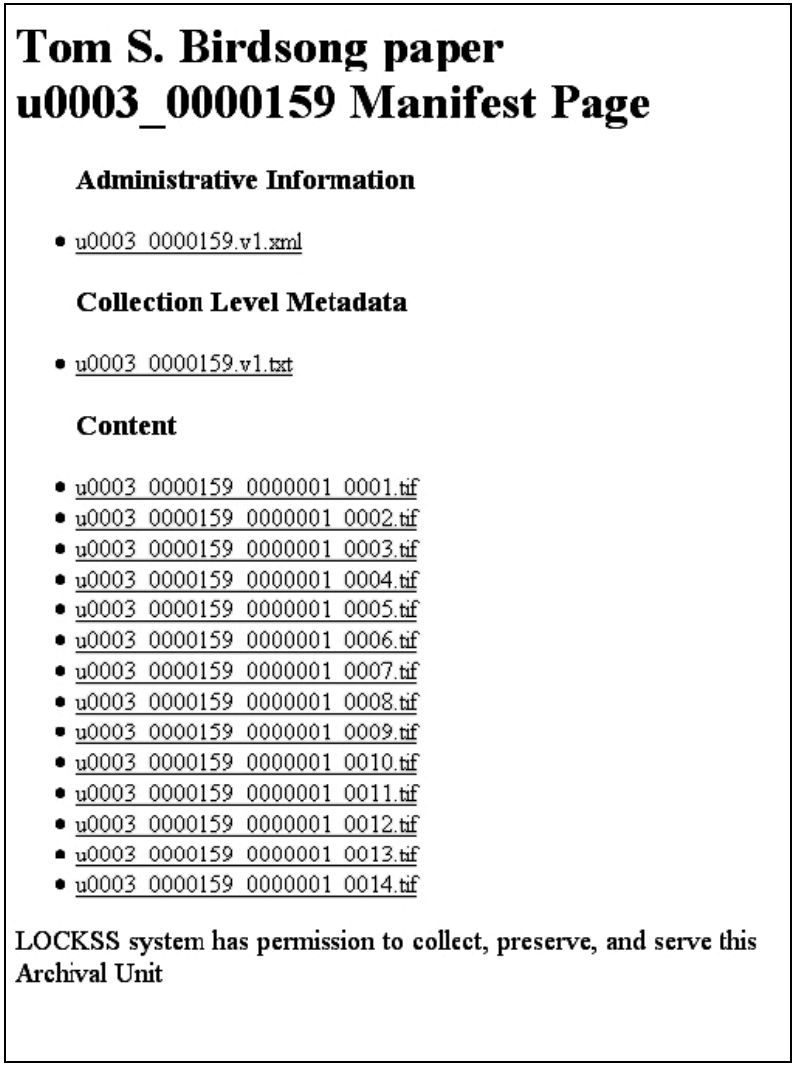
Figure ALAB-3. University of Alabama Libraries Digital File Naming Scheme: Sub-page numbering. (© 2009, University of Alabama Libraries. Used with permission.)

The manifest pages (Figure ALAB-4) link all files to be harvested for duplication across the Alabama Digital Preservation Network (Network of Alabama Academic Libraries, 2009), which is to date the lowest-cost model of digital preservation.

To secure our content further, we capture MD5 checksums upon deposit in our storage system, and verify them weekly prior to full-tape backups. This enables us to ensure that the original file is unaltered over time, and can be restored from a good backup should any corruption occur.

Reuniting Metadata with Archival Content

Software systems come and go, and successful transition between them is dependent upon standardized and coherent content and metadata. Most of our metadata had been altered or added after upload of content into our delivery system, and did not exist with the archival files. Upon examination of the exports from our software delivery system, we found that no single export contained all the metadata for a given archival object. Even in the most complete export option, there was no page-level metadata apart from the title and transcript, and archival file references were sometimes altered, often missing, and always contained reference to the upload directories, which no longer existed.



Tom S. Birdsong paper u0003_0000159 Manifest Page

Administrative Information

- u0003_0000159.v1.xml

Collection Level Metadata

- u0003_0000159.v1.txt

Content

- u0003_0000159_0000001_0001.tif
- u0003_0000159_0000001_0002.tif
- u0003_0000159_0000001_0003.tif
- u0003_0000159_0000001_0004.tif
- u0003_0000159_0000001_0005.tif
- u0003_0000159_0000001_0006.tif
- u0003_0000159_0000001_0007.tif
- u0003_0000159_0000001_0008.tif
- u0003_0000159_0000001_0009.tif
- u0003_0000159_0000001_0010.tif
- u0003_0000159_0000001_0011.tif
- u0003_0000159_0000001_0012.tif
- u0003_0000159_0000001_0013.tif
- u0003_0000159_0000001_0014.tif

LOCKSS system has permission to collect, preserve, and serve this Archival Unit

Figure ALAB-4 University of Alabama Libraries LOCKSS Manifest example. (© 2009, University of Alabama Libraries. Used with permission.)

In the metadata, if indeed there was an identifier, it was stored in multiple different fields in different forms. Sometimes there was nothing at all to indicate what the original file name had been.

We studied the file naming schemes used in all the 30-some existing collections, working out how to rename the files to retain the ordering of delivery, the intended organization of complex files, and the relationships between related items. Analyzing the metadata in the CONTENTdm (OCLC, 2009a) database, we matched up what information we could locate with what little had been retained with the archival files, and slowly filled in the gaps. Whatever sorting and matching could not be scripted was done by hand, often requiring research and investigation.

During this time we explored the use of 7train (Fogel & Hetzner, n.d.) for transforming exported CONTENTdm Standard XML metadata into METS (Library of Congress, 2009b) files, which reordered and clarified the relationships between complex objects. Based on 7train's method of using the first Dublin Core (DCMI, 2009) identifier field in the export as the file name for the resultant METS file, we selected this field for our file identifier, and tagged it the same in every collection. Our metadata librarian (Mary Alexander) has worked hard to remediate the metadata, entering the correct unique file identifier in the specified location in all records. Only by repairing the descriptive metadata to consistently reference the correct identifier, can we match our exported metadata to the appropriate archival files.

The benefits of using this process are that the Dublin Core metadata assigned to the object at the top level is retained, the organization of the complex digital object is retained, and the transcripts or Optical Character Recognition (OCR) content are included in the resultant METS file. However, what is lost includes any metadata which does not map to Dublin Core, value-added labels which clarify the content in the fields, and page-level metadata beyond the title and transcript. If no archival file was used in upload, it is not referenced in the METS; those referenced may have their file name altered, and always refer to the upload directory. If the location of this directory is not corrected after upload to reference the actual location of the archival file, this is useless for reuniting metadata and archival content.

This version of METS was designed for web delivery, not for preservation. California Digital Library created 7train to enable repositories in the state that were using CONTENTdm to participate in their state-wide digital federated search service (Fogel, 2006), which requires METS (California Digital Library, 2009). The METS file contains links into the CONTENTdm software for access to the thumbnails and service derivatives. For preservation, these links will be useless, as delivery systems change rapidly, if indeed the content is still online. Additionally, the 7train METS contains no technical or preservation metadata, as none was created by, or exported from, CONTENTdm.

Seeking to make our METS files more useful, we determined how to add technical metadata, and analyzed the database structure and storage directories to identify the actual location and name of thumbnails and service images. Scripting to replace the derivative links and archival file reference with full path links to the actual files proved to be more trouble than it was worth.

However, as our storage system began taking shape, another option emerged. Our storage structure reflects the compound file structure, creating an unambiguous arrangement which will survive any delivery software. A digital archivist of the future should have no trouble reconstructing our content. The METS file itself may be redundant. Rather than altering the 7Train METS file to meet our preservation needs, we decided instead to leverage our organizational scheme to meet the challenge. All we really need to do is to create the technical metadata, name it correctly, and drop it in the right directory (Figure ALAB-5). Then we will compile the metadata and content links for an item via script into a preservation-ready METS file for long-term storage.

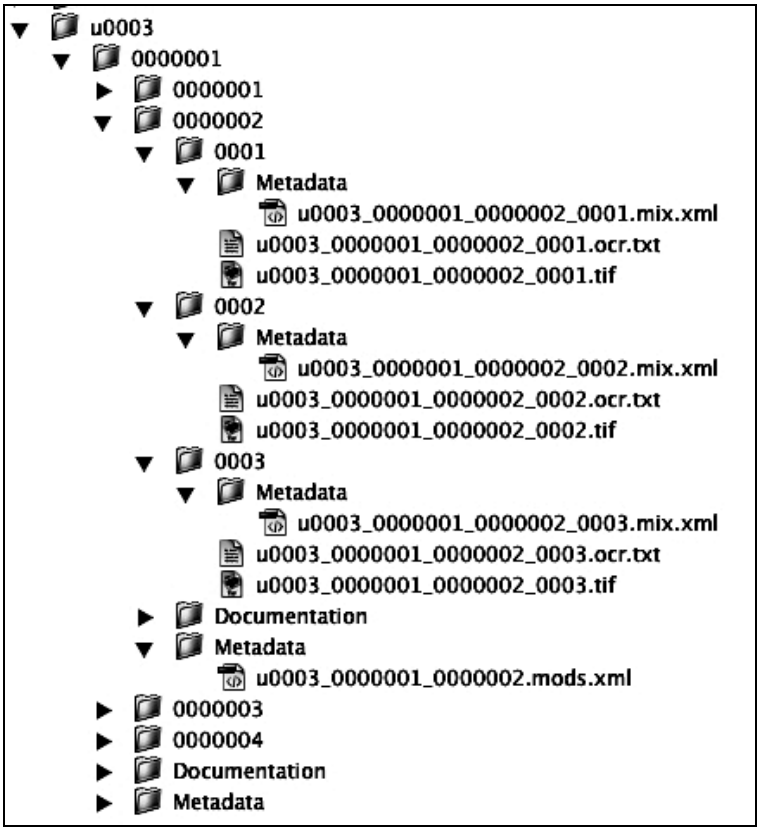


Figure ALAB-5. University of Alabama digital file organization for preservation. (© 2009, University of Alabama Libraries. Used with permission.)

Software Limitations on the Number of collections

The expansion of Digital Services had a major impact on Hoole Special Collections. Archivists were scrambling to find sufficient content to digitize at first; small collections were easiest to pull together, so we suddenly were digitizing many tiny collections. What an archivist considers a collection is determined by provenance, not by quantity.

However, CONTENTdm (version 4.3) was designed to support no more than 200 collections. It was clear that we could no longer define our collections the same way our delivery software did. Therefore we needed to reorganize our content in a way that provided the user

access desired by our archivists, while meeting the constraints of our current delivery system.

We met with the archivists to ask what kind of grouping made sense to them. After much debate, they finally selected date ranges for most of our content, which corresponded to particular eventful periods in American History. Access to each of the digitized analog collections would be managed by canned links searching for the analog collection name in the Dublin Core relation:isPartOf field. Archivists assigned each collection to a time period, and the metadata librarian and I began to sort out how to merge multiple collections and split combined ones.

When merging collections in CONTENTdm, the two collections being merged must have identical metadata fields. It was during this process that we realized the depth and variety of our metadata across all those 30-odd collections. It is far more than can be captured in Dublin Core, and we realized that we wanted to retain the value-added labels. Knowing that a person is a photographer or a lyricist or a composer or a performer is far more valuable than can be conveyed by Dublin Core "creator."

Our metadata librarian combined all our metadata fields into a single spreadsheet, so that all the CONTENTdm "containers" could handle any of our collections. Since each label requires a different spreadsheet column for ingest, we found we needed 87 columns. However, by hiding the columns that aren't needed, this becomes manageable. Different versions of this spreadsheet are used for each type of material, simply with different hidden columns. Since all the containers have the same metadata configuration, this streamlines uploads. The metadata librarian need only unhide the columns and export the spreadsheet for upload. Recognizing the value of the many tags in use, we decided to map all the fields to MODS (Library of Congress, 2009c) to discover if that metadata standard would be capable of retaining all our descriptive information; and indeed it does.

My associate dean recognized the possibilities for further leveraging our current file organization for better user access and

delivery, sidestepping many of the restrictions placed on us by CONTENTdm. He procured the services of a talented programmer (Tonio Loewald), who proceeded to write a translator, which, given a template, can translate any Unicode tab-delimited spreadsheet into separate XML records of the desired schema. His Archivists Utility (Loewald, 2009b) reads in our 87-field spreadsheets and creates a MODS XML record from each line. This enables us to capture all of our metadata, not just the Dublin Core elements, outside of the delivery software for preservation storage.

We added scripts using open-source software such as ImageMagick (ImageMagick Studio LLC., 2009), LAME (Cheng et al., 2009), and Tesseract_OCR (Google, 2009) to generate derivatives from the archival files. These are placed into web-accessible directories that mirror our archive structure. By adding the MODS, and now our newly emerging EADs (Library of Congress. 2009a) we now have all the components for an open, modular delivery system. Under my associate dean's direction, the programmer has built an XML schema-agnostic delivery system, Acumen (Loewald, 2009b), that reads the XML where it lives in a live directory. Metadata and derivatives can be accessed easily and changed at any time without going through any software system. Web agents and web search engine crawlers can easily access our online content also, as it is not buried in a back-end database. Relationships between files are inferred by the file naming system, so that all components of an item are retrieved together, and all items in a collection can also be retrieved by using the collection identifier. In addition, since the file name communicates the hierarchy and relationships of files, digital content can be reorganized according to work flow or even moved between servers while keeping unchanging URLs.

We're very excited about the possibilities this modular methodology offers. By bringing digital content up to the level of the web, we are setting the stage for semantic web applications and the development of user-friendly tools for access and reuse of our content.

Staffing Gap for Creation of Item-level Metadata

Shortage in funding support for archivists is widespread. With the current attention to digitization of archival content, and the lack of funding for archivists, a critical gap is created. The archivists are well-versed in the knowledge of the time periods, issues, relationships, and people related to the archival content we seek to digitize; they are the personnel best suited to describe the material in a way which will enable users the best possible access. A knowledgeable archivist will be able to identify important personages of the time, particular buildings and locations, and can provide biographical, historical, and cultural context which gives meaning to the documents we seek to digitize. Since the metadata provides the necessary information for successful retrieval, correct descriptive information may be the most valuable knowledge capture we could provide. However, the archivists are hard pressed to provide even minimal processing to the burgeoning mass of incoming content. In order for us to move forward in digitization without more funding for archivists, we began to devise pilot projects to seek out alternative possibilities.

Our first pilot project involved students creating item-level descriptive metadata as they digitized content. The collections chosen were small image collections containing a little over 200 photographs. To test for consistency, we assigned only one student to one of the collections, and four different students worked on the second collection. Within a few days it was clear that our careful instructions to the lone student were clearly insufficient. Her metadata spreadsheet was rife with errors; the primary focal person in the collection had his name misspelled seven different ways, and many words were abbreviated (and usually incorrectly). In 56 entries, we located 217 spelling and abbreviation errors. Grammar, punctuation, capitalization and spelling errors were abundant, and the captions and descriptions created were vague and unhelpful.

In the jointly described collection, each student had his own methodology and focus, and the variations between the choices for descriptions and captions were sufficient to impair search and retrieval even within the collection itself. Where one student might

use the terms “infant” and “woman,” a second student would use the terms “baby” and “lady.” While the errors and misspellings were far less frequent, it was clear that for consistency, we needed either stricter guidelines and controlled vocabularies, or to limit the number of employees assigned to create metadata for any one collection. Additionally, since the students had no background information on the context of the collection, the time period, or the content, they were unable to identify well-known buildings, well-known personages, or other significant content of interest. This point became extremely clear when our students could not identify such historically important figures as Governor George Wallace or the locally beloved Paul “Bear” Bryant.

A second pilot project involves the reuse of existing MARC metadata for sheet music which had been cataloged over a period of years by various catalogers. After extraction of the MARC records and analysis, we were dismayed at the variations in how the MARC fields were used. For example, we found the first line of text in fields 590, 500, 740, and 246, along with other types of content. The arranger, lyricist, musician, and composer names were interspersed in field 245c with various textual prefixes, preventing a systematic method of separation. Without remediation, transformation to another metadata scheme would not be recommended, as we cannot safely crosswalk these fields. Metadata librarians and catalogers are currently involved in repairs to the original records.

A third pilot project involves having an archivist and metadata librarian each developing their own version of item-level metadata for the same photograph and correspondence collections. Each worker gathered time measurements, and neither party was allowed to view the others’ descriptions until finalized. Both versions of each collection will be put online in the same web interface, and usability tests will determine to what extent the differences impact user experience. Questions for users will be derived from the known metadata in the finding aid, since the information in the series description would lead the user to expect success in searching for the item-level material that fit the description. If, for example, the series description mentions letters about boarding schools prior to 1900, a

query in the usability study would be to locate a letter about boarding schools written prior to 1900. This method of creating queries was not communicated to the metadata creators, so it could not skew the results. It is possible that neither metadata version will provide helpful results to the queries. In the final analysis, the level of usability for each version will be weighed against the time cost and availability of the personnel in the department that created it. We hope to be able to relieve the archivists of the burden of item-level metadata creation. While archivists know more about the content, and provide more detailed and informed metadata, their available time is very limited. Over the years, we have assumed that the expertise of the metadata creators is of primary importance, but new findings by Paul Conway (Conway, 2008) have thrown doubt on that belief. Since "the proof is in the pudding," we will determine if it is to our advantage to reduce the quality of metadata in order to get more of our content online.

A fourth pilot project will depend solely upon EAD finding aid descriptions at the series and sub series level to provide findability and context for digitized items. Our Archival Access Coordinator (Donnelly Lancaster Walton) suggested that we seek to recreate online the experience of a patron exploring material in the reading room: a folder is opened, and the patron goes through the documents one at a time. No other information is available to describe the material. This methodology will enable archivists to focus on EAD finding aid descriptions, and our digitization team will take box after box and simply digitize content in order for web delivery. As content is digitized, links will be added to the online finding aids from the folder level, out to web directories which contain the digitized items, ordered as they were encountered in the folder. In this manner, we will be able to provide online access to huge collections for which we have insufficient resources to provide item-level descriptions.

We have just been funded by NHPRC to demonstrate this low-cost, scalable model (University of Alabama Libraries, 2010a). Already we have developed scripts (University of Alabama Libraries, 2010b) to add links into the EADs and create minimal MODS records as quantities of scans become available. The software we develop to support this mass digitization method will be made available open

source. As soon as the content begins to appear online, we will conduct usability studies to compare the user experience between accessing content via the finding aid, versus item-level search and retrieval, and will publicize the results.

The fifth pilot project we expect to undertake in the coming months will involve users in tagging our materials online. To the extent possible, we will repurpose available open source software, capturing the tags and free-text descriptions in a database for vetting by our metadata librarians. Tags may be made automatically visible in the web interface and included in indexing for search and retrieval. After review, acceptable entries will be added to the to the descriptive metadata record on file. Our first foray into this venue will be with photograph collections for which we have almost no descriptive metadata. Hence, any apparently valid contributions will be accepted. In this manner we also hope to build user interest and support as we build interaction into our interface.

Summary

In the real world, digitization support within an institution may be fragmented across several departments, only partially funded by donor gifts, and may suffer seemingly arbitrary restraints imposed by delivery software. Some of the departments upon which success depends may be understaffed and unable to meet the demands for digitization support. In addition, most digitization is undertaken with little thought for the future, either in terms of transitioning to alternative delivery software, or long-term access to digital content.

Obstacles are opportunities for creative solutions. We leveraged expertise available across the library to build a successful cross-departmental collaboration. We are developing open source software support for an open, scalable, modular digital content delivery system with consideration for long-term preservation. Our file organization patterns alone may prove to be a life raft for digital content which funding cuts have left unsupported. This or a similar patterning will make reconstruction of digital content far easier. By possibly adding a BagIt manifest (Boyko, Kunze, Littman, Madden & Vargas, 2009), it

becomes potentially feasible to zip up entire digital archives for long term storage in a safe repository or LOCKSS until such time as funding support returns.

The delivery system we are developing will be free and open source, requiring minimal technical expertise. An underfunded institution will be able to use it to raise their content to the level of the web where search engines can promote their materials and web agents can provide greater usability. Since the actual metadata and delivery content is not ingested, but left in the web directory, it can be changed as needed, without risk of harm to the online delivery. Metadata schemes change regularly, and this XML schema-agnostic, modular solution is a low-cost, scalable, simple approach to building online digital libraries with the future in mind.

Thus far, we still use CONTENTdm for participation in multi-site search systems that depend upon this software, such as our state-wide Alabama Mosaic (Network of Alabama Academic Libraries, n.d.). However, the winds of change are blowing. During an open discussion at the 2009 AlabamaMosaic Annual Meeting in Montgomery, Alabama, the director reported a conversation with an OCLC technician who stated that FirstSearch would replace the CONTENTdm Multi-Site server. Thus, continued support of the CONTENTdm Multi-Site Server (OCLC, 2009c) is suspect, as OCLC moves toward FirstSearch (OCLC, 2009b) for cross-database search support. Many small collaborative digitization efforts depend upon CONTENTdm, but they need not be held hostage by a proprietary system. Acumen could potentially fill the gap, offering a low cost option providing much of the same functionality without the constraints.

We continue to explore how best to fill or reduce the gap created by the apparent need for item-level metadata for access and retrieval, and hope that the solutions we adopt in that area will help others as well. We have certainly found that working together with mutual respect and consideration has brought both challenges and unexpected benefits. By bringing everyone to the table, struggling together with the chaos before us and facing the same goals, we have

unleashed passion and creativity, enabling us to make astounding strides into clarity, organization, and hope for the future.

References

Boyko, A., Kunze, J., Littman, J., Madden, L. & Vargas, B. (2009). *The BagIt file packaging format* (V0.96). Retrieved November 14, 2009, from http://www.cdlib.org/inside/diglib/bagit/bagitspec.html

California Digital Library. (2009). *Online Archive of California (OAC)*. Retrieved November 15, 2009, from http://www.cdlib.org/inside/projects/oac/

Cheng, M., Taylor, M., Hegemann, R., Leidinger, A., Tominaga, T, Shibata, N., et al. (2009). *The LAME project*. Retrieved November 14, 2009, from http://lame.sourceforge.net/index.php

Conway, P. (2009). The image and the expert user: a qualitative investigation of decision-making. Paper presented at Archiving 2009, Arlington VA. In *Archiving 2009, Vol. 6*. (pp. 142-150). Society for Imaging Sciences and Technology.

DCMI. (2009). *Dublin Core metadata initiative*. Retrieved November 15, 2009, from http://dublincore.org/

Fogel, P. (2006). CDL 7train Profile – *CONTENTdm simple and complex objects in METS Metadata encoding and transmission standard*. Retrieved November 15, 2009, from http://www.loc.gov/standards/mets/profiles/00000010.html

Fogel, P. & Hetzner, E. (n.d.). *7train METS Generation Tool*. Copyright the University of California Regents. Retrieved November 15, 2009, from http://seventrain.sourceforge.net/

Google. (2009). *Tesseract-ocr*. Retrieved November 14, 2009, from http://code.google.com/p/tesseract-ocr/

ImageMagick Studio LLC. (2009). *ImageMagick*. Retrieved November 14, 2009 from http://www.imagemagick.org/script/index.php

Library of Congress. (2009a). *EAD Encoded archival description (Version 2002)*. Retrieved November 14, 2009, from http://www.loc.gov/ead

Library of Congress. (2009b). *METS Metadata encoding & transmission standard*. Retrieved November 14, 2009, from http://www.loc.gov/standards/mets/

Library of Congress. (2009c). *MODS metadata object description schema*. Retrieved November 14, 2009, from http://www.loc.gov/standards/mods/

LOCKSS. (2008). *What is the LOCKSS program?* Retrieved November 14, 2009, from http://www.lockss.org/lockss/Home

Loewald, T. (2009a). *Acumen*. Retrieved November 15, 2009, from http://acumen.lib.ua.edu/

Loewald, T. (2009b). *Archivists utility*. Retrieved November 15, 2009, from http://lb-416-003.lib.ua-net.ua.edu/notes/?f=Archivist%20Utility.txt

Network of Alabama Academic Libraries. (n.d.). *Alabama Mosaic*. Retrieved November 14, 2009, from http://www.alabamamosaic.org/

Network of Alabama Academic Libraries. (2009). *The Alabama Digital Preservation Network (ADPNet)*. Retrieved November 14, 2009, from http://www.adpn.org/

OCLC. (2009a). *CONTENTdm Digital collection management software*. Retrieved November 15, 2009, from http://www.contentdm.com

OCLC. (2009b). *FirstSearch Online reference*. Retrieved November 15, 2009, from http://www.oclc.org/firstsearch/

OCLC. (2009c). *Multi-site server*. Retrieved November 15, 2009, from http://www.oclc.org/firstsearch/

University of Alabama Libraries. (2010a). *Septimus D. Cabaniss Papers digitization project*. Retrieved March 9, 2010 from http://www.lib.ua.edu/libraries/hoole/cabaniss

University of Alabama Libraries. (2010b). *UA libraries digital services planning and documentation*. Retrieved March 9, 2010 from http://www.lib.ua.edu/wiki/digcoll

W3C. (2005). *XML:id Version 1.0*. Retrieved November 15, 2009, from http://www.w3.org/TR/xml-id/