Automating a Digital Special Collections Workflow Through Iterative Development

Brian Dietz and Jason Ronallo

A recent trend in archival practice stresses developing flexible and efficient procedures in order to work through physical backlogs of inaccessible collections. An emerging approach to the digitization of archival and manuscript materials calls for a focus on getting materials online in the aggregate. Both emphasize new workflows in order to make large amounts of resources available for research. Through quick, iterative, and targeted development, we have started to create procedures that will allow us to scale up our digitization production and enable greater access.

The NCSU Libraries Digital Program for Special Collections (DPSC) has been using tools and workflows that were created to facilitate an "exhibits" approach to digital projects. Projects have been designed around a theme, for instance, historic architecture in North Carolina or the importance of tobacco for the state's economy. The projects featured materials, selected from multiple collections, that were described with a fair amount of detail, including devised and transcribed titles, subject headings, creator names and notes, genres, and geographic locations.

These tools and workflows were adequate for our approach to digitization, but a new approach had emerged that emphasized higher rates of production. We determined that the overall digital infrastructure for special collections materials did not accommodate the more efficient and largely automated workflows that were likely to be required by this new approach. We began finding inspiration in the institutions that were advocating,[1] implementing,[2] and funding[3] large-scale approaches to digitization of special collections.

Large-scale digitization is a systematic approach that exposes resources at the collection level by scanning and putting online large quantities of archival and manuscript materials. By organizing digitized objects by archival collection, rather than by themes and across collections, it maintains an individual object's context within its archival context and collection's arrangement and description. It also redirects the work of selection from curating items to entire series or collections.[4] Placing access at the forefront, it arguably closes the distance between researchers and reading rooms[5] by bringing more materials to where the researchers already are, online.

Large-scale digitization also fits within the context of "More Product, Less Process" (known as MPLP), the influential essay, by Greene and Meissner, that urges archivists to deploy flexible and efficient processing and descriptive procedures in order to work through backlogs[6] and make materials accessible. MPLP, like large-scale digitization, asserts a "Golden Minimum… to maximize the accessibility of collection materials to users"[7] and a "baseline level of access."[8] In a time of dropping budgets, increasing collections size, and increased use,[9] MPLP-like processing may provide guidelines to special collections units for meeting the challenge of staff and resource allocation.[10]

NCSU Libraries Special Collections Research Center (SCRC) was an early adopter of flexible, ef-

Brian Dietz is Digital Program Librarian for Special Collections at NCSU Libraries, e-mail: brian_dietz@ncsu.edu; Jason Ronallo is Digital Collections Technology Librarian at NCSU Libraries, e-mail: jason_ronallo@ncsu.edu

ficient processing of archival material, using an approach similar to that advocated in MPLP. Each collection is processed at a level of detail that is appropriate to its research value and that is necessary to make it available to users. Preliminary inventories, usually made at the time of collection accession, are created to prevent a backlog. We currently have no collections with no description, and very few collections without some form of published description, either as a catalog record, preliminary inventory, or full collection guide.

The DPSC wanted to streamline parts of our digitization process in a similar fashion in preparation for large-scale digitization. One area that has been repeatedly identified as a place to gain efficiencies and sustainability in the process of digitization is descriptive metadata.[11] The manual creation of individual metadata records under large-scale digitization is not scalable to the amount of expected output. For instance, by March 2009 the Archives of American Art had digitized approximately 600,000 objects.[12] Describing such quantities manually and individually would clearly not be an option. If description is not being created by hand, how is it created?

Materials with a collection guide or finding aid have already been described.[13] Description is often applied to archival materials at the aggregate level: a folder of materials are all described with the same folder title, with no item-level description. Collection-level description has been deemed by processing staff and some researchers[14] to provide an adequate level of access, and should be considered sufficient for the description of the digital surrogates of the same materials.[15] Descriptions can be harvested from existing collection guides[16] and repurposed for metadata elements, like titles,[17] for individual digital objects. We thought it was achievable to make the reuse of description more efficient through automation.

Rather than build a grand system or purchase one,[18] our program chose iterative, incremental, and targeted development of our existing digitization systems and workflows. For us, iterative development meant planning in small steps where each cycle of development released a product that was immediately useful while also getting us closer to our larger vision. The work referred to in this paper was also done by an informal small team. These experimental side projects took place in a way that was incidental to other work. These limitations on development led us to look for ways through minimal effort to make greater gains in efficiency. Each stage of development allowed us to have real cases to which we could apply these approaches and for which we could develop and test software as we progressed. The stages of development were a database migration, a filename convention, and a metadata service.

Database Migration

In November 2009, we completed an interdepartmental project[19] that integrated the DPSC production digital projects metadata database with two legacy databases, and then migrated all data from the joint MS Access database to a single MySQL database.[20] The goals of the project included making our data available in a format that was more programmatically accessible in order to enable automated workflows, and supporting development of access layers that relied on data reuse. Once the data was migrated to MySQL we created a new digital assets web application, to interact with the new data store.[21] Much of what is detailed in this paper would not have been possible, or as easily possible, without this migration. Data now exists in a format that is accessible to web-based services. This allows us to be responsive to opportunities such as data sharing and reuse.

Though our data is stored in a locally schemed relational database we can export or request data in various formats. The schema has been mapped to the Metadata Object Description Schema (MODS), Version 3.3, which has allowed us to develop an export of MODS metadata. The digital assets application over our metadata store has enabled us to create custom "shareable metadata"[22] for particular uses. Staff has implemented generic XML and JSON outputs, which can be requested through a simple web API. For instance, one JSON response is used to extract a subset of digital collections metadata which is then combined with other information on NC State University history in the portal Historical State.[23] An XML API response was crafted to allow the Libraries IT department to showcase digital collections on a Microsoft Surface. For an ongoing project with other institutions we were able to export custom XML.

The database migration and development of a quickly scaffolded application also allowed us to more quickly iterate over the schema of our database. As we saw what was possible in our access layer, and not possible or as easy as we had liked, we were able to

make modifications. For instance we once stored geographic locations only by their headings, for example, "United States—North Carolina—Wake County—Raleigh." Most of our resources have geographic locations within North Carolina, so we wanted to be able to facet by county. We added county, town, and latitude and longitude columns to the "geographic locations" table in the database and were able to make a first pass at auto-populating these new fields through a simple script. Metadata creators then reviewed and made corrections as needed. Many small quick changes like this were made simpler, easier, and as we needed them through this migration and subsequent development. Other tasks of migrating schema or data have allowed us to automate some of these changes. The flexibility of this locally developed solution continues to allow us to remain agile in meeting emerging needs.

Filenaming

With a data store under our control, the next problem to emerge in devising an automated approach to the reuse of archival description was how to get the metadata out of the EAD XML documents. Our goal was to deploy a means to programmatically find an object's physical and intellectual location in a collection, automatically extract existing descriptive metadata, and use it to populate a brief, or stub, metadata record. The problem was how to associate a digital object with its place in the archival hierarchy.

This connection could be accomplished by resolvers or a relational database,[24] but these solutions did not fit into our current context. We had to find another solution. We were already handling our filenames as unique identifiers, so we decided to take a closer look at whether using a semantic filename could express an object's place in a collection and its EAD XML document.[25]

We piloted the filename with a new digitization project before we had developed the ability to make use of any extracted data. The filename implementation was planned to be a lightweight solution that would not commit us to a bigger development project. It did not fundamentally change the workflow, even though it was done in anticipation of doing so. At this point, we had only proposed a different convention for filenames/identifiers.

The DPSC, in consultation with SCRC processing staff, devised a convention for naming digital objects. Consultation included making sure our identifier elements were exhaustive for creating unique filenames, and that processing practices would not change the EAD XML document in a way that would impact the stability of the identifiers.

The identifier includes the collection number, series number, container and folder numbers, and an arbitrary sequence number, followed by, in text-based resources or those with sections or divisions, a page extension number. An example of such an identifier is mc00126-001-bx0001-006-001_0001,[26] where:

- mc00126 is the EAD identifier (contents of EAD element <eadid>)[27]
- 001 corresponds to series one (<c01>)
- bx corresponds to a container code for box and 0001 corresponds to the container number for the box (<container type="Box">1</container>)
- 006 corresponds to a folder number within a container (<container type="Folder">6</container>)
- 001 corresponds to the sequence number of an item within a container and folder (not encoded in a collection guide)
- _0001 corresponds to the first page of a multipage resource (not encoded in the collection guide)

We anticipated using this identifier to harvest collection, series, box, and folder titles from the EAD XML document. Towards this end we developed a small, independent software library, called MEAD,[28] which can parse these identifiers, retrieve the associated EAD XML document, parse the XML, locate the relevant component part, and return the associated metadata. The software allowed us to test that we could auto-generate identifiers and parse them to reliably retrieve metadata from an EAD XML document before implementing the new filenaming scheme. The same software can also be used to test EAD XML prior to digitization to validate that all component parts would create unique identifiers.

Stub Metadata and the Stub Record Metadata Service

At this point, we had a new way to create identifiers and filenames, as well as a piece of software (MEAD) that would harvest data from EAD XML documents using those filenames, but no method to reuse the harvested description for digital object metadata records. The next step was to integrate the MEAD library into

our digital asset management system so that we could create metadata in batches based on the filenames. We named this tool the stub record generator (SRG).

This first iteration of the SRG service was developed as a web form within our digital asset system. The SRG is initiated through an upload of filenames of digital objects that need associated descriptive metadata. The SRG first performs a well-formedness check on all submitted filenames. To be well-formed, an identifier must be able to be parsed and conform to the rules of what is allowable and required in an identifier. If all pass this check, the service then retrieves the appropriate EAD XML document and performs two tasks. In referring to the archival description contents in a collection guide (EAD element <archdesc>), it validates all filenames, so that, for instance, a filename that refers to a series or a container that is not present in the XML will be returned as invalid.[29] Once all filenames are considered well-formed and valid, the service will retrieve the relevant description in the descriptive identification element (<did>) at any component level (<c0x>) up through the nested hierarchy. Information retrieved include the unit title and date (<unittitle> and <unitdate>). These values are then written into a stub record for the digital object in the digital assets management database. The digital object filename becomes the natural, unique identifier in the stub metadata record. The archival description from the collection guide then forms the basis of a descriptive metadata record for a digital object.

The data returned is used to create brief metadata records called "stub records." The title of a stub record is a concatenation of all of the unit titles and unit dates from the file level up to the series level. Using the example filename (mc00126-001-bx0001-006-001_0001) the title would result in "Tri-State Tobacco Cooperative-Articles, publications and notes :: Writings and Publications Series." All objects from the container Box 1, Folder 6 in MC126 would inherit this title; all digital resources, when put online, would be as discoverable as the physical materials they represent. Description created for an aggregation of objects would be supplied to multiple individual items. The SRG is able to derive a physical object's physical location (box and folder it is stored in) from the semantic information contained in the filename and write it out to a location note in the resource's metadata record. The service also supplies the record with a title source, i.e. "Title derived from collection guide," and a record status (stating the extent of description, where "0" denotes a stub record).

While initially developed to leverage the new filenaming convention for the creation of stub records featuring existing archival description, we quickly saw the opportunity to expand the functionality to accommodate multiple methods for creating brief metadata records. Titles can be created directly from filenames, a method which has been used in cases where donors have supplied electronic files with meaningful filenames. Titles can also be created through import of a spreadsheet, which has been used to create records from metadata provided by digital project partner institutions.

Beyond creation of titles through metadata harvesting and other means, a simple template form was added to the SRG which allows for us optionally to perform limited semi-automated batch metadata creation. It can be used to assign additional generalizable metadata to stub records using controlled terms.[30] If project staff is creating stub metadata records for a set of resources with the same creator, for instance a set of drawings by an architect, in the SRG template the "creator" field value can be assigned the heading for the architect. There is an obvious benefit in being able to manually fill in metadata fields once, beyond the values that are being harvested from a collection guide, and have those values applied to all records created in that batch. Any stub record can also be assigned to NCSU Libraries' Metadata & Cataloging department for further enhancement, if the content demands further description. In these cases the template approach has reduced the amount of duplicate information metadata creators have had to enter.

The SRG workflow is the basis of a more efficient workflow that leverages existing description. The process reuses existing metadata to quickly create adequate metadata that could enable immediate access online to digital objects. Stub records enable quicker access to digital objects, but they are minimally discoverable as individual items until or unless the records are enhanced. In this way it prioritizes access over discoverability.

Reflections and Further Work

An assumption of this workflow is that collection-level description applied to a digital object makes the object as discoverable as collection guides do. The quality of archival description has implications for staff

time and quality control, as well as for discovery and search engine optimization. Other opportunities that we anticipate exploring include extending an MPLP-like approach to digitization, furthering automation, and the interplay of metadata between archival description and digital objects.

Quality of Existing Archival Description

The burden of pre-publication quality control for description now rests largely with processing staff. We had already begun putting less emphasis on quality control of metadata in an attempt to speed production and publication, but now even less quality control will be done by DPSC staff. DPSC staff cares, initially, about the validity of an EAD XML document for interacting with the SRG, but even SRG-valid collection guides will contain description of varying quality. Below are three examples of series and folder titles pulled from the collection guide to the "University Archives Photograph Collection, College of Engineering Photographs, 1915–1993 (Bulk, 1950–1979),"[31] formatted as stub titles for digital objects.

- Chemical Engineering - General :: Chemical Engineering
- Faculty :: Faculty
- Miscellaneous :: Miscellaneous

These are cases where flattened multi-level, hierarchical description does not work outside of the context of the finding aid.[32]

The above examples suggest this workflow creates a growing interdependence between a institution's digital program and its special collections processing staff. Said another way, this is mainstreaming of digitization as another access-oriented special collections workflow. We will need to negotiate and document how archival descriptive work will may need to be adjusted in consideration of automated workflows for collection level scanning.

Extending MPLP to Digitization

Our DPSC, in consultation with the SCRC's Curator and University Archivist, is developing a queue of archival and manuscript materials to be scanned at the collection level. We anticipate that not all collections will be treated equally regarding metadata. Following the MPLP approach, we believe that some scanning projects will merit more than stub metadata. Just as some collections will merit richer description, so, too, would some digital objects or aggregations of digital objects within a collection. In consultation with SCRC and Metadata & Cataloging staff, we will be developing a matrix and questionnaire to help determine what collections will receive only stub metadata and what metadata records from within a collection will be assigned to Metadata & Cataloging for enhancement.

More Automation

We are at the very early stages of investigating a workflow that extends automation. The exploration is around use of shared network storage in which scanning staff can deposit digital objects to be automatically processed nightly. When new objects are found, a series of processes are set off, including the validation of filenames, the creation of stub records, the creation of access derivatives through our image server, and the moving of files to a gray archive. Project managers would receive notification that a file or set of files had been processed, and further action could be taken.

Scan First, Curate Later

Under this workflow, digitization is becoming a special collections workflow, rather than a set of customized procedures. Digitization projects exist along a continuum that ranges from collection-level scanning with minimally applied metadata and extends to more highly curated exhibits created by library staff and, potentially, library users like students, professors, and public patrons. An collections-level approach to digitization makes available more objects to be repurposed in subsequent curated collections.

Metadata Feedback

An area for experimentation that is opened up by better integration of collection guides and digital objects is what we have begun referring to as the "metadata feedback loop." We harvest metadata from collection guides and write it to metadata records for digital objects. If an item-level metadata record receives any enhancement, is there a role for that metadata in the collection guide? Metadata feedback regards whether any enhanced descriptive metadata created for digital objects is repurposed back to the collection guide. Additional metadata, if integrated or associated with the collection guide, would enable more accurate keyword searching of collection guides, which would then facilitate better chances of discovering collections and their digital surrogates. The answer is less

certain as "digital projects" become less about curated, stand-alone collections and more about integrative special collections workflows enabling access to research collections.

Search Engine Optimization

*T*his concern with metadata quality is also a potential problem with making the digital collections discoverable on the open web apart from their related collection guides. The keywords used in titles are important for search engine optimization (SEO). Since these folder-level titles can be applied to more than one resource, search engines may penalize sites that have duplicate content. Google even warns to "Minimize similar content" and "Avoid publishing stubs" that do not "yet have real content."[33] We have not implemented any potential fixes and do not know yet to what extent these metadata practices will effect the crawling and indexing of our collections or what preventative measures would best serve the SEO needs of our digital collections as a whole. We are actively researching these questions.

SEO becomes a concern and potential cost moved earlier in the special collections workflow and falling to processors to describe physical archival collections with the expectation that the metadata may be reused for digitized content that will be exposed on the web. In keeping with the spirit of an MPLP-like approach, we may have to be satisfied with only optimizing certain collections or select pieces that merit fuller SEO attention.[34] In some cases, like when processing architectural collections, it may be that providing enhancements like controlled terms for geographic locations at the folder-level would be less expensive to do during processing. This data could help improve discoverability of collections on the open web.[35]

Conclusion

Through iterative development, we have quickly built up pieces of our program's approach to digitization of special collections materials. Through a database migration, designing a new practice for creating filenames and identifiers, and a metadata service, we have implemented many components of an automated, efficient metadata workflow. Iterative development and frequent deployment made these accomplishments possible.

At the same time, we have barely scratched the surface in implementing a collections-level approach to digitization. As new challenges and opportunities appear, we will continue planning for new iterations of current and new tools, to the end of providing greater access to our research materials.

Notes

1. Erway, Ricky, and Jennifer Schaffner. *Shifting Gears: Gearing Up to Get Into the Flow*. Report produced by OCLC Programs and Research, 2007. http://www.oclc.org/programs/publications/reports/2007-02.pdf; Erway, Ricky. "Supply and Demand: Special Collections and Digitisation." *Liber Quarterly: The Journal of European Research Libraries* 18, no. 3/4 (December 2008): 324-336.

2. Institutions such as the Southern Historical Collection (UNC-CH), (http://dc.lib.unc.edu/ead/archivalhome.php?CISOROOT=/ead), the Archives of American Art, (http://www.aaa.si.edu/collectionsonline/), and the University of Wisconsin at Oshkosh have undertaken such projects. For information on UW Oshkosh's efforts, see Ranger, Joshua. "Mass Digitization of Archival Manuscripts" presented at the Something Old for Something Old: Innovative Approaches to Managing Archives and Special Collections, The Philadelphia Area Consortium of Special Collections Libraries (PACSCL), December 4, 2008. http://www.pacscl-survey.org/documents/ranger/04ranger.ppt.

3. The National Historical Publications and Records Commission (NHPRC), affiliated with the National Archives and Records Administration (NARA), initiated the Digitizing Historical Records grant program to support collections-level digitization. NHPRC "Digitizing Historical Records" grant announcement, http://www.archives.gov/nhprc/announcement/digitizing.html.

4. Erway and Schaffner, *Shifting Gears*; Erway, "Supply and Demand."

5. "The digital search room experience will mirror the physical search room experience: Users will be able to interact with primary sources directly." Southern Historical Collection, University Library, The University of North Carolina at Chapel Hill. *Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections* (June 2009): 7, http://www.lib.unc.edu/mss/archivalmassdigitization/download/extending_the_reach.pdf.

6. Backlogs have grown as the size and number of post-war twentieth-century archival collections have increased. Greene, Mark A., and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68, no. 2 (Fall 2005): 211.

7. Greene and Meissner, "More Product," 240.

8. Santamaria, Dan. "Guest blogger: Dan Santamaria shares some thoughts on the recent MPLP discussions." *ArchivesNext*, August 21, 2009. http://www.archivesnext.com/?p=332.

9. Dooley, Jackie M., and Katherine Luce. *Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives*. Dublin, Ohio: OCLC Research, October 2010, 23. http://www.oclc.org/research/publications/library/2010/2010-11.pdf

10. Santamaria, "Guest Blogger."

11. "Develop models for *large-scale digitization* of special collections, including methodologies for selection of appropriate collections, security, safe handling, sustainable metadata creation, and ambitious productivity levels" (emphasis in original). Dooley and Luce, *Taking Our Pulse*, 13.

12. Aikens, Barbara, Karen Weiss, and Toby Reiter. "Building a Large Scale Digitization Program at the Archives of American Art" presented at Moving from Projects to a Program: The Sustainability of Large-Scale Digitization of Manuscript Collections, UNC Chapel Hill, March 12, 2009, slide 4, http://www.lib.unc.edu/mss/archivalmassdigitization/download/aikens_weiss_reiter.pdf.

13. This implies that the materials must be processed and described in a collection guide or finding aid before digitization. The NHPRC, for instance, states the expectation that for their "Digitizing Historical Records" grant program, "The materials should already be available to the public at the archives and described so that projects can re-use existing information to serve as metadata for the digitized collection." See NHPRC, "Grant Opportunities" (http://www.archives.gov/nhprc/announcement/).

14. Southern Historical Collection, *Extending the Reach*, 34.

15. "*Stop thinking about item-level description*. Learn from archivists and let go of the obsession to describe items. Think in collections and arrange and describe unique materials in subunits…Putting the most minimal description on the web will not restrict use anywhere nearly as much as limiting discovery to those who are able to show up in person and who know whom to ask for what" (emphasis in original). Erway, "Supply and Demand," 327; "Nothing prevents us from organizing and describing digitized (or born-digital) material at the file or even the series level, except our own fascination with individual documents." Greene, Mark A. "MPLP: Not Just For Processing Anymore." *American Archivist* 73 (Spring/Summer 2010): 194; "If this brave new world of Web access to archives depends on document-level description, archivists are doomed by the sheer mass waiting in the unprocessed stacks. The expense of extracting and entering item-by-item metadata multiplies many times over the huge costs involved in scanning entire collections." Evans, Max J. "Archives of the People, by the People, for the People." *American Archivist* 70, no. 2 (Fall 2007): 388.

16. Evans, "Archives of the People," 392.

17. A title would be a required element to enable the reuse and repurpose of digital object metadata. The title element is also required in our system, and since the development of this project was to be incremental we needed to have an item-level metadata record that could be used within our current system. Looking ahead, having an individual record would be necessary for any future record enhancement. Evans makes the point that, "It is not just minimum metadata; it is *extensible* metadata." Evans, "Archives for the People," 395.

18. In 2007, an NCSU Libraries committee dedicated to digital collections evaluated its then current architecture and software for digitized materials. The evaluation also looked at alternative purchased solutions, including CONTENTdm. The assessment was done against compiled lists of functional requirements (both "required" and "desired") for metadata creation tools and access layers. No software assessed during this period provided the set of required functionality, so we chose to table the issue, which allowed time to consider other options. Afterwards, use cases began appearing that both showed the difficulty of extracting data for reuse between projects and that justified small targeted development for localized problem. Development grew as our problem space emerged and expanded.

19. The database migration was an NCSU Libraries Digital Collections Technical Oversight Committee project carried out by staff in Digital Library Initiatives, Information Technology, Metadata & Cataloging, and the Special Collections Research Center.

20 The main tool we used to migration from the Access database was the MySQL Workbench developed by MySQL which has a nice graphical user interface, is simple to use, and has a translator from Access to MySQL. The tool was located here http://dev.mysql.com/doc/migration-toolkit/en/ but it has reached its end of life. The current version of this tool called the MySQL Workbench (http://dev.mysql.com/downloads/workbench/) does not seem to have similar functionality yet. Archived versions of the MySQL Toolkit can be found here: http://dev.mysql.com/downloads/gui-tools/5.0.html

21. This is written in Ruby on Rails. This process was simplified and expedited by using a script which generated Rails code based on the contents of the existing tables.

https://github.com/ahe/reverse_scaffold

22. Our digital asset system, as Riley et al. put it, provides "a shareable metadata record" which is "a second copy of a descriptive record, designed explicitly for sharing." The record is "generated automatically from the 'master' record." Riley, Jenn, and Kelcy Shepherd. "A Brave New World: Archivists and Shareable Descriptive Metadata." *American Archivist* 72, no. 1 (Spring 2009): 95.

23. Historical State (http://historicalstate.lib.ncsu.edu/).

24. The Archives of American Art uses a relational database which seems to break the EAD into component parts which are associated with digital objects. Aikens, Weiss, and Reiter, "Building a Large Scale Digitization Program," slides 21, 23.

25. Other institutions are also using variations on a semantic identifier that includes information like collection identifier (EAD element <eadid>) and box/folder number, linking an object to its physical location in the collection. See "Project Documentation : Digaccess : Digitization." http://www.lib.unc.edu/mss/archivalmassdigitization/index.html?section=digaccess&page=digitization; and Guegguen, Gretchen. "Digitized Special Collections and Multiple User Groups." *Journal of Archival Organization* 8, no. 2 (2010): 96.

26. The identifier appears complicated. However, it has proven relatively easy to construct, and once scanning staff has been trained, there is very little management needed, unlike the problems that occurred trying to manage an arbitrary and sequential schema.

27. Guide to the Ralph Waldo Green Papers, 1903-1946, MC 126, Special Collections Research Center, North Carolina State University Libraries (http://www.lib.ncsu.edu/findingaids/mc00126)

28. MEAD (Metadata from Encoded Archival Description) is available as open source and can be packaged as a Ruby gem (https://github.com/jronallo/mead).

29. Since the initial iteration of the SRG, as part of the MEAD library we are able to "validate" all of our EAD XML documents for use with the SRG. Knowing which collection guides are "valid" may help staff determine the best potential candidate collections for this workflow.

30. The additional fields include creator, genre, geographic location, rights, repository, and subject, along with technical metadata fields.

31. Guide to the University Archives Photograph Collection, UA 023.013, College of Engineering Photographs, 1915–1993 (Bulk, 1950–1979), (http://www.lib.ncsu.edu/findingaids/ua023_012)

32. "One specific activity that most archives could benefit from is to design multilevel description with an eye towards how it could be 'flattened' in the future to file- or item-level view, outside the context of the full finding aid." Riley and Shepherd, "Brave new world," 110.

33. Google suggests consolidating pages into one or using a noindex meta tag for "placeholder pages" (http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=66359).

34. Some tools from Google may be instructive for determining the best terms to use for description of collections to improve their discoverability on the open web. Google trends (http://www.google.com/trends) indicates which of several terms is most popular or trending popular. Google AdWords Keyword Tools (https://adwords.google.com/select/KeywordToolExternal) can be useful beyond buying advertising as it shows the number of searches for particular terms and the competition for those terms from other sites.

35. The SRG allows for any metadata at any level in an EAD XML document to be harvested and potentially used. For instance, in describing landscape architectural collections as part of our "Changing the Landscape" project (http://news.lib.ncsu.edu/changinglandscape/), supported by the Council on Library and Information Resources "Cataloging Hidden Special Collections and Archives" program, the decision was made, in part based on feedback from scholars and researchers, to create robust structured metadata in the EAD XML. Information such as controlled access names and geographic data has been recorded during processing. While this will more immediately make the collection guides more discoverable and usable for scholars, it would also help with making the digital objects in any future digitization project more discoverable because more metadata could be harvested automatically.