

UTILIZING MPLP IN ARCHIVAL DIGITIZATION PROJECTS

Steve Moray
INFOST 650: Introduction to Modern Archives Administration
December 19, 2016

Since Mark Greene and Dennis Meissner introduced the concept of “MPLP” in their paper “More Product, Less Process: Revamping Traditional Archival Processing” in 2005, the archival world has undergone (or is, perhaps, still undergoing) somewhat of a paradigm shift in the realm of processing.¹ Larisa K. Miller, in her 2013 article “All Text Considered: A Perspective on Mass Digitizing and Archival Processing”, makes the case for what she describes as “Digitizing without Archival Processing”.² What Miller refers to as “without” processing is, perhaps, better described as using minimal processing. This minimal processing she describes is a theoretically similar concept to Greene and Meissner’s idea of MPLP. In fact, Greene went on to write a follow up to his and Meissner’s original article that dealt specifically with using the concept of MPLP in conjunction with digitization (among other topics).³ By using the basic tenets of MPLP, along with ideal archival standards for processing and digitization and a review of existing digitization processes, a theoretical framework of processing by digitization can be devised, much in the vein of Miller, that balances issues of backlog, access, preservation, authenticity, and time and funding constraints.

Greene and Meissner introduced the archival world to the concept of “More Product, Less Process” in their 2005 paper. They make the case that the *raison d’etre* of archives is to allow access of collections to users.⁴ In order to better facilitate access, however, archivists must be willing to make sacrifices in the quality of processing. Processing collections, from the physical removal of paper clips, to the arrangement, description, and creation of finding aids,

¹ Greene, Mark A., and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68 (2005).

² Miller, Larisa K., "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *The American Archivist* 76 (2013).

³ Greene, Mark A., "MPLP: It's not Just for Processing Anymore," *The American Archivist* 73 (2010).

⁴ Greene, Mark A., and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68 (2005): 212.

takes time and effort. And time translates into man-hours, which eventually comes back to funding. In a perfect world of unlimited funding, personnel, and time, MPLP would be unnecessary, but nearly every archive is subject to these constraints, which results in backlog. Backlog, Greene and Meissner contend, benefits no one. By focusing on the quicker processing of backlog, archives are better serving their clientele. Faster processing can be achieved through a variety of means. The first is by cutting down drastically the amount of physical processing through the removal of binders, and refolding and reboxing collections in acid-free materials.⁵ Greene and Meissner argue that environmental controls (low light, heat, humidity, etc.) should serve to sufficiently stabilize the majority of materials against major degradation.⁶ Second, archivists should consider carefully what would be the least amount of information needed for researchers to be able to locate material they need in a reasonable manner. If materials *need* to be described at an item level, there should be sufficient justification for this choice, but there is nothing inherently wrong with it. If, however, materials can be described at a box level or higher while still providing access to researchers, then the cursory description can allow archivists to better spend their time doing reference work or outreach (or processing even more backlog).⁷

Greene's follow up article "MPLP: It's Not Just for Processing Anymore", covers a variety of archival topics in regard to MPLP, including appraisal, reference, electronic records, and privacy. It also addresses concerns specifically focused on digitization. Most of these concerns, however, are the same, or similar, to concerns with MPLP when used on analog collections. Greene reiterates his previous opinion that item level description is not always necessary, even during digitization projects.⁸

⁵ Ibid., 221.

⁶ Ibid., 231.

⁷ Ibid., 217.

⁸ Greene, Mark A., "MPLP: It's not Just for Processing Anymore," *The American Archivist* 73 (2010).

Elizabeth Kelly, in her article “Processing Through Digitization: University Photographs at Loyola University New Orleans” offers a case study in a digitization project. She is also a proponent of the MPLP mentality of creating more access to users, stating that “[v]ast quantities of digitized primary materials will trump a few superbly crafted special collections”.⁹ Kelly deliberately ignores the professional best practices for digitization. Instead of the recommended 400-800 ppi lossless TIFF format for preservation with a JPG access copy, she opts for a lower quality, but faster scanning 300 ppi lossy JPG as the standard.¹⁰ Because there is no finding aid for the collection each photograph gets an item level meta-data description. Every photograph is credited with the Loyola University copyright, the photographer when known, standardized Library of Congress Subject Headings, and keywords based on any description that may have accompanied the original photograph. The photographs were then published using CONTENTdm.¹¹

In their article “Failure Is an Option: The Experimental Archives Project Puts Archival Innovation to the Test, Mary O. Murphy, Laura Peimer, Genna Duplisea, and Jaimie Fritz outline two different projects the “Experimental Archives Project” undertook at the Schlesinger Library at Harvard University. The Experimental Archives Project was a result of workshop focused on improving processing and access, with the more substantial output dedicated to many innovative digitization projects to assist in processing.¹² The first project Murphy et al. describe is “Redaction Redux with the Addenda Papers of Elizabeth Winship”, an attempt to digitize a second group of letters to a newspaper help columnist, including printouts of email

⁹ Kelly, Elizabeth., "Processing Through Digitization: University Photographs at Loyola University New Orleans," *Archival Practice* 1, no. 1 (2014): 2.

¹⁰ *Ibid.*, 3.

¹¹ *Ibid.*, 3.

¹² Murphy, Mary O., Laura Peimer, Genna Duplisea, and Jaimie Frit, "Failure is an Option: The Experimental Archives Project Puts Archival Innovation to the Test," *The American Archivist* 78, no. 2 (2015): 435.

correspondence.¹³ Murphy et al. state that “[d]irect-to-digital processing of a collection meant no refoldering, sorting, file renaming, or any conservation work, such as photocopying acidic materials or flattening oversize materials”.¹⁴ Although they also mention that the collection was in good condition when it arrived,¹⁵ which certainly plays a factor in the ability to minimize physical processing. The authors report that the lessons learned from the Winship project include: flexibility in process when presented with new challenges, the uniqueness of each collection and the need to tailor process to the qualities of that collection, and the standard MPLP idea that decisions must be made to tradeoff detailed description for faster access.¹⁶ The second project described in the paper was “Traditional-to-Tagging with the Records of That Takes Ovaries”. The collection was received with no clues as to the original order, so arrangement and description were able to be completely “born digital”.¹⁷ The faster and easier manipulation of data in the digital realm allowed the processing to become more experimental and iterative, fostering creative solutions to arrangement problems.¹⁸ One of these flashes of creativity came in the use of an explanatory video instead of the traditional written “scope notes”.¹⁹ Lessons learned from the second project included: the possibility of skipping physical arrangement and description completely in favor of digital, acknowledging to users they “may have to dig” to locate certain items, creative solutions to a lack of institutional resources (such as using Flickr to host images), and the importance of obtaining user feedback.²⁰

¹³ Ibid., 440

¹⁴ Ibid., 441.

¹⁵ Ibid., 442

¹⁶ Ibid., 444.

¹⁷ Ibid., 445.

¹⁸ Ibid., 447.

¹⁹ Ibid., 447.

²⁰ Ibid., 448-449.

Closer to home, at the University of Wisconsin – Oshkosh, Joshua Ranger makes the case for not using best practice recommendations for digitization in his presentation “More Bytes, Less Bite: Cutting Corners in Digitization”. In his 2008 presentation at the American Archivists Conference, Ranger outlines a study comparing two different digitization projects at the UW-Oshkosh archives. The first utilizing more traditional, best practices for digitization and description, the second “cutting corners” using lower quality techniques.²¹ Ranger then performed a user study of undergraduate and graduate students for both collections, with directed interviews after users had interacted with the collections.²² Ranger found that the cost savings accrued by photocopying then quick scanning, and applying higher level description and metadata was significant over high quality scanning original copies of items and applying item level meta-data.²³ Users, however, were turned off by the description level of the experimental group. They felt it was much harder to navigate and there should be some sort of search function, instead of just being browseable.²⁴ When explained about the cost savings factor, however, users were understanding, and indicated that they would rather have access to the low quality digital collection, then no access or limited access by having to visit the archives in person.²⁵ Ranger’s study provides some possible cost savings measures that archivists might take when working on digitization projects. It also may indicate that explanations of how digitization projects were performed and the motivation behind choices archivists made might increase user satisfaction and add some degree of returning to use the archives again.

²¹ Ranger, Joshua, “More Bytes, Less Bite: Cutting Corners in Digitization.” (presentation, Annual Meeting of the Society of American Archivists, San Francisco, CA, August 24-31, 2008): 5-12.

²² *Ibid.*, 13-23.

²³ *Ibid.*, 12.

²⁴ *Ibid.*, 17-20.

²⁵ *Ibid.*, 21.

Tracy M. Jackson, in her paper “I Want To See It: A Usability Study of Digital Content Integrated into Finding Aids”, discusses the results of a usability study of digital finding aids using CONTENTdm.²⁶ Jackson addresses the issue that not only do modern users expect collections to be made available digitally, but there must be some way for users to navigate such collections. While other options are possible (such as full content searching), digital finding aids combine the traditional archival theory with modern digital technologies to present users with (theoretically) a comfortable, yet functional way to use digital collections.

Based on the results of the usability study performed of the finding aids of the North Carolina Collection Photograph Archives (NCCPA) at the University of North Carolina – Chapel Hill, Jackson separated the users into three groups; “novice”, “intermediate”, and “advanced”.²⁷ She then analyzed each of these groups responses to a standardized questionnaire based on a user session with the online finding aids. Results were similar to what one might expect. Novice users had not really used finding aids before, and about half rated the as difficult to use in digital form. Intermediate and Advanced users, who had previous experience with analog finding aids, found the digital finding aids to be helpful in navigation the digital collections.²⁸

The Smithsonian’s American Archives of Art website publishes a guide to digitizing entire collections. This guide outlines best practices for digitization, including such aspects as file format, PPI, initial meta-data, workflow, creation of finding aids, and quality control.²⁹ The guide is reflective of the high quality reputation of the Smithsonian as an institution and is representative of what the ideal standards might look like in practice in the real world.

²⁶ Jackson, Tracy M., "I Want to See It: A Usability Study of Digital Content Integrated Into Finding Aids," *Journal for the Society of North Carolina Archivists* (2012): 40.

²⁷ *Ibid.*, 40.

²⁸ *Ibid.*, 54-55.

²⁹ Aikens, Barbara, Erin Kinhart, Robin Holladay Peak, and Judy Ng, “Digitizing Entire Collections at the Archives of American Art,” *Archives of American Art*, October 27, 2011, Accessed November 5, 2016, <http://www.aaa.si.edu/collections/documentation/digitization>.

In her article “User Impact on Selection, Digitization, and the Development of Digital Special Collections”, Alexandra Mills makes the case for increasing the involvement of users in the selection of collections that will be digitized. Mills believes that because increasing user access is one of the primary reasons for digitization to take place, users’ considerations are vital to that spirit of that goal.³⁰ She describes a variety of ways to take into account users’ interests, including directly, by surveying or interviewing patrons or being open to digitizing on-demand, or indirectly by keeping up to date on the latest research on user demand, or analyzing internal data about collection use.³¹ Mills does, however, make the caveat that the selection process for digitization is complex, and a variety of other institutional factors come in to play besides user wants or needs. An archivist must be able to balance what has been learned about users with considerations of funding, mission, complexity, consistency, and so on.³² She also warns that while recent scholarship (much like Miller) indicates that users expect more and more (or even all) collections to be digitized, and users’ expectations should absolutely be taken into account, it behooves the archivist to educate users about the reasons collections have been chosen to be digitized or not digitized in order to manage those expectations.³³

Jeffery Beall, in his article “The Weaknesses of Full-Text Searching” creates a counterpoint to Miller’s advocacy for full-text searching as a replacement for processing in some cases. Beall outlines a number of reasons why relying solely on full-text searching can be problematic. The first reason he gives is that of synonyms. If users use search terms that may have the same meaning as text content, but different terminology they are limited in their

³⁰ Mills, Alexandra, "User Impact on Selection, Digitization, and the Development of Digital Special Collections," *New Review of Academic Librarianship* 21, no. 2 (2015): 161.

³¹ *Ibid.*, 163-165.

³² *Ibid.*, 165-166.

³³ *Ibid.*, 162-163.

results.³⁴ The same is true for terms with variant spellings or terminology, especially between British and American English (“color” vs “colour”), and the shortening of terms through such things as acronyms.³⁵ Beall also argues that homonyms can make search precision problematic. By searching for a term whose spelling has more than one meaning, a large amount of unwanted search results could occur, drowning out the intended results.³⁶ Other problems that may occur while searching is that users may not know a specific term that would best help in a full-text search, or that a document may be describing a certain topic the user wishes to know about, but never actually have the name of the topic explicitly in the text.³⁷ While Beall’s argument is not directly solely toward digitized archive collections, but rather the larger library database community, the principles he puts forth in his paper are still valid in the scope of the former.

A good place to start in discussing a theoretical methodology of incorporating digitization into processing is with the ideal best practices of both archival processing, and digitization. With regard to digitization, the Smithsonian Archives of American Art has published online, its guide for best practices in digitizing entire collections. All scanning is done at 400 to 600 PPI, with a set of standard meta-data at creation of the digital object, including file name, file format, dimensions, date of capture, camera serial number and firmware, white balance, make & model of camera, capture software, shutter speed, and ISO speed rating.³⁸ Interestingly, the guidelines indicate that item level description is not necessary, and that “[t]he EAD finding aid created by the processing archivist serves as the only descriptive metadata for the digital files associated

³⁴ Beall, Jeffrey, "The Weaknesses of Full-Text Searching," *The Journal of Academic Librarianship* 34, no. 5 (2008): 439.

³⁵ *Ibid.*, 440.

³⁶ *Ibid.*, 440.

³⁷ *Ibid.*, 442.

³⁸ Aikens, Barbara, Erin Kinhart, Robin Holladay Peak, and Judy Ng, “Digitizing Entire Collections at the Archives of American Art,” Archives of American Art, October 27, 2011, Accessed November 5, 2016, <http://www.aaa.si.edu/collections/documentation/digitization>.

with each fully digitized collection”.³⁹ Aikens et al. seem to have familiarity with the process of MPLP as the first major tenet of deciding what not to scan states:

If a collection is prioritized and scheduled for scanning, most of the material within the collection should be scanned. The decision not to scan materials should NOT be made on an item by item basis, but rather on the folder and/or series/subseries level. For the most part, it is simply easier to scan the materials than to flag and make notes for not scanning. However, not everything can or should be scanned in a collection.⁴⁰

Decisions on what to scan or not to scan, and how they are scanned seem, very much, to be in order with Greene and Meissner’s concept of MPLP, prioritizing less description at more detailed levels, and emphasizing saving time above all else.

When digitizing, however, other considerations must be taken into account other than just access. These other concerns may influence how decisions are made in digitization projects.

After access, one of the primary concerns of digitization is preservation. While it may be more efficient to photocopy original documents, then mass scan the copies at a low resolution as in the project at UW-Oshkosh⁴¹, and it may allow researchers access more quickly, there is another cost in the authenticity and preservation of the collections. If something were to happen to the originals, would the low quality copies of copies be sufficient replacements? Would future researchers be able to understand the documents in the same way or are there details they may be missing out on? The argument to provide access first, then worry about higher quality images defeats the purpose of the original goal of efficiency. As the old saying goes “measure twice, cut once”. Scanning documents twice, when a high quality project the first time would have been sufficient, or even preferred, makes little sense. Ranger’s or Kelly’s methodologies are too cavalier with the concerns of preservation and authenticity in favor of access. Therefore, the

³⁹ Ibid.

⁴⁰ Ibid.

⁴¹ Ranger, Joshua, “More Bytes, Less Bite: Cutting Corners in Digitization,” (presentation, Annual Meeting of the Society of American Archivist, San Francisco, CA, August 24-31, 2008).

theoretical digitization process outlined by this paper will prioritize the industry standards and best practices of high quality scans of 400-800 PPI saved to a lossless file format (PDF or TIFF), from which web copies may be created.

With this decision out of the way, it is evident that many of the aspects of MPLP can be applied directly to a digitization processing project to decrease processing time and increase access. As is exemplified in the American Archives of Art, it is perfectly possible to apply minimal description to collections. While some meta-data can be automatically generated upon creation of a digital object, other, more “subjective” meta-data can be applied a higher levels (such as the EAD finding aid for the collection) instead of at an item level.⁴²

As Miller suggests, Optical Character Recognition (OCR) and full-text searching may also offer a way to save time with detailed descriptions.⁴³ Beall, however, makes the case that full-text searching should not be a replacement for description due to the variety of factors stated above.⁴⁴ Miller, however, argues that full-text searching is often better than finding aids.⁴⁵ Ironically, Beall and Miller makes some similar arguments. Terms or subjects a user may be searching for may not be either in the text (as Beall states), or not in the finding aid/description (as Miller states). The problem then becomes trying to create the most access for the user, while simultaneously saving the archivist the most time and effort. A full-text search, in conjunction with detailed item level description could (theoretically) solve each problem. Allowing users to navigate through finding aids, subject headings, detailed descriptions and meta-data, and

⁴² Aikens, Barbara, Erin Kinhart, Robin Holladay Peak, and Judy Ng, “Digitizing Entire Collections at the Archives of American Art,” *Archives of American Art*, October 27, 2011, Accessed November 5, 2016, <http://www.aaa.si.edu/collections/documentation/digitization>.

⁴³ Miller, Larisa K., "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *The American Archivist* 76 (2013).

⁴⁴ Beall, Jeffrey, "The Weaknesses of Full-Text Searching," *The Journal of Academic Librarianship* 34, no. 5 (2008).

⁴⁵ Miller, Larisa K., "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *The American Archivist* 76 (2013).

searching the full-text of documents would create the ultimate access system, the best of both worlds for researchers. This, however, is not realistically feasible due to time and funding constraints. A reasonable compromise would be to utilize Miller's OCR and full-text searchability of the uncorrected document, coupled with the MPLP practice of description at the highest level feasible. Individual items need not be described, but brief descriptions of series may be desired. Again, this would be discretionary according to the project and the archivist. This is, of course, assuming that OCR on documents would even be possible, as the vast majority of digitization projects will likely involve collections that are not limited to a standardized type that would be OCR readable. Archival collections often include many photographs and handwritten documents that would not even be applicable to the situation.

While there can be no one-size-fits-all solution to balancing archival concerns in digitization projects, perhaps some general guidelines may be inferred from the scholarly literature and using common sense.

1. While digitization at large resolution into lossless file formats may not be the best choice in every situation, it should be the preferred standard if at all possible. This alleviates concerns of preservation and authenticity, allowing remote researchers to the best possible copy of an item, while maintaining efficiency in the future by saving collections from having to be digitized twice.
2. MPLP has worked efficiently to provide guiding principles in the arena of digitization in a variety of projects. The most important aspect of MPLP to follow during digitization project is that descriptions can be kept to a minimum, with little to no need for item level description, and providing detailed descriptions only where

- necessary. As Murphy et al. also pointed out, digital arrangement can be much more efficient, taking the place of physical arrangement.⁴⁶
3. While not always possible, performing Optical Character Recognition, and allowing full-text searching is a great way to add access functionality to a digitized collection without much effort by processing staff.
 4. As evidenced by Murphy et al., Ranger, Jackson, and Mills, user input is vital to the both the selection of collections to digitize, and just as importantly, what users expect, which will, in turn, dictate choices in how to perform digitization projects.

⁴⁶ Murphy, Mary O., Laura Peimer, Genna Duplisea, and Jaimie Frit, "Failure is an Option: The Experimental Archives Project Puts Archival Innovation to the Test," *The American Archivist* 78, no. 2 (2015): 441.

Bibliography

- Aikens, Barbara, Erin Kinhart, Robin Holladay Peak, and Judy Ng. "Digitizing Entire Collections at the Archives of American Art". *Archives of American Art*. October 27, 2011, Accessed November 5, 2016.
<http://www.aaa.si.edu/collections/documentation/digitization>.
- Beall, Jeffrey. "The Weaknesses of Full-Text Searching." *The Journal of Academic Librarianship* 34, no. 5 (2008): 438-444.
- Greene, Mark A. "MPLP: It's not Just for Processing Anymore." *The American Archivist* 73 (2010): 175-203.
- Greene, Mark A., and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68 (2005): 208-263.
- Jackson, Tracy M.. "I Want to See It: A Usability Study of Digital Content Integrated Into Finding Aids." *Journal for the Society of North Carolina Archivists* (2012): 20-77.
- Kelly, Elizabeth. "Processing Through Digitization: University Photographs at Loyola University New Orleans." *Archival Practice* 1, no. 1 (2014).
- Miller, Larisa K. "All Text Considered: A Perspective on Mass Digitizing and Archival Processing." *The American Archivist* 76 (2013): 521-541.
- Mills, Alexandra. "User Impact on Selection, Digitization, and the Development of Digital Special Collections." *New Review of Academic Librarianship* 21, no. 2 (2015): 160-169.
- Murphy, Mary O., Laura Peimer, Genna Duplisea, and Jaimie Fritz. "Failure is an Option: The Experimental Archives Project Puts Archival Innovation to the Test." *The American Archivist* 78, no. 2 (2015): 434-451.
- Ranger, Joshua. *More Bytes, Less Bite: Cutting Corners in Digitization*. Presentation at the Annual Meeting of the Society of American Archivist, San Francisco, CA, August 24-31, 2008.