

## Cost-effective Large Scale Online Access: The National Historical Publications and Records Commission's Digitizing Historical Records Grant Program

Lucy Barber and Rachael Saylor<sup>1</sup>

From 2006-2016, the National Historical Publications and Records Commission (NHPRC) of the National Archives and Records Administration ran a grant program called Digitizing Historical Records. These grants funded projects at archives to digitize large processed records collections while repurposing existing metadata. At the time, many archives felt that digitization projects were too expensive because of costs associated with creating item-level metadata. By offering an alternative mass digitization approach, which did not require any new descriptive metadata, the NHPRC hoped that archival repositories would expose and make accessible more of their collections online. In most cases, NHPRC projects digitized collections at the series or folder level. This report looks at the track record of the grant program based on project reports to the NHPRC and the results from a survey distributed to the grant recipients.<sup>2</sup> The hope is that the background on the results of these grant projects may guide repositories embarking on similar efforts.

### *Background on Archival Digitization Projects*

Digitizing historical records collections has been going on for some time before the NHPRC grant program. The Library of Congress created American Memory in 1990, which became the National Digital Library four years later. Agencies with an interest in the archival field—including the National Endowment for the Humanities and the Institute for Museum and Library Services—had regularly funded digitizing projects. But the NHPRC, with its limited funds, chose instead to make basic processing of collections a priority, along with support for electronic records research and development, among its grantmaking programs. The Commission had not regularly funded any digital reformatting unless it was necessary for preservation purposes because of outmoded machine readable formats.

The Commission's policy on digitization in the early part of the century was based, in part, on some fundamental questions about the cost-benefit value. In "Why Digitize?" a 1999 [Council on Library and Information Resources](#) report, Abby Smith noted:

---

<sup>1</sup> Thanks to the staff of the NHPRC especially Nancy Melley for her data tables and Keith Donohue for his editorial assistance.

<sup>2</sup> NHPRC Digitization Grant Follow-up Questionnaire. Files available from the NHPRC.

What we have found is that digitization often raises expectations of benefits, cost reductions, and efficiencies that can be illusory and, if not viewed realistically, have the potential to put at risk the collections and services libraries have provided for decades.<sup>3</sup>

In addition to emphasizing that digitization was not preservation, Smith's caution stemmed from the concerns about the costs of preparing materials in a method that would make the digital surrogate understandable to the users. She noted the costs of preparing objects for digitization including collected the metadata that she assumed users would need to understand each item. She concluded by encouraging institutions to develop criteria for digitization projects to ensure that valuable and sustainable resources would be prioritized. Fundamentally, the caution that appeared in works written before the start of the program rested on the observation made by one author "we are still learning about optimal digitizing methods."<sup>4</sup>

Another concern was the lack of interoperability among projects. In 2002, The National Initiative for a Networked Cultural Heritage (NINCH) released a "Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials" which began with the observation "Early developers of digital resources often had little thought for how their projects might dovetail with others."<sup>5</sup> The NINCH guide sought to provide guidance and resources to encourage future projects that would be built with more attention to sustainability and interoperability.<sup>6</sup> Sometimes, the caution surrounded questions of copyright of the original materials.<sup>7</sup> More commonly was concerns about the creation of digital resources without adequate appreciation for how technological changes might result in obsolescence or increased costs.<sup>8</sup> A 2003 survey of IMLS-funded digitization

---

<sup>3</sup> Abby Smith, *Why Digitize?* (CLIR, 1999) Available at <https://www.clir.org/pubs/reports/pub80-smith/pub80.html>.

<sup>4</sup> Steven D. Smith, "Cooperative Imaging: Scans Well with Others," in Maxine K. Sitts, Editor, *HANDBOOK FOR DIGITAL PROJECTS: A Management Tool for Preservation and Access* (Northeast Document Conservation Center, 2000). Accessed <https://www.nedcc.org/assets/media/documents/dman.pdf>.

<sup>5</sup> Humanities Advanced Technology and Information Institute, University of Glasgow and NINCH, "Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials," 2002 & 2003 Available from <https://www.researchgate.net/publication/31869587> *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*.

<sup>6</sup> Lopatin in her review of the state of digitization projects in 2006 also noted that most projects were not interoperable, with the exception of the model Collaborative Digitization Project; Lopatin, 282.

<sup>7</sup> Dan Hazen, Jeffrey Horrell, Jan Merrill-Oldham, *Selecting Research Collections for Digitization* (CLIR, 1998) <https://www.clir.org/pubs/reports/hazen/pub74.html>.

<sup>8</sup> Paul Conway, *Overview: Rationale for Digitization and Preservation* in Maxine K. Sitts, Editor, *HANDBOOK FOR DIGITAL PROJECTS: A Management Tool for Preservation and Access* (Northeast Document Conservation Center, 2000). Accessed

projects found that 86 per cent of respondents had used item-level metadata,<sup>9</sup> striking a note of caution over costs, despite a general enthusiasm for digitization as a way to increase access to collections.

In fact, the countervailing arguments for digitization were buttressed by increasing levels of expertise and technological advances, particularly via the Internet and the nature of web-searching which made resources more readily discoverable through search engine optimization. Scanning equipment had improved and was more affordable. As more archivists experimented with digitizing successfully, there was momentum to support digitization projects that would provide access to collections much like the researcher used collections in the research room. The Smithsonian Archives of American Art received funds in 2005 from the Terra Foundation with the explicit mission to digitize entire collections and make them available at the folder level. Their work provided inspiration for the NHPRC as it designed its program.<sup>10</sup>

Max Evans, then Executive Director of the NHPRC, articulated the framework for the Digitizing Historical Records projects in a series of talks that culminated in an article for the *American Archivist*, “Archives of the People, by the People, for the People.” Evans expressed the idea that repositories should begin to expose their collections on the web using the existing descriptive information and by digitizing complete sets of material at the folder level. He imagined a time when researchers could conduct their research much like in a research room at any time and in any location.<sup>11</sup> Similar ideas appeared the same year at a conference and subsequent report by OCLC on the importance of digitizing for special collections in the wake of mass digitization of books. The participants argued that new approaches be tried. Pointing to evidence that researchers increasingly expected to find their sources online, the authors of the report, Ricky Erway and Jennifer Schaffner endorsed the idea that “For collections known to be of great interest and that lend themselves to digitization, think about scanning the entire collection rather than making decisions about which bits.”<sup>12</sup> By 2010, Mark Greene and Dennis Meissner were arguing the management strategies behind “More Product, Less

---

<https://www.nedcc.org/assets/media/documents/dman.pdf>. Howard Besser, “Digital Longevity” in same.

<sup>9</sup> Laurie Lopatin, "Library Digitization Projects, Issues and Guidelines." *Library Hi Tech* 24:2 (2006), 281

<sup>10</sup> “About the Terra Foundation Center for Digital Collections,” February 12, 2005, <https://www.aaa.si.edu/inside-the-archives/about-the-terra-foundation-center-for-digital-collections>

<sup>11</sup> Max Evans, “Archives of the People, by the People, for the People” *American Archivist* 70:2 (Fall/Winter), 387-400

<sup>12</sup> Ricky Erway, and Jennifer Schaffner, “Shifting Gears: Gearing Up to Get Into the Flow. Report produced by OCLC Programs and Research”, 2007. Published online at: [www.oclc.org/programs/publications/reports/2007-02.pdf](http://www.oclc.org/programs/publications/reports/2007-02.pdf)

Process” archival projects should also apply to digitization projects.<sup>13</sup> An OCLC survey in 2010 found that nearly every institution had done at least one digitization project, though they tended to be time sensitive and very selective, highlighting individual items.

In this atmosphere, the Digitizing Historical Records program began with the first applications considered in November 2006. By the time the program was folded into another grant initiative in 2016, there had been 211 applicants and 47 grantees. Now, virtually every NHPRC-funded records project has some element of digitizing historical records to increase public access.

### *Results of Projects:*

#### *Size of Projects and Costs of Digitization*

A major goal of this grant program was to encourage applicant to use cost effective methods for digitizing the materials by foregoing expensive item-level metadata creation in favor of repurposing existing metadata for the descriptive part of the project.

Based on the grant reports, 39 projects have been completed to date. They range in scope from a large-format map project (which digitized 5,241 items) to a collaborative project on county court records (which digitized 884,475 items.) Overall, there were 4,684,409 digitized images created by the grantees, with an average of 120,113 images per project.

Federal funds also vary widely. The largest expenditure was \$150,000 (for the court records) to \$13,884 for an oral history digitization project at a university. The average amount of grant funds spent was \$64,622. In total, \$2,520,255 was spent by projects in NHPRC grant funds.<sup>14</sup>

Grantees were required to provide cost share contributions equal to at least 50% of the total project budget. They actually exceeded the requirement, providing a total of \$4,159,656, with an average amount of \$106,658. Federal dollars were a catalytic investment in these digitizing projects.

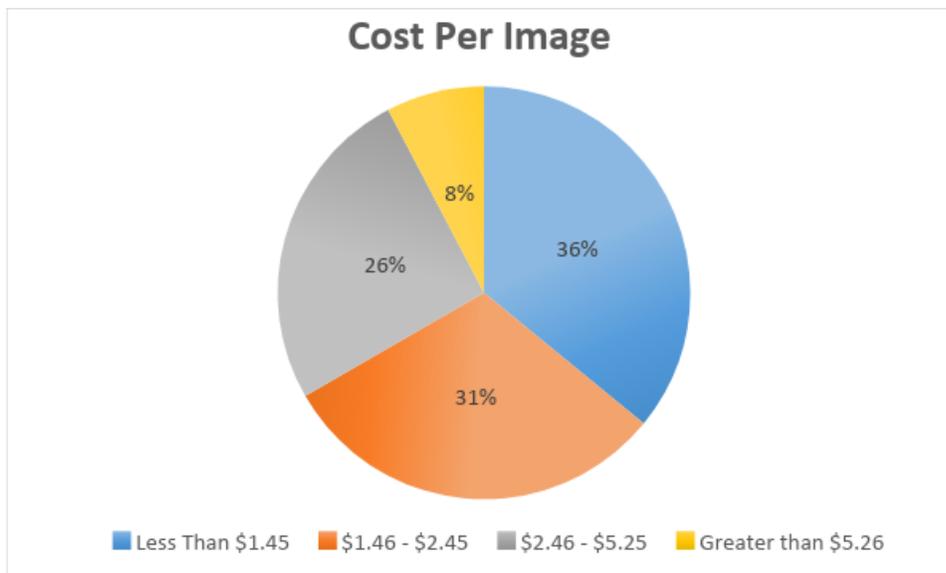
The average cost per image—including all grant funds and cost share contributions—was \$2.71 per image. Twenty-six projects had costs below this number, and 14 projects kept their costs below \$1.45 per image. The University of Alabama conducted a detailed study comparing the costs of

---

<sup>13</sup> Dennis Meissner and Mark A. Greene, “More Application While Less Appreciation: The Adopters and Antagonists of MPLP,” *Journal of Archival Organization* 8.3/4 (July–Dec. 2010).

<sup>14</sup> All figures are based on final reports on grant funds spent and cost share provided. Many projects spent less than the total amount of grant funds they had been awarded for these projects. The reason tended to be unexpected efficiencies in conducting the digitizing and quality control work by vendors or scanning technicians.

digitizing their Samuel Cabaniss collection for access through their finding aid with procedures used for digitizing at the item level. They concluded that the costs for the Cabaniss project were \$0.79 per scan in contrast to \$2.47 per scan for what had been their “normal” level of work, including the creation of item-level metadata.<sup>15</sup>

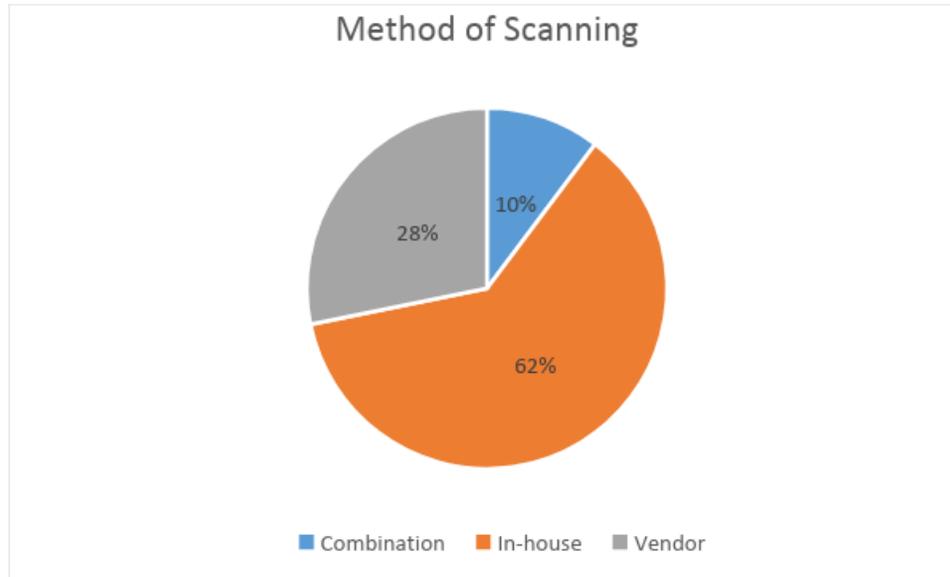


#### *Methods of Digitization: In-House v. Vendors*

The emphasis of the grant program on repurposing existing metadata and completing the projects in a cost-efficient manner affected the manner that institutions approached the practicalities of scanning archival materials. One crucial question was whether to digitize in house or to contract for the service. Of the completed projects, 62% did the scanning in house, sometime using distinct digital services units. 28% used vendors for the scanning. Four projects (10 %) used both vendors and in-house methods to scan their materials. Even when projects used vendors for digitizing, they were still responsible for managing the digital objects and creating the online displays.

---

<sup>15</sup> Jody L. DeRidder, Amanda Axley Presnell, and Kevin W. Walker, “Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis” *American Archivist* Vol. 75 (Spring/Summer 2012) 157.



Princeton University, which had included developing a sustainable digitization model as part of its project, attempted to compare the effectiveness of each approach. They had planned to use a contractor for some of the collections and then do other portions in house. They discovered that the vendor outsourcing required some initial investments of time to prepare the folders, but was otherwise more cost-efficient than the methods they tried in house. Including preparation, oversight, and indirect costs, the materials scanned by a vendor cost \$0.76 per page. For the same activities, when the project scanned from microfilm, they had per page costs of \$1.07. With a Zeutschel scanner, the cost went up to \$1.39 per page. The project director concluded:

If funds are available, vendor-supplied digitization significantly decreases the amount of staff time required to complete large-scale projects. The convenience of outsourcing is enhanced by its surprising cost effectiveness, a consequence of the relatively little oversight required for outsourced projects to run smoothly. More initial preparation of collections is required for vendor scanning, but, in all other respects, the time savings is significant for both student and professional staff.<sup>16</sup>

The University of California, San Diego also had success using a vendor for their digitizing project, though the calculated cost per page was significantly higher than Princeton at \$3.78. The University of Kansas for its Digitizing Kansas Sanborn Fire Insurance Maps, 1883-1922 project found the experience of using a vendor to be a “joy.” Because of the size and detail of the maps required more

---

<sup>16</sup> Final report, “Digitizing the Origins of the Cold War: Developing a Sustainable Digitization Model at the Seeley G. Mudd Manuscript Library,” NHPRC grant file RD-10118-13.

intensive scanning, costs were expected to be high. They digitized 5,241 maps at a cost per map of \$11.41.<sup>17</sup>

However, not all projects that outsourced digitization had such smooth experiences. The American Institute of Physics found that their vendor had significant quality control issues and that items often had to be redone up to four times. Because the Institute had contracted at a fixed price, this did not drive up the amount spent on digitization, but additional time was necessary for staff time conducting quality control.<sup>18</sup> This increased the amount of staff time was included in their cost-share contribution. The Railroad Commission of Texas also found that its vendor was challenged by the diverse size of materials in their collection and had to develop special fixes in order to render the images accurately.<sup>19</sup> The Archives of Michigan also experienced trouble with its vendor taking more time than predicted because the vendor had not counted on the fragility of the materials.<sup>20</sup> In general, it appears that those who confronted problem with vendors and digitization had contracted with companies that were not used to archival materials. In contrast, the Library of Virginia used a vendor that it had contracted with before on similar projects and found that “no unforeseen issues arose and the project completed successfully and ahead of schedule.”<sup>21</sup>

Projects that did the work in house also had to make adjustments to produce cost efficiency. A challenge for some projects was the allocation of staff time on equipment. All projects who conducted the scanning in house had to ensure that the scanning technicians were properly allocated. Georgia State University analyzed its methods in detail for handling the project of creating scans, creating the derivatives, running the scans through Optical Character Recognition (OCR) software, and loading them into the content management system. The scanner software combined with a project tracking spreadsheet allowed the project staff to divide the work among different assistants, reducing the amount of a bottleneck when the scanner was being used to create the more than 179,000 images in 2,510 distinct folders. It was still necessary to increase the pace of production of the project to complete the project on schedule. The solution was to adjust the schedules of the people working on the project so they worked outside of the traditional 8 am - 5

---

<sup>17</sup> Final report, “Digitizing Kansas Sanborn Fire Insurance Maps, 1883-1922,” NHPRC grant file RD-10032-10

<sup>18</sup> Final report, “Digitizing the Samuel A. Goudsmit Papers,” NHPRC grant file RD-10029-09.

<sup>19</sup> Final report, “Digitizing Historic Oil and Gas Hearing Files, 1932-1972,” NHPRC grant file RD-10019-09.

<sup>20</sup> Final report, ““Thank God for Michigan:’ Digitizing Historic Records,” NHPRC grant file RD-10005-07.

<sup>21</sup> Final report, “Augusta County, VA, Chancery Records, 1745-1912,” NHPRC grant file RD-10060-11.

pm, Monday-Friday work week. By assigning staff to earlier shifts and to weekend work, they added about 24 hours per week of availability for the single scanner.<sup>22</sup>

Troup County in Georgia had two scanners at their disposal, but limited staff as well. They found it advantageous to hire interns from a local college to ensure constant use of the equipment.<sup>23</sup> The University of Alabama staff involved with the Septimus D. Cabiness project noticed early in the project that digitization rates by student workers were slower than had been predicted. They staff emphasized the need for an increased rate of scanning and shifted some of the scanning operation to a faster piece of equipment to meet the required production rate.<sup>24</sup> At Cornell University, staff noted that student assistants tended to perform better when assigned shorter shifts.<sup>25</sup> Slow image processing of negatives at the Getty Research Institute inspired creative new uses of equipment. The Lead Imaging Technician “explored an alternative direct capture process and designed what he called a ‘scamera’: a camera and digital back captured transparencies on light box.” This different technique, which is still in use at the Getty, allowed the project to increase from scanning 6-12 transparencies an hour to scanning 40 images an hour.<sup>26</sup> The purchase of an i2s Copibook scanner helped the University of Minnesota project to digitize the American Social Health Association records at a much faster rate than expected. They noted some variability among their students’ scanning rates with some scanning 126 images per hour to a high of 242 images per hour.<sup>27</sup>

Turnover in staff could also contribute to delays. A number of projects had to ask for extensions because of staff turnover. Some found it difficult to retain students for longer than six months, requiring the advertising of positions again.<sup>28</sup> Some, however, benefited from the costs and efficiencies of student workers. The University of Washington noted in its final reports that “an exceptional student was ended up doing the majority of the work. She was able to use a significantly more efficient workflow . . . Her method involved performing two different task simultaneously, where while the scanner was working on one image, she had time to upload the previously-completed image to CONTENTdm.”<sup>29</sup> At the University of Massachusetts, Amherst, the

---

<sup>22</sup> Barbara Petersohn, Traci Drummond, Melanie Maxwell, and Kelly Pepper, “Resource Leveling for a Mass Digitization Project,” *Library Management* 34:6/7 (2013), 493-495.

<sup>23</sup> Kaye Lanning Minchew, “Digitizing Entire Collections in Georgia,” *Archival Outlook*, September/October 2008, p. 6.

<sup>24</sup> Final report, “Digitizing the Septimus Cabiness Papers,” NHPRC grant file, RD-10033-10.

<sup>25</sup> Final report, “Cornerstones of the American Middle Class: The Historical Collective Bargaining Agreements Project,” NHPRC grant file RD-10150-14.

<sup>26</sup> Final report, “Living the American Dream: Housing and Urban Development in Los Angeles, 1936-1997,” NHPRC file RD-10072-11.

<sup>27</sup> Six month interim project report, August 1, 2009 - January 31, 2010, “Digitizing the Historical Records of the American Social Health Association,” NHPRC grant file RD-10023-09.

<sup>28</sup> Petersohn et al, 495.

<sup>29</sup> Final Report, “J. Willis Sayre Photographs of Performers Digitization Project.” RD-10083-12.

availability of work-study students, whose salaries were subsidized, to conduct the scanning meant that the project returned a substantial amount of funds at the end of the project period. Efficiency as the grant staff grew more experienced were also noted at the Minnesota Historical Society (MHS) and contributed to the project scanning 101,936 images in comparison to the 32,000 they had estimated at the beginning of the project.<sup>30</sup>

Efficiency at the MHS project was also gained because it was one of the few projects able to use sheet feeding to digitize the archival records. The project director made the judgement that certain of the Hubert Humphrey speech files were in good enough order on strong enough paper that they could be digitize with the feeder.<sup>31</sup> Princeton tested sheet feeding for one of the collections in their project where they had duplicates of the collection in question. Unfortunately, the equipment they were using could not be set to the necessary standards for creating sufficiently high quality images, so the collection was digitized by a vendor instead.

Some grant recipients also had to decide on policies for dealing with items with sensitive or copyright information. At Texas Tech University, the “Digitizing the Vietnamese-American Immigration Experience” project discovered that there were social security numbers and medical records interspersed with material being scanned. They quickly trained their student scanners to redact such sensitive information from the facsimiles.<sup>32</sup> Georgia State University in digitizing the PATCO records made the decision to scan just the first page of publications that might still be in copyright. In this manner, they continued to give the researcher the context of the original material, but avoided infringing on the rights of others. They also informed all users that if they found material that they believed was personal in nature, the users could apply for it to be removed. Such takedown policies were common at institutions whose records dated to the later part of the 20th century.

#### *Availability of Digitized Materials: Reactions and Methods*

One of the conditions of the NHPRC grant was that the grantee’s digitized images would have to be available without charge after completion. Based on a survey conducted of grant recipients and the grant reports filed at the end of project; this goal was achieved. Of the respondents to the survey, all but one said the material was available on the repository’s website. The exception was a project that instead made the scans available through the Connecticut History Online. (Other projects also

---

<sup>30</sup> Minnesota Historical Society

<sup>31</sup> Final report, “Digitizing the Hubert H Humphrey Speech Text Files,” NHPRC grant file RD-10090-12.

<sup>32</sup> Final report, “Digitizing the Vietnamese-American Immigration Experience,” NHPRC grant file RD-10097-12.

contributed materials to statewide repositories, including the Digital Library of Georgia. The Herman Baca project at the University of California, San Diego made digital links available from the finding aid on the Online Archive of California.)

Based on the questionnaire distributed to grantees, most projects report a favorable response to the availability of the digitized collections. We asked “How would you rate the response from the people who use your digital collection?” More than three quarters of the responding institutions (77.8%) have received a very positive response to their digitization project, and an additional 14.8% have experienced a mostly positive reception. While no one reported any negative response, one respondent (3.7%) reported a neutral or mixed response, and another (3.7%) responded, “We have not collected responses from people using the digital collection, but use of the collection has been high. So maybe that is in itself a positive response?”<sup>33</sup> A few projects conducted surveys with users as part of their projects. Duke University actively solicited response to the ROAD 2.0 digitized images. The responses were sometimes very enthusiastic. “Thank you for this wonderful resource” and “This website is awesome” were responses to the Duke site. But users also shared opinions about desired features and slow responsiveness which grantees had to take into account when responding to the NHPRC survey.<sup>34</sup>

Digitizing at this scale and in this manner resulted in organizations testing innovations. Many were in the methods of generating the descriptive metadata for linking to the digitized images. Some converted finding aides to spreadsheets and then used those to attach the file names of the scans before ingesting into a content management system.<sup>35</sup> For the Troup County project, the Digital Library of Georgia created Perl scripts to generate Dublin Core metadata.<sup>36</sup> Likewise, the University of Alabama constructed several Perl scripts to help with conversion and uploading of materials for their Cabaniss project; they were also working simultaneously on an open source collections management system.<sup>37</sup> Marist College needed a method to convert its detailed Word based finding aid to EAD. To make this task possible for students with little experience with encoding, they built tool that made this process intuitive for student workers. They also used jQuery to develop an interface where users could expand and collapse series.<sup>38</sup> One project encouraged the addition of

---

<sup>33</sup> NHPRC Questionnaire to Digitizing Historical Grantee Recipients.

<sup>34</sup> Final report, “ROAD 2.0: Digitizing Outdoor Advertising,” NHPRC grant file RD-10017-09.

<sup>35</sup> See, for example, Interim Report, January 30, 2015, “Digitizing American Soviet Jewry Movement Collections,” NHPRC grant file RD-10153-14.

<sup>36</sup> Minchew, p. 6.

<sup>37</sup> Final report, “Digitizing the Samuel Cabaniss Papers,” NHPRC grant file: RD-10033-10.

<sup>38</sup> Monish Singh, “EADfying Lowell Thomas Papers,” Encoded Archival Standards Section, Society of American Archivists. Accessed:

<https://www2.archivists.org/groups/encoded-archival-standards-section/eadfying-lowell-thomas-papers>.

crowd-sourced transcriptions to enhance the discovery of resources. At the Abraham Lincoln Presidential Library and Museum, when the institution digitized the records of an Illinois governor, they turned to a tool called Scripto from the Center for History and New Media at George Mason University to permit people to transcribe the materials online and then have the transcriptions searchable. Three hundred people registered. In approximately six months, the volunteers transcribed more than 3,000 documents.<sup>39</sup> The University of Florida developed a tagging feature to use with their digitized collection and planned to work with history professors to have them assign tagging exercises to their students.<sup>40</sup>

For projects dealing with modern typescript, the use of optical character recognition increased the ability of researchers to find material by keyword. The University of Iowa noted that for its Henry Wallace Papers project “we repurposed metadata from the microfilm inventory and performed optical character recognition (OCR) scanning on the digital images to allow full text searching.”<sup>41</sup> Similar approaches were used by the University of Delaware in its project to digitize the George S. Messersmith Papers, the Rutgers Law School Library project on Housing Policy and Reform, Georgia State University for its Professional Air Traffic Controllers Organizations (PATCO) Records project and the Minnesota Historical Society for its “Digitizing the Hubert H. Humphrey Speech Text Files. The University of Minnesota noted that its original plan to use a desktop version of OCR software proved too inefficient, so they invested in a server side product that meant they had an enterprise-level solution.<sup>42</sup> For its Collective Bargaining Agreements digitizing project, Cornell University staff noted that some originals were of poor legibility or handwritten, so OCR could not work. They are considering transcribing the hand-written agreements as a separate project.<sup>43</sup> In general, though for projects with typescript materials, the use of OCR increased the searchability of the digitized images. This approach has been endorsed by an article in the *American Archivist* by Larisa Miller who argued digitization and OCR of modern archival collections would be an effective way to prevent further backlogs and make collections key word searchable.<sup>44</sup>

---

<sup>39</sup> Final report, “Digitizing the Richard Yates Papers,” NHPRC grant file RD-10139-13

<sup>40</sup> University of Florida Everglades

<sup>41</sup> Final report, “Henry A. Wallace Digital Collection Project,” NHPRC grant file RD-10020-09.

<sup>42</sup> Final report, “Digitizing the Historical Records of the American Social Health Association,” NHPRC grant file RD-10023-09.

<sup>43</sup> Final report, “Cornerstones of the American Middle Class: The Historical Collective Bargaining Agreements Project,” NHPRC grant file RD-10150-14

<sup>44</sup> Larisa Miller, “All Text Considered: A Perspective on Mass Digitizing and Archival Processing,” *American Archivist*, 76:2 (Fall/Winter 2013), 521–541.

## *Usage Statistics*

Usage statistics were not always available in grantees' final reports because the digital collections were only made available towards the end of the project period. But when asked in the follow up survey about whether they had noticed any change in the number of people who access the records (either in original or digital form), the overwhelming majority of respondents (96.3%) indicated that they had noticed an increase in access. Final reports mostly supported this point of view. At Texas Tech University, the archives digitized over 350,000 pages for its 18,835 files in the collection of the Families of Vietnamese Political Prisoners/Vietnamese American Heritage Project Collection. These files were mainly drawn from the series of the Orderly Departure Program Application, which documents efforts to support the migration of Vietnamese people held in reeducation camps. The project reported that by the end of the project over 5,000 searches in the collection and the accessing of 46,430 pages of the collection. In addition, researchers downloaded files nearly 19,000 times.<sup>45</sup> The University of Kansas reported that by the final report that the Kansas Sanborn Map Collection was "the most popular collection available" on their content management site. They contrasted the average 76 maps sets that had been examined each year in the research room to the 12,999 page views of the digital images from just four months after digitization was complete in 2011.<sup>46</sup> After posting digital images of Supreme Court Cases from the 1820s through the Civil War, the Missouri State Archives reported that the number of views of their Supreme Court of Missouri Historical Records database increased from a monthly average of 4,800 web hits to 37,000 web hits each month.<sup>47</sup> The Library of Virginia found that adding the Augusta County chancery records resulted in an increase of 22.3% in the use of a Chancery Record Index site.<sup>48</sup> In just over two years, Marist College found that average monthly use of the Lowell Thomas Graphic Materials series had increased by 377% to 4,736 users per month.<sup>49</sup> On the other hand, Princeton University reported that page views of the Allen W. Dulles Papers finding aid had not increased a year after the digital images were linked to them.<sup>50</sup>

Three projects have publicly accessible longitudinal statistics on use since the digitized collections were released on the web. The University of North Texas makes available statistics on its collection

---

<sup>45</sup> Final report, "Digitizing the Vietnamese-American Immigration Experience," NHPRC grant file RD-10097-12.

<sup>46</sup> Final report, "Digitizing Kansas Sanborn Fire Insurance Maps, 1883-1922," NHPRC grant file RD-10032-10.

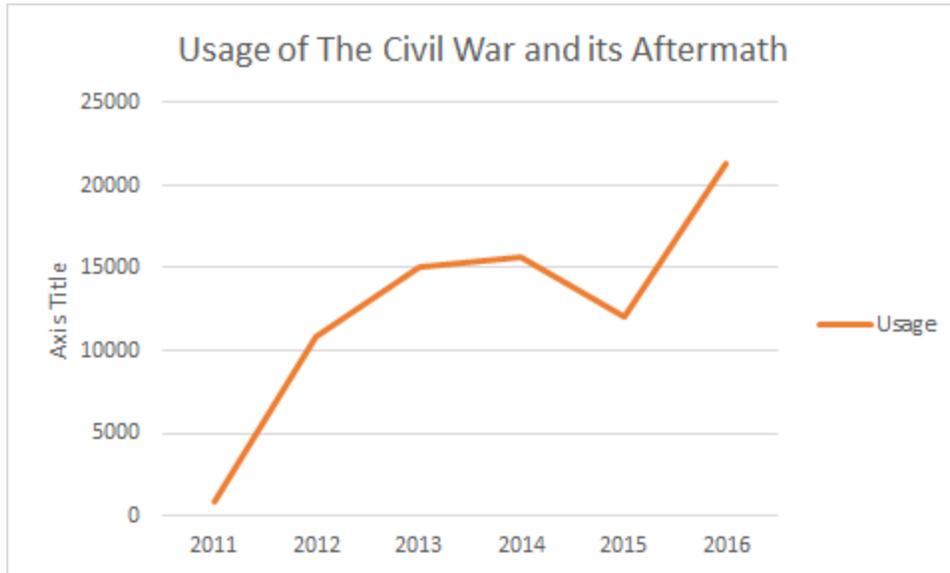
<sup>47</sup> Final report, "Supreme Court of Missouri Case File Digitization Project, Early Statehood through the Civil War, 1821-1865," NHPRC grant file RD-10051-10.

<sup>48</sup> Final report, "Augusta County, VA, Chancery Records, 1745-1912," NHPRC grant file RD-10060-11.

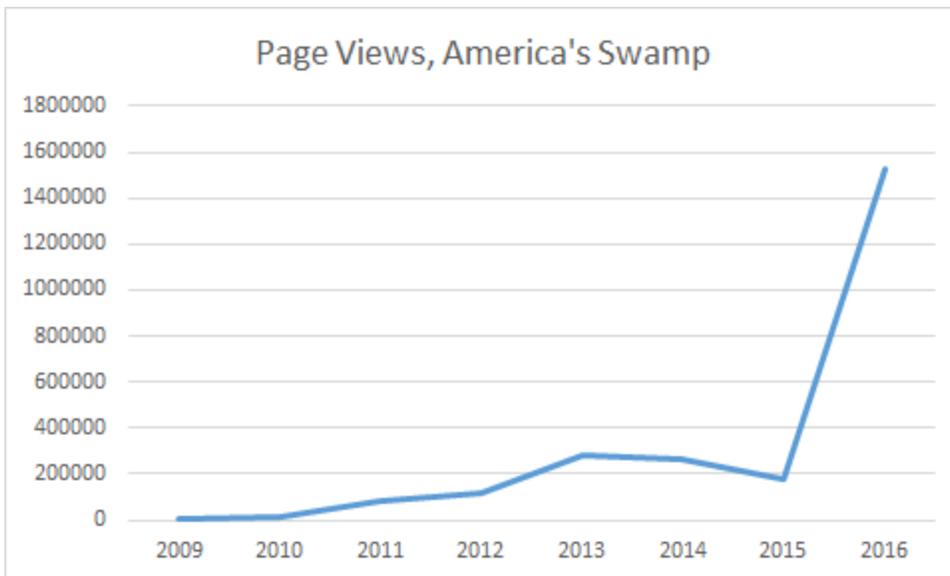
<sup>49</sup> Final report, "Lowell Thomas Papers Digitization Project," NHPRC grant file RD-10067-11.

<sup>50</sup> Final report, Digitizing the origins of the Cold War: Developing a Sustainable Digitization Model at the Seeley G. Mudd Manuscript Library," NHPRC grant file RD-10118-13.

“The Civil War and its Aftermath.” At the project beginning, there were just 855 users, by 2016 that had grown to 21,367.<sup>51</sup>



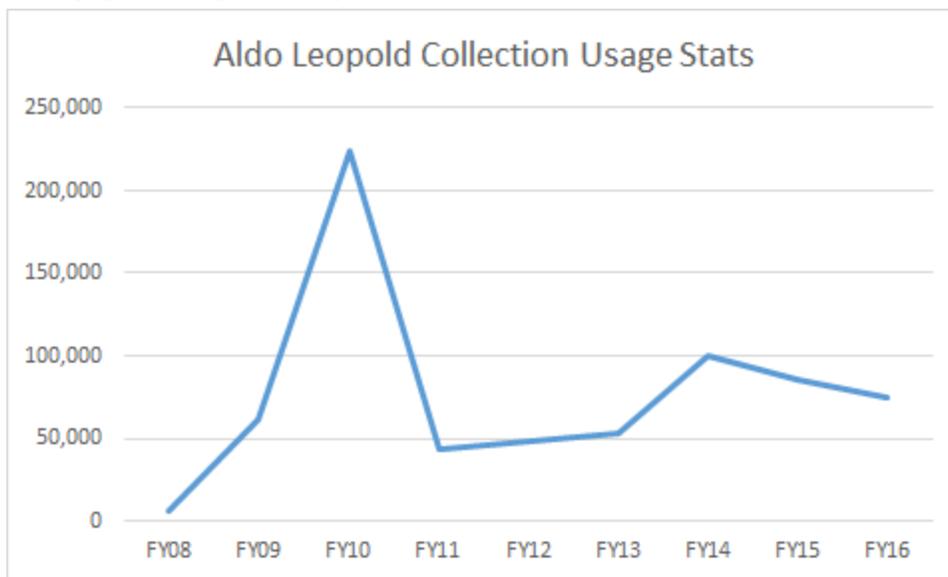
The America’s Swamp project on collections related to the Everglades also continues to reports usage statistics. The University of Florida combined seven collections into a virtual collection of documents concerning a variety of issues related to managing and studying the Everglades. Though use averaged 24,000 page views from 2013-2015, it jumped considerably in 2016 to 1,527,017.<sup>52</sup>



<sup>51</sup> <https://texashistory.unt.edu/explore/collections/CWADP/stats>

<sup>52</sup> <http://ufdc.ufl.edu/stats/usage/history/SWAMP>

The Aldo Leopold collection demonstrates the increased use of the collection online once the digital content was available. The chart indicates page views by fiscal years. Before there was digitized content, the finding aid received under 6,000 views. Once all the digital content was available in 2010, there was a spike of views to over 200,000. In subsequent years, the usage has ranged from 43,000 to 99,000 page views per fiscal year.<sup>53</sup>



Although these three examples illustrate the variability of usage statistics, they, like most of the statistics from project’s grant files indicate an increase in the use of the online materials. It is a fair assumption that the numbers of online visitors exceed the number of in-person researchers for these collections previously.

At some institutions, the ways in which researchers accessed the materials changed because of the digitization projects according to the survey we distributed. We asked “Are the original records also available for researchers to use, or do researchers use the digitized versions exclusively?” Just under two-thirds (63%) of the institutions still use the original records in addition to the digital images, while just under one-third (29.6%) use the digital collection exclusively. The two respondents who selected “other” here indicated that the digital images are used initially, but that the originals would still be made available to users if necessary. The substitution of digital collections should not be considered a strict preservation methods, but it does suggest that these kind of projects can contribute to a reduction in wear and tear on originals while still increasing access.

#### *Outreach connected to the Digitized Collections*

Projects used these collections to undertake a variety of forms of outreach. In the follow-up survey, we asked “Have you conducted any outreach projects or educational programs related to the digital

<sup>53</sup> University of Wisconsin Digital Collections, Usage Statistics, <https://www.library.wisc.edu/zero/uwdcc/usage-statistics/>

collection?” More than half (55.6%) of the responding institutions have performed some sort of education or outreach, including creating classroom assignments/projects, working with other organizations, and designing promotional material. At the University of Florida, for its America’s Swamp project, project staff designed four lesson plans for different grade levels to take advantage of the digitized materials.<sup>54</sup> The University also developed lesson plans for their subsequent project on “Pioneer Days in Florida.” For this project, staff reported that seven research projects based on the collection were in progress and students had completed five transcription projects. The collection received a Primary Source Award for Access from the Center for Research Libraries.<sup>55</sup> The Herman Baca project staff at the University of California, San Diego took the collection on the road, attending an outdoor event for the Chicano movement. They demo-ed the site and handed out buttons with Baca’s image.<sup>56</sup> The University of Iowa sent a flyer publicizing the Henry Wallace Collection to 1,600 schoolteachers to encourage use in school projects.<sup>57</sup> Marist College used the digitization of the graphic materials in the Lowell Thomas Papers to spur an on-campus travel photography contest, which attracted over 350 submissions from the student community.<sup>58</sup> Bowdoin College held a forum for faculty at the completion of its Oliver Otis Howard Papers project that highlighted how the digitization of the material opened up new possibilities in the digital humanities. In a subsequent course, parts of the collection were used for student projects in an Interactive Data Visualization class.<sup>59</sup>

#### *Facilitating New Accessions and Initiatives*

The projects had effects outside of increasing access to the digitized materials. In our survey, we asked “Has the digitization project you completed inspired your organization to undertake any related or similar projects?” 81.5% replied that it had spurred further work, including more digitization, processing, tagging, and system development. Other effects included that the repositories acquired new related collections because the materials were made more accessible. In the case of Aldo Leopold Foundation, additional donations were made to their collaborator, University of Wisconsin, Madison, of Leopold related materials in time to include them in the digitization project. Others found that providing online access to collections attracted the attention of donors. An event at the University of Massachusetts, Amherst related to the Horace Mann Digitization Project resulted in contacts with the Africa America Institute, which agreed to donate its

---

<sup>54</sup> Final report, “America’s Swamp: The Historical Everglades Project,” NHPRC grant file RD-10018-09.

<sup>55</sup> Final report, “Pioneer Days in Florida,” NHPRC grant file RD-10114-13.

<sup>56</sup> Final report, “A Time for Resistance: Chicano Activism in San Diego and the American Southwest, 1964-2006,” NHPRC grant file RD-10086-12.

<sup>57</sup> Final report, “Henry A. Wallace Digital Collection Project,” NHPRC grant file RD-10020-09.

<sup>58</sup> Final report, “Lowell Thomas Papers Digitization Project,” NHPRC grant file RD-10067-11.

<sup>59</sup> Final report, “Oliver Otis Howard Digitization Project,” NHPRC grant file RD-10149-14.

records to the University.<sup>60</sup> Similar occurrences took place at Atlanta Fulton Public Library, Texas Tech University, Cornell University, and the University of Florida.

The projects also convinced institutions to invest in future projects and influenced the nature of subsequent efforts. The Missouri State Archives decided to continue to scan Supreme Court Files after completing the grant-funded project focused on the earliest files.<sup>61</sup> Likewise, the Railroad Commission of Texas was able for a while to add to the historical Oil and Gas files collections that they had started with the grant. The Railroad Commission also changed their current business projects so that new filings were digitized.<sup>62</sup> Georgia State University realized after completing their PATCO project that they should invest in hiring additional permanent staff to conduct future projects. The University of Iowa digitized a related collection of microfilm.<sup>63</sup> Taking on this approach to digitizing inspired some repositories to build more robust digitizing programs. For their “Digitizing the Samuel Goudsmit Papers” project, the American Institute of Physics contracted with an outside vendor because the project was beyond their internal capacity. By the end of the project, the institution had invested in scanning equipment and was prepared to start a scan on demand program.<sup>64</sup> The University of Alabama began applying the techniques and workflows used in their project to subsequent digitization endeavors.<sup>65</sup>

#### *Usability and Concerns about the Quality of Description*

Repurposing existing metadata for the descriptive information of the digitized collection sometimes generated concerns about the discoverability and usability of the digitized materials. Depending on the detail of the existing finding aids, folder titles could be rather opaque. Correspondence without any indication of creator is one example. Institutions took various approaches. Some decided to rearrange the collections and improve the finding aids before embarking on the digitization. The University of Massachusetts, Amherst for its Horace Mann Bond collection rearranged the correspondence from chronologically arranged to an order based on individual correspondents.<sup>66</sup> The University of California-San Diego when faced with folders labeled Correspondence that belonged to sub-series representing organizations, copied the sub-series information to the

---

<sup>60</sup> Final report, “Horace Mann Bond digitization project,” NHPRC grant file RD-10101-12.

<sup>61</sup> Final report, “Supreme Court of Missouri Case File Digitization Project, Early Statehood through the Civil War, 1821-1865,” NHPRC grant file RD-10051-10.

<sup>62</sup> Railroad Commission response, NHPRC Digitization Grant Follow-up Questionnaire, April 21, 20176

<sup>63</sup> Final report, “Henry A. Wallace Digital Collection Project,” NHPRC grant file RD-10020-09.

<sup>64</sup> Final report, “Digitizing the Samuel A. Goudsmit Papers, NHPRC grant file RD-10029-09.

<sup>65</sup> Final report, “Digitizing the Samuel Cabaniss Papers,” NHPRC grant file: RD-10033-10.

<sup>66</sup> Final report, “Horace Mann Bond digitization project,” NHPRC grant file RD-10101-12.

Correspondence folder title to create more descriptive trails.<sup>67</sup> The Schlesinger Library for their Those Extraordinary Blackwells planned to continue to add descriptive “tags” at the item level to improve the searchability of the collection.<sup>68</sup> At the University of Florida, archivists found that repurposing subject terms from their library catalog reflected changing practices in term selection and did not fit well with the use of a faceted browsing display. The Pioneer Days project had the unintended benefit of encouraging the establishment of a more standardized approach to such metadata.<sup>69</sup> In some cases, project staff admitted that they wished to improve on the description outside the bounds of the original grant program design. For a photographic collection that emphasized Chicago’s built environment, the project director speculated that adding geo-references to the collection would increase the value of the digitized materials.<sup>70</sup>

The University of Alabama conducted the most systematic usability study of its digitization project. They conducted focus groups with users who explored the Cabaniss Collection digitized in the context of its EAD finding aid and another collection that had item-level description and could be accessed through a search box. The results of the study suggested that there were significant tradeoffs in usability of the Cabaniss collection, particularly for novice researchers and those with English as a Second Language. Still, some of their studies suggested that ease of use increased over repeated use of the Cabaniss materials.<sup>71</sup> They concluded:

A drawback, however, is that this method of Web delivery may currently be more suitable for scholars than for students. Longitudinal research needs to be performed to clarify the learnability of the finding aid as a Web interface to digitized items for novice users. Additionally, further studies need to be performed to determine what modifications could be made to the finding aid to increase usability for this population.<sup>72</sup>

More detailed research is necessary to weigh the advantages and disadvantages of providing access to collections with minimal metadata. It may be that the collections that use finding aids for discovery have drawbacks, but the Alabama usability study did suggest that users could complete research tasks with such collections with more time. For distant researchers who would otherwise have to travel to visit the repository, evidence suggest the trade-off may be worth it.

---

<sup>67</sup> Final report, “A Time for Resistance: Chicano Activism in San Diego and the American Southwest, 1964-2006,” NHPRC grant file RD-10086-12.

<sup>68</sup> Final report, “Those Extraordinary Blackwells,” NHPRC grant file RD-10126-13.

<sup>69</sup> Final report, “Pioneer Days in Florida,” NHPRC grant file RD-10114-13.

<sup>70</sup> Final report, “Chicago: Photographic Images of Change,” NHPRC grant file RD-10057-11

<sup>71</sup> Jody L. DeRidder, Amanda Axley Presnell, and Kevin W. Walker, “Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis,” *American Archivist* Vol. 75 (Spring/Summer 2012) 143–170.

<sup>72</sup> DeRidder et al, 169.

## *Conclusions*

This review of the Digitizing Historical Records grant program helps to highlight what can be accomplished when grant recipients tried new approaches to digitization. None of the grant recipients were practiced in digitizing in this manner when awarded the grants. They had to try new methods and develop new workflows. Challenges arose from difficulties with vendors and determining the most efficient use of staff and equipment. But the results in terms of cost efficiencies seem undeniable.

The most important results came in the indication of increased access to materials. The usage statistics, where available, indicate that researchers turned to these digitized collections in numbers far beyond what could be expected in most research rooms. Users' needs were met. The Aldo Leopold project collected anecdotal responses to the release of the digitized collection. A scholar reported "The online Aldo Leopold Archives are incredibly valuable resource for scholarly research. They have proven vital to me in my work on a new book about Aldo Leopold, allowing me access to any and all documents without having to travel to the physical archive." Another noted he was in a remote location and was able to read various journals. He exclaimed "It all seems a miracle."<sup>73</sup> In a similar vein, after the University of Delaware advertised that it was digitizing the George Messersmith Papers, archivists could fulfill reference requests from distant researchers with the digitized images.<sup>74</sup> Though usability studies at the University of Alabama suggest that not all users are going to access the materials with the same ease, the collections are now available for research. The positive effects of making materials available online also carried over to the repositories which took part. Some acquired new collections; others invested in further digitization. By enabling these institutions to engage in large-scale digitization in a cost-effective manner, the NHPRC fulfilled its mission to make historical records accessible to the American people.

This kind of digitization at scale does not necessarily represent the last work that will be done with these archival collections. By making them available online, the repositories have the opportunity to continue to assess if more detailed access points are required to satisfy the needs of certain researchers or to improve discoverability through search engines. Cross-collection searching which had been seen as a reason for hesitancy about embarking on digitization projects in the early 2000s remains a challenge. The rise of initiatives, like the Digital Public Library of America, that aggregate items from multiple locations are most functional when the metadata is at the item level. For the DPLA, it is harder to integrate materials digitized at the series or folder level as was done in many of these Digitizing Historical Records projects. However, researchers used to the diligence required to

---

<sup>73</sup> Final report, "Digitizing the Aldo Leopold Archives," NHPRC grant file RD-10006-07.

<sup>74</sup> Final report, "Moving Democracy Forward: Digitizing the George S. Messersmith Papers for Worldwide Access," NHPRC grant file RD-10046-10.

review materials at the folder level can now with these digitized collections scan for the materials they need online without the constraints of visiting a particular archive. In addition, in the NHPRC's newest grant program, "Access to Historical Records: Major Initiatives," we are encouraging applicants to develop collaborative projects that will bring together collections across repositories. Such virtual unification will increase the access that the public will have to notable collections.