2019

Microfilm, Manuscripts, and Photographs: A Case Study Comparing Three Large-Scale Digitization Projects

Emily Lapworth
University of Nevada, Las Vegas, emilyannelapworth@gmail.com

Sarah Jones
University of Nevada, Las Vegas, jones.sarahrhiannon@gmail.com

Marina Georgieva
University of Nevada, Las Vegas, marinik@abv.bg

Follow this and additional works at: https://elischolar.library.yale.edu/jcas

Part of the Archival Science Commons

Introduction

Libraries, archives, and museums have become accustomed to and adept at creating digital collections of individually described items. Digitization projects and digital collections of cultural heritage materials continue to be supported and used by a variety of audiences with unique needs. Some users may have the illusion that "everything" is available online and therefore there is no need to track down physical resources. Others may know that this is not true, but still have this expectation—after all, there are already vast collections of digitized materials available freely and publicly online.

Although not everything can be digitized and put online, cultural heritage institutions can scale up their operations in a variety of ways. The Digital Collections department at the University of Nevada, Las Vegas (UNLV) Libraries has completed several large-scale digitization pilots in an effort to test and implement specific strategies for increasing efficiency and output. Three of these projects are described in this article to show examples of different approaches to large-scale digitization: (1) the Nevada Digital Newspaper Project (part of the National Digital Newspaper Program), (2) the Culinary Union Workers Local 226 Photographs collection, and (3) the Entertainment Project. The UNLV projects digitized: (1) microfilmed historic newspapers, (2) photographic prints and negatives, and (3) mixed-materials manuscript collections.

In this article, different aspects of UNLV's large-scale digitization projects are discussed and compared: roles and responsibilities; grant funding; copyright, privacy, and confidentiality; arrangement; formats; and metadata. Although every institution, collection, and project is unique, this article provides examples of a variety of approaches and strategies that can be employed to increase digitization. Implementation of large-scale digitization can help archives and special collections meet user needs and expectations for online access to cultural heritage materials.

Literature Review

For over two decades, digitization of special collections materials was performed in a "boutique" style: items were hand-selected and intensively described. Many large-scale digitization efforts were focused on bringing book collections into the digital world, and the transition to large-scale digitization of archival collections was slow. OCLC's "Shifting Gears: Gearing Up to Get in the Flow"[1] appeared in 2007 and was one of the first articles advocating for large-scale digitization of special collections materials. In 2010, OCLC surveyed special collections institutions and found that 38 percent of 169 respondents had undertaken large-scale digitization, prompting the authors to call for action to develop models for large-scale digitization.[2] That same year, Mark Greene suggested that the archival theory of MPLP (More Product, Less Process) should be applied to digitization to increase the availability of archival resources on the Web,[3] and Oya Rieger outlined some of the challenges and benefits of applying large-scale digitization to special

---

[1] Ricky Erway and Jennifer Schaffner, "Shifting Gears: Gearing Up to Get Into the Flow," OCLC Programs and Research, 2007, https://www.oclc.org/content/dam/research/publications/library/2007/2007-02.pdf.

[2] Jackie M. Dooley and Katherine Luce, "Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives," OCLC, 2010, https://www.oclc.org/content/dam/research/publications/library/2010/2010-11.pdf.

[3] Mark A. Greene, "MPLP: It's Not Just for Processing Anymore," *The American Archivist* 73, no. 1 (April 1, 2010): 175–203, https://doi.org/10.17723/aarc.73.1.m577353w31675348.

collections materials.[4] Articles detailing how to optimize resources for large-scale digitization,[5] principles for managing vendor relationships,[6] and case studies on rapid-capture equipment[7] followed as large-scale digitization of special collections materials ramped up across the profession.

Literature specifically about the large-scale digitization of photographic and manuscript formats is available but not comprehensive in scope. The Southern Historical Collection described the establishment of its large-scale manuscript digitization program[8] and the Archives of American Art published its institution-specific workflows for digitizing entire manuscript collections.[9] In 2012, Shan Sutton compared vendor digitization of microfilm with in-house digitization of manuscript materials and demonstrated that MPLP could indeed be effectively applied to digitization in support of large-scale initiatives.[10] Other articles have explored the process of mass digitization of photographs specifically.[11] There are multiple articles and books related to the National Digital Newspaper Program, but the majority of these articles focus on specifics of the program itself, including challenges (budgeting, title selection,[12] project management, image quality and newspaper titles,[13] etc.), technologies used (microfilm and digitization equipment and

---

[4] Oya Y. Rieger, "Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 11, no. 1 (2010): 11–22, https://doi.org/10.5860/rbm.11.1.328.

[5] Cory Lampert, "Ramping Up: Evaluating Large-Scale Digitization Potential with Small-Scale Resources," *Digital Library Perspectives* 34, no. 1 (February 12, 2018): 45–59, https://doi.org/10.1108/DLP-06-2017-0020.

[6] "Principles to Guide Vendor/Publisher Relations in Large-Scale Digitization Projects of Special Collections Materials," Association of Research Libraries, June 2010, http://www.arl.org/storage/documents/publications/principles_large_scale_digitization.pdf.

[7] Ricky Erway, "Rapid Capture: Faster Throughput in Digitization of Special Collections," OCLC Research, 2011, https://www.oclc.org/content/dam/research/publications/library/2011/2011-04.pdf.

[8] "Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections," Final Grant Report Prepared by the Southern Historical Collection, University Library, The University of North Carolina at Chapel Hill, for the Andrew W. Mellon Foundation, June 2009, https://docsouth.unc.edu/watson/archivalmassdigitization/download/extending_the_reach.pdf.

[9] "Digitizing Entire Collections," Archives of American Art, accessed August 7, 2018, https://www.aaa.si.edu/documentation/digitizing-entire-collections.

[10] Shan C. Sutton, "Balancing Boutique-Level Quality and Large-Scale Production: The Impact of 'More Product, Less Process' on Digitization in Archives and Special Collections," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13, no. 1 (2012): 50–63, https://doi.org/10.5860/rbm.13.1.369.

[11] Mark A. Vargas and Jessica Bright, "Rev Your Engines: Racing Ahead with Mass Digitization," *Computers in Libraries* 37, no. 7 (September 2017): 4–8; Craig Harkema and Cheryl Avery, "Milne En Masse: A Case Study in Digitizing Large Image Collections," *New Review of Academic Librarianship* 21, no. 2 (May 2015): 249–55, doi:10.1080/13614533.2015.1034806.

[12] Donnie Summerlin, "Selecting Newspaper Titles for Digitization at the Digital Library of Georgia," *D-Lib Magazine* 20, nos. 9–10 (September–October 2014), doi:10.1045/september2014-summerlin.

[13] Kopana Terry, "The Digitization of Historic Newspapers on Microfilm: The Kentucky Experience," *Microform & Imaging Review* 38, no. 2 (July 7, 2009): 54–63, https://doi.org/10.1515/mfir.2009.006.

OCR [optical character recognition] software),[14] metadata schema, and quality review technologies.[15]

This case study described below is unique because it compares the large-scale digitization of three common types of archival materials: microfilmed newspapers, photographs, and manuscript materials. By comparing and contrasting different strategies for large-scale digitization, it highlights commonalities that can contribute to the establishment of best practices while also taking into account the diversity of materials cared for by special collections and archives.

Nevada Digital Newspaper Project (NvDNP)

The Nevada Digital Newspaper Project is an extension of the National Digital Newspaper Program—a partnership between the National Endowment for the Humanities and the Library of Congress. It is a grant-funded large-scale digitization project that runs on two-year grant cycles and UNLV has been funded through two cycles (2014–2016, 2016–2018). The state's largest newspaper digitization project consists of selecting, digitizing, and making publicly accessible 100,000 pages per grant cycle of historic newspapers from Nevada. Newspapers that were already available on microfilm were selected and digital objects were created from second-generation (2N) duplicate silver negative microfilm. The master negatives are preserved at the Library of Congress and the digitized newspapers are readily available via the country's largest newspapers repository, Chronicling America.[16]

Culinary Workers Union Project (CWU)

The Culinary Workers Union Local 226, the largest local union of UNITE HERE (the Union of Needletrades, Industrial, and Textile Employees and Hotel Employees and Restaurant Employees), is located in Las Vegas, Nevada and represents over 55,000 hotel, food service, and hospitality workers in southern Nevada.[17] The Culinary Workers Union Local 226 donated their photographic archives (1950s–2006) to UNLV Special Collections and Archives in 2014 and 2016. The photographs document the Frontier Strike, one of the longest lasting strikes in United States history, along with other strikes, demonstrations, public events, work activities, and social events of the union. The collection was digitized due to the union's need for easier access to its own photographs and its desire to make them available to the public. UNLV Special Collections and Archives wanted to improve access to this important collection (which is minimally described at the file level), and the acquisition coincided with the purchase of rapid capture digitization equipment that the library was eager to put to use. With input from the union's director of communications and digital strategy, UNLV's Digital Collections department

---

[14] Kenning Arlitsch and John Herbert, "Microfilm, Paper, and OCR: Issues in Newspaper Digitization; The Utah Digital Newspapers Program," *Microform & Imaging Review* 33, no. 2 (January 9, 2008): 59–67, https://doi.org/10.1515/MFIR.2004.59.

[15] Thomas McMurdo and Birdie MacLennan, "The Vermont Digital Newspaper Project and the National Digital Newspaper Program," *Library Resources & Technical Services* 57, no. 3 (2013): 148–63, http://dx.doi.org/10.5860/lrts.57n3.148.

[16] "Chronicling America: Historic American Newspapers," Library of Congress, accessed February 4, 2019, https://chroniclingamerica.loc.gov/.

[17] Culinary Workers Union Local 226 Las Vegas, Nevada Photographs, 1950s–2006, PH-00382, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, accessed February 4, 2019, http://n2t.net/ark:/62930/f16g7f.

digitized and provided online access to the collection (thirty-seven linear feet of photographic prints and transparencies) in a year and a half, completing the project in December 2017.

Entertainment Grant Project (ENT)

The main goal of this grant project was to help Nevada cultural heritage institutions gain skills and build statewide capacity for large-scale digitization by demonstrating models and teaching techniques to empower digital project teams across the state. The second goal of the project was to enable UNLV Special Collections and Archives to digitize and provide access to some of the most frequently and widely used archival collections related to entertainment: the Jerry Jackson Papers,[18] the Donn Arden Papers,[19] and the Sands Hotel Public Relations Records.[20] Using previous experience with large-scale digitization, UNLV's Digital Collections team translated theory into practice and evolved workflows to meet the needs of a large-scale digitization project. The findings were then shared in a free, statewide workshop at the conclusion of the project. By creating one large entertainment-related digital collection, the team has made it possible for patrons to access a comprehensive research archive on entertainment in southern Nevada. This venture was funded by a one-year (2017–2018) Library Services and Technology Act (LSTA) grant administered by the Nevada State Library, Archives and Public Records. The grant proposal was titled "Raising the Curtain: Large-Scale Digitization Models for Nevada Cultural Heritage."

Table 1. UNLV project comparison

	ENT	CWU	NvDNP
Length of project	1 year	1.5 years	2 years
Volume of materials digitized (linear feet)	134	37	11,378 *128 microfilm reels*
Digital items created	47,338	38,750	100,391
Digital compound objects online	1,406	1,015	25,097
Digital single objects online	234	55	0
Project websites	http://d.library.unlv.edu/digital/collection/ent	http://d.library.unlv.edu/digital/collection/cwu	https://chroniclingamerica.loc.gov/newspapers/Nevada/

---

[18] Jerry Jackson Papers, 1953–2009, MS-00573. Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada. Collection guide: http://n2t.net/ark:/62930/f1rp63.

[19] Donn Arden Papers, 1918–1994, MS-00425, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada. Collection guide: http://n2t.net/ark:/62930/f1k61t.

[20] Sands Hotel Public Relations Records, 1952–1977, MS-00417, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada. Collection guide: http://n2t.net/ark:/62930/f1mw5j.

Roles and Responsibilities

Each of the three large-scale digitization projects at UNLV had a unique set of project staff due to the variety of funding resources available for each project. Teams were each configured differently but had similar staff roles with analogous responsibilities.

Table 2. UNLV project teams

ENT	CWU	NvDNP
● Principle investigators ● Project manager ● Project technician ● Student assistants	● Project manager ● Student assistants	● Principle investigators ● Project manager ● Project technician ● Digitization vendor

Both ENT and NvDNP were grant-funded projects proposed and authored by principal investigators (PIs). The PIs identified the main goals for the projects and then wrote grant proposals in order to secure funding. They were responsible for budgeting once the funding was secured and for any reporting required by the funding institution. Additionally, in NvDNP, the principal investigators and project manager were responsible for communicating with outside vendors.

Each project had a project manager who helped develop and maintain efficient workflows and procedures; managed digitization activities, staff schedules, and quality control; and enforced project deadlines. The project manager oversaw the daily activities of the team and assisted in digitization and metadata creation to ensure the project remained on track to meet its goals. In CWU and ENT, project managers established workflows for digitization, metadata creation, OCR, and final upload of digital objects into the digital collections management system.[21] For ENT and NvDNP, the project managers were hired as temporary contract positions, which allowed the PIs to take a step back from the daily management of the project. For CWU, the project manager was the digital special collections librarian who had additional job duties outside of the project. The in-house digitization of a single photograph collection (CWU) was more straightforward than ENT and NvDNP, which enabled student assistants to complete the bulk of the project.

Student assistants were responsible for a large part of ENT and CWU, including digitization, metadata creation, and uploading items to the digital collections management system. Undergraduate student workers perform an important part of the work in UNLV Special Collections and Archives, but their career goals do not always align with their job duties, and coursework is naturally prioritized over work in the library. This results in a very dynamic working schedule, which can mean limited staff during exam weeks and holiday breaks, and an abundance of student hours during summer when schedules are more flexible. Adjusting project

---

[21] UNLV Digital Collections uses CONTENTdm, an OCLC supported software.

workflows and managing deadlines is made more challenging for the project manager due to students' changing schedules.

Working with student staff to determine their strengths and weaknesses was critical and ensured a solid foundation moving forward with additional projects.[22] By assigning a variety of tasks at the beginning of a project, the project manager was able to determine what areas of the project students excelled in: digitization, metadata creation, or quality control, for example. If students continued their employment in Digital Collections for a long period of time, they developed their own unique areas of expertise or interest and became a leader within the team. Some students became personally interested and invested in their work, taking on more complex assignments, while others were most productive with straightforward and very clearly explained tasks that did not require critical thinking.

An alternative to part-time student labor is a dedicated project technician. The project technician for NvDNP was responsible for executing the daily production activities and staying on task in order to meet the deadlines. Production activities included collation of microfilm, metadata creation, and quality control. In NvDNP, the project technician was responsible for the majority of the project's daily tasks. Historically in NDNP projects, full-time project technicians were chosen instead of student workers due to their consistent schedule and reliability, and thus UNLV's NvDNP principal investigators adopted the same structure.

Most of the staff involved in UNLV's large-scale digitization projects were hired on temporary contracts. If the funding for additional staff is available, this can be a great way to scale up digitization activities. However, training temporary or contract staff is a significant investment of time and resources. Permanent staff should always remain involved and engaged throughout the project in order to institutionalize the best practices and workflows as they are developed. There is also always a possibility that temporary staff may leave before their contract is over. For example, the ENT project manager transitioned to a different project within the library five months into their nine-month contract. The ENT digitization technician replaced them, but the technician had less experience and archival training. As a result, one of the PIs took on additional management responsibility in the project and an additional student assistant was hired to keep up with the project goals and grant budget.

It is not uncommon in a large-scale digitization project to use outside vendors to complete project components that the institution cannot, often because of a need for specialized equipment or expertise, or due to the large volume of material. NvDNP outsourced two of the project's major components, which freed up more internal staff time than if the tasks had been completed in-house, but this presented additional communication and coordination challenges and responsibilities for the project manager. The project manager and PIs had to communicate to the vendor all of the digitization specifications outlined in the National Digital Newspaper Program 2016 Technical Guidelines for Applicants.[23] The project manager developed digitization requirements, internal workflows, and deadlines at the beginning of the grant cycle for the

---

22 Emily Gainer and Michelle Mascaro, "Faster Digital Output: Using Student Workers to Create Metadata for a Grant Funded Project," *Provenance, Journal of the Society of Georgia Archivists* 30, no. 1 (2012), https://digitalcommons.kennesaw.edu/provenance/vol30/iss1/3.

23 "The National Digital Newspaper Project (NDNP) Technical Guidelines for Applicants," Library of Congress, September 29, 2015, https://www.loc.gov/ndnp/guidelines/archive/NDNP_201618TechNotes.pdf.

vendor to follow but needed to remain flexible as the vendor completed work in their queue. Due to the large volume of material being digitized, NvDNP project staff enforced strict quality control measures and sent back reels to be re-digitized if necessary to ensure that the vendor adhered to the grant's technical requirements. Utilizing outside vendors requires open and clear communication regarding delays, issues, or questions. Any setback needs to be addressed immediately and communicated to all project stakeholders, especially when there are granting agency deadlines to be met.

Communication and organization was essential to the success of all the UNLV large-scale digitization project teams. ENT and CWU project managers and student staff supplemented spreadsheets with Trello, a Web-based project management application. NvDNP staff communicated their activities using Basecamp, a project management and team communication software. Managers assigned and tracked tasks, established deadlines, and monitored project activities using these project management applications. Utilizing these applications kept all team members, including outside vendors, informed, updated, and accountable during UNLV's large-scale digitization projects.

Grant Funding

For many institutions, grants are sources of fresh money that enable special collections to reduce their backlog or try something new and innovative. Several grant programs target digitization and archival processing, and the number of institutions taking advantage of extra funding is on the rise.[24]

Just as with everything else, however, being awarded a grant has its benefits and challenges. On the one hand, it is a wonderful opportunity for digitizing and making publicly available under-utilized archival materials; sometimes digitization is critical, as it is the only way to preserve fragile items. On the other hand, all grant projects come with restrictions, limitations, and rigid deadlines that can lead to tension and frustration due to lack of flexibility and independent decision-making. Those who consider applying for grant programs need to be aware of all caveats that come with such external funding and must be prepared to work within the guidelines, policies, and deadlines established by the funding agency.

Applying for external funding requires familiarity with an institution's collections and a strategy to match the collections with appropriate funding opportunities. Grant funding can serve as a catalyst for digitization priorities, which can result in a scattershot portfolio of digital collections rather than a programmatic approach addressing the needs of collections that have a high research value or a high preservation risk. Applying for grants also takes much planning, from choosing the collection, to creating an inventory of the materials, calculating budgets, and selecting project staff. Ideally, the principal investigators should consider all potential expenses in the application, but this is not always the case; many unforeseen circumstances arise.

Two of the three projects in this case study were grant funded. The NvNDP was funded as part of the National Digital Newspaper Program (NDNP), a nationwide initiative and collaborative effort for digitizing historic newspapers, while the ENT project was funded as part of the Library

---

[24] "NEH Announces $18.6 Million for 199 Humanities Projects Nationwide," National Endowment for the Humanities, April 9, 2018, https://www.neh.gov/news/press-release/2018-04-09.

Services and Technologies Act (LSTA). The ENT project grant was slightly more flexible than the NvDNP grant. It came with an advantage in that the project team had freedom in decision-making in important areas such as workflow organization, metadata schema and metadata element set, and online repository. It did have stipulations and deadlines, such as quarterly reports required by the Nevada State Library, Archives, and Public Records (NSLAPR) that documented project progress, challenges, and a detailed budget. The project deadline was originally set for June 30, 2018, and the project manager was able to assign multiple interim deadlines and set milestones for the project. However, the final project deadline was extended to July 31, 2018, after the unexpected early departure of the project manager created a budget surplus.

The NvDNP grant was more structured. The Library of Congress stipulated the final project deadline and interim deadlines, due dates for midterm and final reports, and the quantity, quality, and format requirements of the deliverables. There were also established guidelines for workload segmentation, shipping of deliverables, metadata element set, encoding schema, selection of newspaper titles for digitization, essay writing, and essay submission due dates. The NvDNP team had the flexibility to draft the monthly/weekly scheduling. These schedules complemented and supported the formal Library of Congress deadlines and were tailored to fit the project needs and to keep the project on track. The structure provided a helpful roadmap for project managers and was complemented by outstanding partnership and collaboration with the Library of Congress and National Endowment for the Humanities; their guidance and support was essential for the success of the project.

Another contrast between the ENT project and NvDNP was the advisory board. The NDNP program funds a Statewide Advisory Board to select titles for digitization. The board drafted their decisions and submitted them for approval in the form of a title list. The Advisory Board was assisted when selecting titles for digitization by title selection criteria outlined in the Content Selection Guidelines;[25] the Library of Congress had the final word and could send the proposed title list back for revisions if the newspapers did not meet the selection criteria. The statewide advisory board was also responsible for assigning digitization priorities; they ranked the titles from the approved list and established the digitization order, considering multiple factors such as microfilm quality, team availability, vendor operation schedule, and researchers' demand.

The CWU project was a different scenario. It started as a library-funded project that developed into a partnership with the Culinary Workers Union Local 226, which contributed to its timely completion. Being internally funded allowed more freedom and flexibility in decision-making and project management. The union's only requirement was that the project team employ students on the project and work with the union's director of communications and digital strategy, who reviewed all images before they were published online.

Copyright, Privacy, and Confidentiality

The three projects described in this case study represent three different copyright scenarios. Evaluating the copyright status of materials for digitization and online public access projects is an important step in the planning process. Some cases may be relatively clear-cut, while others

---

[25] "Content Selection," Library of Congress, last updated August 4, 2016, http://www.loc.gov/ndnp/guidelines/selection.html.

may require a more careful analysis of risks and benefits. Some institutions are more risk averse than others, and other factors, such as privacy and confidentiality concerns, should be evaluated too. This article is not legal advice, and there are more comprehensive guides addressing fair use and copyright available,[26] but below are some specific examples of copyright evaluation for large-scale digitization projects.

United States copyright law does not restrict the use of materials in the public domain. For that reason, many institutions consider digitizing materials in the public domain to be relatively risk-free in relation to copyright. NvDNP is an example of a project that focused only on materials in the public domain. Newspapers are also published publicly at the time of creation, so there are few concerns about exposing private or confidential information by making them publicly available online.

When the copyright holder for an entire collection is known and can be easily contacted, the institution can get a clear answer as to whether or not the copyright holder grants permission, or the institution can at least document due diligence in trying to obtain permission. The Culinary Workers Union Local 226 photographs are an example of this scenario. The photographs were created by the union itself and the union claims copyright. The union not only gave permission for digitization, but also were partners in the project to provide public online access to the collection. The Jerry Jackson Papers that were digitized as part of the ENT project were another case in which UNLV Special Collections and Archives had clear and documented permission in the form of a gift agreement. The agreement was signed by the creator himself, Jerry Jackson, and included the terms that "the Collection may be reproduced and used in any way without restriction to further and support the mission of UNLV." Writing terms into gift agreements that allow for digitization and public online access is a good strategy to preemptively obtain permission for these kinds of projects.

The Sands Hotel Public Relations Records, digitized as part of ENT, presented a slightly different scenario: the collection contains materials with many different creators, often all mixed together. In this case, the collection was digitized and put online according to the principle of fair use under US copyright law. Fair use can be assessed by considering the following questions, summarized in the "Statement of Best Practices in Fair Use of Collections Containing Orphan Works for Libraries, Archives, and Other Memory Institutions":

> 1. Did the use "transform" the material taken from the copyrighted work by using it for a broadly beneficial purpose different from that of the original, or did it just repeat the work for the same intent and value as the original, in effect substituting for it?
> 2. Was the material taken appropriate in kind and amount, considering the nature of the copyrighted work and of the use?[27]

Digitization of entire archival collections presents a reasonable case for fair use because the use is transformative; digitizing individual items and keeping them within the context of an archival collection transforms them from their original intended use into a new research use. Scholars and the public can use the Sands Hotel Public Relations Records to understand the history of the

---

[26] For example, Patricia Aufderheide et al., "Statement of Best Practices in Fair Use of Collections Containing Orphan Works for Libraries, Archives, and Other Memory Institutions," Center for Media & Social Impact, 2014, http://www.cmsimpact.org/sites/default/files/documents/orphanworks-dec14.pdf.
[27] Ibid., 18.

hotel, the city, gaming, and more. They can also use the collection in new computational ways, including full text search, text mining, visualization, and so on. For researchers, context is very important, so it is appropriate that the entire collection (or at least entire files or series) is accessible. Since the use is different from the original intended use, it is also unlikely to have a negative economic impact on the copyright holders.

The Sands Hotel Public Relations Records also contain personally identifiable information like social security numbers that should not be shared with the public. Laws such as HIPAA (the Health Insurance Portability and Accountability Act of 1996) and FERPA (the Family Educational Rights and Privacy Act) protect health and student information, but allowing access to other information in archival collections may be up to the discretion of the archives. It is good practice for an institution to have a policy on sensitive information, whether it is in physical, born-digital, or digitized materials. In the Sands Hotel Public Relations Records, only a few documents contained social security numbers, so it was decided to physically cover the numbers when capturing the digital image. Electronic redaction is also an option, but it must be done before processes such as OCR and must be secure enough so that the redaction cannot be stripped away by software.

Dealing with copyright and sensitive information relies on the ability to balance risks and benefits, and to consider not only the institution, the public, and the donor, but also any copyright holders or third parties. When undertaking large-scale digitization it is best to choose materials that do not pose significant risks in these areas. Even collections that may be completely open in the reading room may not be appropriate for digitization and public online access. It is wise to create a takedown policy to establish guidelines for anyone who may request removal of items.[28]

When undertaking large-scale digitization, it is necessary to survey the collection during the planning phase to determine the copyright status of the materials and if there is any sensitive information in the collection. It is important to communicate with internal and external stakeholders to decide how copyright and privacy issues will be addressed and to document all decisions made and workflows created. Issues with individual items may not be discovered until mid-workflow, so it is important for staff to be able to identify these issues and know what to do without causing unnecessary interruptions or delays. Along with a takedown policy, information on how the materials can be used should be made easily available. Copyright information should be included in the metadata, ideally in standardized form using RightsStatements.org or Creative Commons. Given the significant amount of resources required to undertake large-scale digitization, it is desirable to encourage use and reuse as much as possible and to make it easy for users to find information about allowable uses.

Arrangement

When undertaking a large-scale digitization project, it is important to decide how you will reflect the hierarchy of an archival collection and the "feel" of browsing through physical materials. While traditional digital collections are described at the item level, it is much more difficult to achieve the same amount of detail in a large-scale project if the collection has not already undergone item-level processing. Projects like ENT, CWU, and NvDNP reused existing

---

[28] UNLV's take-down policy is available on this webpage: "Using the Digital Collections—Copyright, Use Guidelines," UNLV Digital Collections, accessed May 29, 2018, http://digital.library.unlv.edu/request-images.

descriptions and groupings to maintain efficiency. Instead of creating many single objects for a large digital collection, complex digital objects were created to mirror the existing arrangement of the collection's folders or intellectual groupings.

UNLV Libraries uses CONTENTdm as their digital collections management system. In this system, digital files can be uploaded and viewed individually as single objects, or in groups, referred to as compound or complex digital objects. A single object is a single digital file that often represents an individual piece of intellectual content. A good example of a single-item object comes from ENT, where a single costume design sketch has a descriptive record and a single digital image.[29] A complex digital object consists of more than one digital file and can represent any number of physical items that are grouped together.[30] A compound object's descriptive metadata consists of one aggregate "parent" record that represents all "child" digital files or items. The compound object may also contain descriptions for each child item, but that is often not the case in large-scale digitization.

The Digital Public Library of America's (DPLA) Archival Description Working Group recommends that DPLA's partners "consider creating aggregated objects when existing description is at an aggregated level" for both practical and philosophical reasons.[31] This trend toward utilizing aggregate description is increasingly the standard for archives and special collections since Mark Greene and Dennis Meissner's "More Product, Less Process" article in 2005.[32] The same aggregate descriptive practices can also be applied and utilized in the digital environment. The context provided by aggregations is useful for some researchers, who can then mimic the reading room experience of viewing an entire archival folder or other intellectually meaningful grouping rather than looking at individual items separated out of context. For other Web users, complex objects can prove challenging if they are not expecting to retrieve results that require browsing several subsets (child records).

It is important to survey the collection during the planning phase of a large-scale digitization project to determine how the physical materials are arranged and how the digital objects will reflect that arrangement. Any rearrangement or processing activities should be done (if possible) before digitization. This ensures that the arrangement of the digital surrogate mirrors or builds upon the intellectual arrangement of the physical materials, and the context and hierarchy of the collection material is preserved.

For projects like NvDNP that consist of uniform units with titles, the arrangement of the digital objects is determined by the existing arrangement and description; in this case, that of newspaper publication, individual issue, and page. All digital objects in this project are compound objects equivalent to one newspaper issue and they follow the same page sequence. The challenge for the NvDNP team was that they were not working with the original materials but with microfilm copies. Usually the newspaper issues were filmed in their publication page sequence and in those

---

[29] #4 for Kristine Kappeh: Original Rough Sketch, Jerry Jackson Papers, 1953–2009, MS-00573, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, http://n2t.net/ark:/62930/d13d19.
[30] Costume Design Drawings by Nolan Miller, Jerry Jackson Papers, 1953–2009, MS-00573, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, http://n2t.net/ark:/62930/d1v63t.
[31] Digital Public Library of American (DPLA) Archival Description Working Group, "Aggregating and Representing Collections in the Digital Public Library of America," November 2016, http://bit.ly/dplaCollections.
[32] Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 208–63, https://doi.org/10.17723/aarc.68.2.c741823776k65863.

cases digitizing and grouping digital objects was straightforward. However, the team sometimes experienced complicated scenarios that had to be resolved prior to digitization so that the digital objects reflected the correct arrangement of original newspaper publications.

If newspapers were filmed out of sequence, additional research of the original page sequence was required. The majority of newspapers from the 1800s and early 1900s do not have page numbering, so this was a particularly burdensome task that slowed down the project. Occasionally, newspapers with missing pages required the team to determine if these missing pages were randomly filmed toward the end of the reel, on a different microfilm reel, or not filmed at all. Finally, newspapers are frequently microfilmed with duplicate pages. Often the duplicates are appended to the newspaper issue to which they belong. The decision in this case was easy: the vendor digitized the better copy. In rare cases, duplicate pages appear out of sequence on the same or a different reel. In situations like this, the team would track and determine which copy had better quality and notify the vendor so that only one copy of the page was digitized.

For both ENT and CWU, the archival arrangement of the physical materials determined the digital object aggregations. The majority of the materials in the ENT and CWU collections used complex digital objects to reflect the file-level archival arrangement and description. It was most efficient for each individual folder (which consisted of a variety of material formats) to become one complex digital object made up of all the individual items within that folder. In the ENT collections, an intellectual file-level description in the finding aid referred to a single physical folder. In the CWU collection, a file-level description often referred to multiple physical folders, each containing hundreds of items. Instead of creating huge digital objects that corresponded to the file-level description in the finding aid, digital objects were created for each individual physical folder. File titles from the finding aid were repurposed to create digital object titles, mirroring the physical arrangement of the materials. For example, "Photographs of First rally of the Frontier Strike, Culinary Union, Las Vegas (Nev.), 1991 August 07 (folder 2 of 3)" (see the metadata section of this article for more information about description).

In ENT, the three different collections were all processed with varying degrees of specificity and hierarchy. The Donn Arden Papers and the Sands Hotel Public Relations Records are both arranged into a number of series and subseries with folder-level inventories. This made it relatively simple to have the digital arrangement reflect the physical. The third collection, the Jerry Jackson Papers, was processed to a very fine degree—but only in certain series.[33] These item-level descriptions became single objects once they were digitized. Single item objects were only created when there was an item-level description already available in the finding aid.

---

[33] This reflects the adoption of MPLP at UNLV Special Collections and Archives: a "golden minimum" level of processing for all collections, and finer-level processing for specific collections, series, or items as their value merits.

Figure 1. Screenshot of part of the Jerry Jackson Papers collection guide with "view online" links to digital objects at both the folder and item level.

16th Edition		
Presentation, outline, running order, budgets, 1983 (view online)	box 01	folder 24
"Old Fashioned Opening": hand written sheet music, 1983 (view online)	box 01	folder 25
Music notes, lyrics, speeches, 1983 (view online)	box 01	folder 26
Choreography notes, 1983 (view online)	box 01	folder 27
Fly plot and cue sheets, 1983 (view online)	box 01	folder 28
Set design drawings by Charles Lisanby		
Souvenir Card and Wings: Drawing Number 1, 1983 (view online)	flat file 21	
Paris Electric Drop: Drawing Number 2, 1983 (view online)	flat file 21	
Starburts and Swings: Drawing Number 3, 1983 (view online)	flat file 21	
Diamond and Star Electric Drop: Drawing Number 4, 1983 (view online)	flat file 21	
Opening Finale: Drawing Number 5, 1983 (view online)	flat file 21	
Intro to Clock: Drawing Number 6, 1983 (view online)		
- Page 17-		

Formats

UNLV's large-scale projects have evolved over time from digitizing homogenous materials to encompassing more complex collections and workflows.

Since 2014, NvDNP has digitized 100,000 historic newspaper pages per grant funding cycle (two years). Newspapers that were already available on microfilm were selected and digital objects were created from second-generation (2N) duplicate silver negative microfilm. Although duplication and digitization were outsourced, in-house staff used microfilm readers for collation. The uniformity of the source material (both the newspaper and microfilm formats) meant that all materials were digitized, described, and presented in the same way. The digitization vendor applied OCR software to produce PDF files in addition to TIFF files, and encoded metadata in METS/ALTO schema.[34] Uniformity is ideal for large-scale digitization and even more so when outsourcing work.

---

[34] METS (Metadata Encoding and Transmission Standard) and ALTO (Analyzed Layout and Text Object) are XML standards maintained by the Library of Congress. The combination of METS and ALTO (often written METS/ALTO) is the current industry standard for newspaper digitization used by hundreds of modern, large-scale

Many institutions have in-house experience digitizing flat reflective materials, such as photographic prints or documents, using flatbed or overhead scanners. Equipment is one factor that can make a big difference in scaling up digitization projects. Having the Special Collections and Archives curator of visual resources, who is also a professional photographer, involved in the early planning stages of large-scale digitization gave Digital Collections unique insight into and expertise on what equipment was best suited for large-scale projects:

> To undertake truly large-scale digitization, time is everything. Using a digital camera mounted above materials on a copy stand can speed up the process of digitizing a variety of materials. What takes a flatbed scanner 1-2 minutes to capture, takes a camera or digital back a split second. These kinds of setups can vary greatly depending upon your needs and budget. Using a consumer level mirrorless or DSLR camera (like a Canon, Nikon, Sony, or Fuji) mounted on an inexpensive copy stand with lights (like those made by Beseler or Kaiser) can get your project rolling faster. More sophisticated high quality rapid capture systems with camera, high megapixel digital back, copy stand, lights, and integrated software designed specifically for cultural heritage work (like those offered by Phase One through Digital Transitions) can be purchased and will undoubtedly supercharge your large-scale digitization program while elevating your project's FADGI[35] 4-Star compliance. Regardless of the size of your organization, using stand-mounted cameras instead of flatbed scanners will increase your productivity.[36]

The CWU and ENT digitization projects at UNLV Libraries used a Digital Transitions Phase One cultural heritage rapid capture system. Most of the photographs in the CWU collection are prints, but there are also many negatives and other transparent photographic materials mixed in. Working with collections of mixed materials is challenging because different physical formats require different equipment and workflows to digitize and provide access to them. Transparent materials can also be digitized using an overhead camera but additional equipment is needed, such as a light box and film carriers. Whether a flatbed scanner or overhead camera is used, photographic negatives will always take more time to digitize than prints. The FADGI Still Image Working Group explains:

> Original photographic negatives are much more difficult to scan compared to positive originals (prints, transparencies, slides, etc.), with positives, there is an obvious reference image that can be matched and for negatives, there is not. . . . Digitizing negatives is very analogous to printing negatives in a darkroom and it is very dependent on the photographer's/technician's skill and visual literacy to produce a good image. There are

---

newspaper digitization projects. To learn more about METS/ALTO visit https://www.veridiansoftware.com/knowledge-base/metsalto/.

[35] Federal Agencies Digital Guidelines Initiative. See Thomas Rieger, ed., "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Files," Federal Agencies Digital Guidelines Initiative, September 2016, http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf.

[36] Aaron Mayes, UNLV Special Collections and Archives Curator of Visual Materials, email to author, June 11, 2018.

few objective metrics for evaluating the overall representation of digital images produced from negatives.[37]

The CWU project team digitized photographic prints first, keeping track of any transparent materials to be digitized later. Physically and intellectually, prints and transparencies were mixed together within the same folders and within digital objects, so it was important to keep track of arrangement and file-naming. Periodically the capture system would be reset in order to digitize the transparencies in batches. Many of the negatives were already housed in transparent sleeves, and these sleeves were placed on a light box to create digital proof sheets. Negatives were digitized as proof sheets instead of individually to save time, but editing the digital proof sheets proved to be challenging. Strips of negatives from different shoots were grouped together within a single proof sheet so often strips still needed to be edited individually because the lighting or coloring differed. Once the master digital images (TIFFs) of the prints and transparencies were created, the workflows for description and access were more straightforward because all of the items were photographs and there was no text to transcribe.

The three manuscript collections digitized for the ENT project contained multiple formats mixed together, for example, photographic prints, negatives, and text documents all within the same folder. Rather than physically sorting materials into like groups before digitization, original order was maintained and all reflective flat materials smaller than 12" x 16" were digitized first. Although in other situations text is usually captured at 300ppi, all reflective flat materials smaller than 12" x 16" were captured at 600ppi so that the student assistants digitizing did not have to switch back and forth between settings. Any materials that could not be captured during the first pass (such as transparencies and oversized materials) were identified and later digitized in batches.

For the ENT project, negatives and other photographic transparencies were digitized individually using a film kit and light box with the Phase One rapid capture system. Oversized materials smaller than 30" x 40" were also digitized on the Phase One system at 300ppi. Larger oversized materials, such as set design drawings, were plentiful in the ENT collections. These extra-large materials were digitized using a ColorTrac SmartLF Gx+ T56 large format scanner. However, the team ran into technical difficulties due to digitizing such a large volume of materials in a short span of time on equipment not meant for high-speed, high-volume production. Large-scale digitization depends on efficiencies in many different aspects of the workflows, and deciding between using in-house equipment or outsourcing to a company that has the appropriate equipment is a fundamental consideration.

The ENT project team conducted a cost-benefit analysis and decided to skip certain infrequent formats, at least in the large-scale grant phase of digitization. The team discussed having a photo shoot for costume pieces and other three-dimensional artifacts such as trophies, plaques, and signs, but decided it was not worth the resources at that time for the small number of costume pieces and the low research value of the other items. Audiovisual materials were also not digitized; UNLV Special Collections and Archives outsources audiovisual materials for digitization at scale, but this was not accounted for in the original grant budget. Entire folders of

---

[37] Federal Agencies Digitization Initiative (FADGI)-Still Image Working Group, "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files," August 2010, 52, http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf.

newspaper clippings were also skipped, and newspaper clippings grouped with other materials in a folder/digital object were photographed in a pile with the following notice:

> This item has not been digitized in its entirety. The original item is available for research and handling at the UNLV University Libraries. Additional digitization is available upon request. Please contact Special Collections to request additional digitization or with any questions regarding access at special.collections@unlv.edu.[38]

Magazine covers were also digitized with the above notice.[39] This approach allowed UNLV Digital Collections to focus on providing access to the most valuable unique archival materials. Newspaper and magazine content is published elsewhere, and the clippings in this collection were often folded up (requiring flattening and possibly rehousing) and oversized (requiring additional tracking and a different digitization equipment setup). Digitizing the clippings individually and entire issues of magazines would have required significantly more time and presented copyright challenges. The notice allows users to see that more physical content exists, and to either view it in person or request additional digitization.

All ENT digitization resulted in the same type of digital files: digital images (TIFF masters and JPEG access copies). There are a number of choices for digital file formats, but when digitizing at a large scale it is important to consider file size in addition to quality. Planning for the storage and preservation of digital files, not just their creation and dissemination, is necessary.

Even though not all of the materials were textual, text and images were often mixed together within the same compound digital object. No effort was made to distinguish images from text during the OCR process because it was faster and less error-prone for student assistants to process everything the same way. All JPEGs belonging to a compound digital object were loaded into Abbyy OCR software (some digital objects had hundreds of child items, and JPEGs load faster than TIFFs) and the OCR transcripts were not corrected. A text file of the transcript was outputted for each individual image file, and a PDF/A file was created for the entire compound digital object. The text files were used to add the transcripts into the metadata for each image file. The JPEG images are available online along with metadata that includes the uncorrected transcripts. PDF/As of the digital objects are also available, but because of UNLV Digital Collections' current online access system, this is not readily apparent to users. Ideally, multiple available formats would be available online side by side so users could choose the format that is most useful to them.

Metadata

Metadata is an important component of every digital object. Whether it is basic or rich, each object needs description. Embracing large-scale digitization brings metadata challenges as decisions are made to prioritize either quality or quantity in output. The new large-scale digitization approach changed UNLV Libraries' perspective on metadata. Projects yielding

---

[38] For an example, see "Nude Dancers Controversy, Sands Hotel Public Relations Records," 1952–1977, MS-00417, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, http://n2t.net/ark:/62930/d1kw84.

[39] For an example, see "Song lyrics to 'Red Hot Mama'—show unknown," Donn Arden Papers, 1918–1994. MS-00425, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, http://n2t.net/ark:/62930/d18g6f.

thousands of new digital objects per month require a completely different approach for describing the items; in this case, the focus was on automation. UNLV has taken this approach to aim for efficiency and to capture metadata at the most granular level that exists. This results in more basic descriptions, the reuse of existing metadata from finding aids, and aggregate object description. All three projects described in this case study focused on efficiency and reused metadata (where possible) for describing digital objects. The metadata provides succinct but sufficient information, which is enough for users to discover and retrieve the objects they need.

The ENT and CWU collections were uploaded to CONTENTdm with descriptive metadata created according to UNLV's metadata application profile and mapped to the Dublin Core Schema. UNLV's digital collections are harvested by the Mountain West Digital Library (MWDL) and the Digital Public Library of America (DPLA), so certain metadata fields are required in order to align with both MWDL and DPLA's metadata application profiles.

The NvDNP metadata element set is governed by the National Digital Newspaper Program 2016 Technical Guidelines for Applicants,[40] so project staff dedicated the most time to adding value to page-level metadata. While considering the program's metadata guidelines, the project manager was able to design a metadata set that meets the Library of Congress requirements and yet was tailored to reflect the richness of the physical objects and their content. Unlike most digital collections where metadata applies to objects only, the NvDNP element set is a three-tier metadata structure describing the batch itself, the physical reel, and the items.

It is interesting to compare how the different UNLV large-scale digitization projects created titles for digital objects. In the CWU project, file titles from the finding aid were enhanced to describe digital objects that are the equivalents of physical folders: for example, the finding aid file title "First rally of the Frontier Strike" and corresponding date were repurposed for the digital object title "Photographs of First rally of the Frontier Strike, Culinary Union, Las Vegas (Nev.), 1991 August 07 (folder 2 of 3)."[41] Similarly, the NvDNP team used the original publication title with a slight modification so it conforms to the Library of Congress official title record. Usually, the changes were related to punctuation and/or word capitalization. For example, "Elko independent." is in the METS/ALTO field <ndnp:titles> and the Chronicling America website displays "Elko independent. : (Elko, Nev.) 1915-current" as the newspaper publication title and "Elko independent., July 09, 1915" as the issue title. The ENT project also reused titles from the finding aid, but without enhancement: for example, the finding aid title was the same as the digital object title, "Awards and honors: Dance Educators of America, press clippings and telegram" and the date was placed in a separate field.[42] This allowed Digital Collections to create a script that matched up the digital objects with the corresponding archival description and added digital object links[43] to the finding aid.

---

[40] "The National Digital Newspaper Project (NDNP) Technical Guidelines for Applicants," 19–35.

[41] "Photographs of First rally of the Frontier Strike, Culinary Union, Las Vegas (Nev.), 1991 August 07 (folder 2 of 3)," Culinary Workers Union Local 226 Las Vegas, Nevada Photographs, 1950s–2006, PH-00382, Special Collections, University Libraries, University of Nevada, Las Vegas, Nevada, http://d.library.unlv.edu/digital/collection/cwu/id/292/rec/1.

[42] "Awards and honors: Dance Educators of America, press clippings and telegram," Donn Arden Papers, 1918–1994, MS-00425, Special Collections and Archives, University Libraries, University of Nevada, Las Vegas, Nevada, http://n2t.net/ark:/62930/d1cg6c.

[43] UNLV used Archival Resource Keys (ARKs), which are URLs that are persistent identifiers.

Table 3. Comparison of titles

	Finding aid titles	**Digital object titles**
CWU	First rally of the Frontier Strike	Photographs of First rally of the Frontier Strike, Culinary Union, Las Vegas (Nev.), 1991 August 07 (folder 2 of 3)
NvDNP	Elko independent	Elko independent. : (Elko, Nev.) 1915-current
ENT	Awards and honors: Dance Educators of America, press clippings and telegram	Awards and honors: Dance Educators of America, press clippings and telegram

Archival description is often inherited, so Digital Collections also elected to add any existing parent description (e.g., series and subseries titles) from the finding aid to the ENT and CWU digital objects. While this is not as ideal as individually customizing the title and description of each digital object, it is an efficient way to ensure that relevant information is included and searchable in the digital object metadata.

Figure 2. Part of a metadata record for a digital object from the ENT project

Digital ID	ent000406
Title	1950s rough costume design sketches: originals
Description	Professional work in the United States: show production materials -- Las Vegas shows: production materials -- Fire and Ice, Hacienda Hotel
Physical extent	6 Items
Collection Identifier	MS-00573
Original Collection	Jerry Jackson Papers
Collection Guide	http://n2t.net/ark:/62930/f1rp63
Date	1980; 1981; 1982; 1983; 1984; 1985; 1986; 1987; 1988; 1989
Subject (FAST Topical)	Costume design Dancers Showgirls Theater--Production and direction Revues-Librettos
Individual as Subject	Jackson, Jerry
Group as Subject	Hacienda Resort Hotel & Casino (Las Vegas, Nev.)
DC Type	Still Image

UNLV Special Collections and Archives uses ArchivesSpace, a Web-based archives information management system, to create and store finding aids. Staff exported collection inventories into spreadsheets for Digital Collections to reuse for digital object metadata for ENT and CWU. Titles and dates were the main elements reused for the digital object metadata, but physical extent was also added if available. Information about the archival collection, including a link to

the collection guide, was included in the metadata so that users can easily navigate between the digitized materials available online and the archival description of the entire collection.

Subject terms were added to digital objects for the ENT collection but not for CWU or NvDNP. While the bulk of the other metadata for ENT and CWU was readily available from the finding aid and only needed to be enhanced or reformatted, subject terms for each digital object required analysis and selection by project staff (in this case, student assistants). CWU did not invest time into adding subject terms, instead this will be a separate project at a later point in time. NvDNP did not use controlled vocabularies for normalizing the metadata content due to the lack of subject fields in the Chronicling America repository. The metadata editor used by the students (CONTENTdm Project Client) for ENT already contained a local selection of Faceted Application of Subject Terminology (FAST) subject headings for students to choose from, but to make the process more efficient for the ENT team, a smaller list of relevant terms was created for students to prioritize. Controlled terms for individual and group names were also preloaded so that students only needed to choose the relevant terms rather than constructing them from scratch or searching name authority files.

Metadata for child items within complex digital objects varied across the UNLV large-scale digitization projects. The only child-item metadata created for digital objects in the ENT project was a digital identifier and the uncorrected OCR transcript. For the CWU project digital identifiers and titles were created at the item level, but the child-item titles were basically the same as the digital object title, for example, "Ladies from Europe pose with Jim Arnold, Culinary Union, Las Vegas (Nev.), 1990s (folder 1 of 1), image 1." UNLV prefers the CWU approach to creating item-level titles since the child-item title is used as the webpage title by CONTENTdm. Unfortunately, the constraints of uploading and creating child-item metadata in CONTENTdm caused the Digital Collections department to choose a simpler and faster approach for the ENT project with the intention that the metadata can be enhanced in the future.

The NvDNP case is quite different, not only because it uses a completely different system for displaying the digital objects, but also because the browser tab title combines the values from several metadata fields (title, publication place, newspaper time span, date/year of issue) and consistently displays them on parent and child levels, for example, "Elko independent. (Elko, Nev.) 1915-current, July 09, 1915, Image 1." The majority of the metadata fields on batch, reel, and item level host technical metadata (image quality, reel quality, file sizes, density readings, etc.), which is valuable information for the hosting institution, the Library of Congress. The fields designed for researchers encode titles, publication, and spatial information. In addition to the user-centered metadata fields, all digital objects undergo OCR processing and support full-text searching.

Using OCR software to create full-text transcripts of textual materials, even uncorrected transcripts, is a powerful and efficient way to enhance item-level access to materials. It can partially compensate for a lack of subject headings and other descriptive metadata that is often forgone in large-scale digitization projects like NvDNP and ENT. Unfortunately, OCR cannot be used for the photographs in the ENT and CWU projects, but technological innovations may eventually result in applications as widely used as OCR software that recognize and generate subject terms for images.

Conclusion

The three different large-scale digitization projects had very different project life cycles, team compositions, and materials, yet their deliverables were similar and equally beneficial for researchers: digital objects accessible online. Comparing the experiences and discussing the peculiarities and differences of the projects demonstrates the diversity of digitization methods, the numerous project management options, and the complexity of decision-making. Unfortunately, there is no "one size fits all" model—all projects are unique and the digitization approaches should be carefully selected to best fit the project, the archival materials, and the project team. Assessment and iteration are critical components of any large-scale digitization project that lead to efficiency and deliverables with outstanding quality.

Iterative or extensible processing is an important concept in archives that supports the argument for doing a minimal amount of work for basic access to a maximum number of collections, and doing additional work later to enhance access to specific materials based on user demand and other ongoing assessment.[44] The same concept can be applied to digitized archival materials: efficiently create and preserve high-quality digital surrogates with enough basic metadata to provide access to a large amount of unique materials that would otherwise only be available physically within an institution's reading room. Later on, additional description and other enhancements can always be added. Promotion and communication with researchers is an important follow-up to large-scale digitization to determine if the materials warrant additional description or exhibition; for example, UNLV Digital Collections monitors the use of online materials using Google Analytics.[45] By tracking online users and communicating with researchers, frequently used collections can be flagged for further descriptive enhancement.

Large-scale digitization projects will be different for individual institutions, depending on their size, available technologies, and financial opportunities. However, all large-scale projects require thorough documentation and project planning in order to be successful. UNLV hopes that by piloting different types of large-scale projects it can share strategies and lessons with the wider archival community and contribute to the development of best practices that will allow other institutions to scale up their own digitization programs. Ultimately, the core values of access and use of archival collections[46] are furthered by large-scale digitization, making more unique primary sources digitally available to a global audience.

---

[44] Daniel A. Santamaria, *Extensible Processing for Archives and Special Collections: Reducing Processing Backlogs* (Chicago: Neal-Schuman, an imprint of the American Library Association, 2015).

[45] See appendix A.

[46] "SAA Core Values Statement and Code of Ethics," Society of American Archivists, May 2011, https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

Appendix A

Web analytics for UNLV large-scale digitization projects

	ENT	CWU	NvDNP
Date range	December 22, 2017–October 21, 2018	June 9, 2017–October 21, 2018	October 1, 2015–April 15, 2018
Numbers of days analyzed	303	499	928
Page views	16,906	5,236	172,872
Average page views per day	55.8	10.49	186.28
Average time on page	00:00:18	00:00:27	00:01:12