

Maximizing Assets and Access through Digital Publishing

Opportunities and Implications for Special Collections

Ellen Engseth and Marguerite Ragnow

The future is now. Technology is driving us forward into new ways of thinking about how we interact with our constituents and how we surface content. New opportunities to maximize our institutional assets, such as collections of distinction, present themselves every day. How we respond to these opportunities is one measure of our flexibility and adaptability—in effect, our prospects for successfully traversing an ever-changing technological landscape.

Long recognized as important to libraries of all types, high-impact holdings are often central to our scholarly and cultural contributions, and libraries have a strong tradition of both sharing and making institutional use of these collections. Archivists and librarians are building on this strong tradition by capitalizing on new technologies and innovative forms of teaching and learning tools that facilitate scholarly communication and quick global reach. One option for realizing this reach is to publish and license library materials through third-party digital

publishers. Already viable for many libraries, this will become an increasingly possible and technologically transformative option that will, among other things, extend the digital life of special collections. Digital outreach and user accessibility through digital products is the future for special collections.

The prominent “inside-out” model for today’s library features, in part, archives and special collections. In this model, the library distributes institutional assets and research expertise to both local and global audiences, juxtaposed with a traditional “outside-in” model wherein a collection is bought and gathered in primarily for a local user community. For example, Dempsey et al. assert that for academic libraries, the “inside-out orientation will become more important as universities focus attention on distinctive institutional assets and libraries direct increased curatorial attention toward special collections, new scholarly products, research preprints, and pedagogical resources.”¹ Concurrently, archives and special collections are receiving more attention from those outside of special collections for their distinctive roles in teaching, learning, and research—something long known to those who daily work with these materials. This wider interest drives demand for more curated digital materials, which supports and often directs the process of scholarship. In the inside-out model, curatorial expertise is required and “turned ‘outward’ to help contextualize and characterize the value of institutional holdings.”² However, the necessary staff, time, and technical resources to effectively curate and host the digital scholarly conversation, though certainly developing on the academic campus and in the nonprofit environment, are limited.

The greater attention generally enjoyed by special collections in the recent era will continue, in part due to the increase in the number of libraries adopting the inside-out model. The current digital age presents striking opportunities for these special collections: for digital preservation and access; in outreach, branding, and revenue streams; and for more public roles in knowledge production and the evolving research ecosystem. This environment “requires new ways of thinking”—both now and in the future.³

Publishing digital content through a third party to reach goals of outreach and asset building demonstrates a new way of thinking. Third-party vendors and ventures now present libraries with options for digitizing, licensing, and sharing high-value or high-impact holdings. They also provide opportunities for realizing revenue streams, reaching new or wider audiences, and relieving libraries of some of the technically prohibitive or difficult tasks associated with digital projects. The resulting digitized special collection is a next-generation institutional asset, one that is useful for preservation, sharing, and reuse; building reputation; and providing rationale for digital infrastructure. This chapter shares our experience and analysis

of partnering with a nonlibrary vendor for digitizing and publishing as one example for building institutional assets and enhancing access.

Further, although libraries are seen generally to operate with different values than those of profit-driven organizations, and there may be an aversion to intentional asset building inherent in library culture, strategic use of our collections falls squarely within the framework of library entrepreneurship. The continuing era of funding shortages demands additional and alternative sources of funding. While challenging, this environment clearly presents opportunities for innovation. Third-party partners or commercial enterprises have long proved useful to other cultural heritage institutions and will increase in value and usefulness to libraries interested in pursuing innovation in the future.⁴

The technical arena adds crucial layers to today's already complex and multifaceted library environment. At the University of Minnesota (UMN) Libraries, the Archives and Special Collections Department utilizes a complicated matrix of digital delivery modes, platforms, and products all to extend access and reach. Some components were developed or provided by internal skills, staff, and resources. Others are collaborative at some group level or interact with non- and for-profit third parties. Table 12.1 reveals the array of our practices at the time of the digital publishing project described in this chapter.

TABLE 12.1

A partial high-level view of digital platforms and discovery/delivery tools for archives and special collections. The third-party commercial publishing platform discussed in this chapter is italicized.

Audience	Scale		
	<i>Institutional</i>	<i>Group/noncommercial</i>	<i>Commercial third-party</i>
Local	Shared network drives		Sharing software (e.g., Google Drive) FTP servers
Group	Shared network drives		Sharing software
Web/global	Curated exhibits (e.g., Omeka, bespoke platforms) Digital collections and research data repositories Web-based collection management and access systems (e.g., ArchivesSpace, legacy tools)	Catalog records (OCLC) Specialized portals (e.g., Minnesota Digital Library, DPLA)	<i>Platform provided by third-party vendor</i>

*The concept for this table was adapted from Lorcan Dempsey, Constance Malpas, and Brian Lavoie, "Collection Directions: The Evolution of Library Collections and Collecting," *Portal: Libraries and the Academy* 14, no. 3 (2014): 416.

We reached our decision to digitize, license, and publish our materials through a third-party digital publisher in light of the environment sketched above and only after much consideration.

The authors of this chapter are the curators who head two repositories within the Archives and Special Collections Department: the Immigration History Research Center Archives (IHRCA) and the James Ford Bell Library (Bell Library). The IHRCA is an archive and library for the study of immigration, ethnicity, and race; the Bell Library documents the history and impact of international trade and cultural exchange prior to circa 1800 CE. Both strive to preserve collections and make them accessible and discoverable locally and globally, with an emphasis on research and learning. As our discussion of the inside-out library affirms, it is increasingly important to find innovative and cost-effective ways to share our materials digitally. The UMN Libraries plays an integral role in campus life and is a significant contributor of resources and programs to the academic community and the general public. We and our colleagues provide a highly collaborative environment that is distinguished by new models for teaching and learning, research support, and scholarly communication. As the UMN Libraries is known for advanced web services, significant collaborative digital library development, and a record of innovative partnerships, it had the infrastructure to consider the third-party vendor projects we were interested in pursuing: the IHRCA's "Migration to New Worlds" and the Bell Library's "Age of Exploration."

TECHNOLOGY

Digital projects come in all shapes and sizes, and each one may require different technological parameters. They all have these requirements in common, however: (1) technology necessary to convert analog material to digital format, along with a computer and software to ensure quality control; (2) a platform and software through which to make the digitized material available to the target audience; and (3) storage for the digital objects both in the short term and for the long term.

All digital conversion projects, regardless of the technology used, have multiple phases. Once the items for conversion have been selected and prepared for digitization, most projects will include the following steps:⁵

- scanning text or image files or the conversion of video or audio files
- creating digital master files
- digitally processing the captured data and producing derivative files

- collecting and recording metadata for each digital object or set of objects
- ingesting the digitized objects and their associated metadata into a digital management system
- preserving and storing the digital objects created
- creating a user interface for making the digital objects accessible

As this list suggests, multiple technologies, many working in concert, are required to successfully mount a curated digital primary source database with an effective public interface. This is true whether the project is small, featuring a single collection, or large, bringing together multiple collections. For the projects discussed in this chapter, many of these steps were the responsibility of our publishing partner, Adam Matthew Digital (AMD), which is an imprint of SAGE Publishing and an award-winning commercial publisher of digital primary source collections for the humanities and social sciences. Sourced from leading libraries and archives around the world, their unique research and teaching collections cover a wide range of subject areas, from medieval family life to twentieth-century history, literature, and culture. The investment AMD has made in developing attractive and very user-friendly public interfaces for their products was one reason we were drawn to them as partners.

While the IHRCA project did include the conversion, translation, and transcription of some audio sources, we'll focus here on the technology used for the conversion of paper-based analog objects that were common to both projects: books and bound and unbound documents and maps. Key to the success of many digitization projects is the use of a rapid-capture process for quickly digitizing materials. Raw rapid capture was first created specifically for the National Archives and Records Administration. Rapid capture refers to the speed of workflow and was initially used primarily for flat single-sheet documents. New technology, however, enables even rare book digitization to be accomplished more rapidly. Some projects have resulted in 1,000 to 1,500 scans per day for unbound documents, whereas others have scanned from 150 to 300 pages per day of bound material. Bound material presents more of a challenge than unbound material. The tightness of the binding may require images to be photographed from particular angles, for example, or page size within a single binding may vary. Each of these conditions may require frequent camera adjustments.⁶

We reached an agreement with AMD that the digital conversion would be outsourced but that it had to take place at the UMN. We weren't comfortable sending our materials off-site; on-site conversion also would enable us to more closely monitor progress and deal with special handling issues as they arose. AMD

contracted with LUNA Imaging Inc. to undertake the conversion. In consultation with LUNA, AMD, and the UMN Libraries' digital experts, we agreed that it made sense to use the same rapid-capture system for both projects. Despite the format differences, both collections were best suited to planetary, overhead image capture rather than a sheet-fed system. LUNA hired an experienced local digital photographer to do the image capture using a proprietary rapid-capture system with components integrated and configured by LUNA. The basic components were a vertical copy stand, a full-frame camera as the capture device, and UV-filtered light fixtures containing high-intensity discharge lamps. The camera height on this system is adjustable, and the cameras can be switched out based on the requirements of the individual objects. Bound items were captured by also employing a LUNA-designed ninety-degree book cradle. LUNA shipped all of its equipment to the UMN Libraries for the project.

There are other vendors as well as similar equipment, less the proprietary elements developed by LUNA, that are available for both internal and outsourced projects. For example, other types of planetary scanning systems, such as the i2S Copibook HD with freestanding fluorescent lights, have been used at the UMN for a variety of projects; we also have used a Fujitsu fi-6230 Color Duplex Flatbed scanner for materials that can be placed in a high-speed sheet feeder.⁷

Both of our AMD projects included a variety of types of analog formats: single-sheet documents; documents held together by metal clasps, staples, and paper clips; bound documents; rare books; vellum manuscript documents; oversized flat maps; and books and scrapbooks with inset images or maps that also were often folded, in addition to the audio files noted above. The LUNA technician worked closely with the project manager on the IHRCA project to prepare the documents for scanning, including the removal of clasps, staples, and other temporary bindings. He also worked with the Bell Library curator to ensure the proper handling of rare materials, foldouts, and vellum manuscripts that had to be photographed through clear Mylar sleeves.

LUNA staff worked closely with AMD staff to run quality checks and create master and derivative files. Master files were captured as 300 dpi TIFF files and processed to produce 300 dpi JPEG files for display.⁸ As curators, we participated in image checks and file-name convention development, but the balance of the digitization process was handled by AMD for ingesting into their proprietary data management system.

AMD partners with federated platform providers such as Ex Libris, ProQuest, EBSCO, and OCLC to ensure discoverability within their products, including

through Primo Central, Metalib, SFX Knowledge Base, Summon, EBSCOHost, WorldCat, and WorldCat Local. Their curated and contextualized online databases also utilize 360-degree object viewers, interactive maps, and a proprietary federated searching tool (a widget) that can be added to a library's website to facilitate searches. Metadata in MARC 21 format is fully downloadable.

Short- and long-term storage issues for the digital assets were faced by both the publisher and our library. The publisher uses a storage vendor, Portico, and its D-collection preservation service for long-term dark storage and data preservation. The master files supplied to our library under the contract with the publisher will be stored on our storage area network (SAN) storage servers, which provide for both redundancy and tape backup as well as off-site storage for all files.

CASE STUDY

Our goals in undertaking these projects were to expand global reach and increase access to support teaching and research. For this chapter, however, we will concentrate on our decision-making process and the technological aspects of the projects.

Adam Matthew Digital representatives first approached each of us separately about contributing to "Migration to New Worlds" and "Age of Exploration" products, and independently of one another, we were intrigued by the opportunities their proposals presented. When we discovered a shared perspective on both potential benefits and concerns regarding subscription-based products, we determined that it would be ideal to work together to (a) evaluate the benefits of participating in such projects, (b) explore issues of access and cost, and (c) steer the proposals through the contract development and negotiation phases. The UMN Libraries had not yet participated in any such projects. Our strategy was to work together as copilots for the libraries' first foray into a royalty-producing digital publication of primary sources. This strategy proved fruitful. For example, together we consulted with our internal copyright specialist on contractual issues related to access. Although much of the IHRCA materials required rights considerations whereas the Bell Library materials are public domain, we were both concerned about licensing and subscription access.

As librarians generally committed to the ideal of open access, we were initially resistant to participation in subscription-based products. As we learned more about how this public-private partnership would work, we also had serious concerns about a paywall, which would be part of the license agreements. The materials

we contributed could not be used in other large-scale digitization projects until a certain number of years elapsed; in other words, they would remain behind a paywall for the duration of the licenses. Additionally, only people who had access to the products through an institutional subscription would have digital access to our contributed content. However, through more discussions with AMD and with other UMN Libraries staff, we realized that these issues were not the insurmountable obstacles they initially seemed to be.

We came to understand that the investment in technology made by AMD to create these products could not be matched by our institution at this time. Nor did the content to be included in these products fall within the existing parameters of our internal strategic digitization plans. We realized that if we did not contribute our materials, then only those people who could physically visit our collections would have access to them—a much more limiting prospect. Moreover, the paywall would in no way interfere with our ability to make the materials available to our researchers on the same basis as we always had. Individual researchers, regardless of their ability to access the AMD online products, would still be able to use and exploit the analog items, including requesting digital copies on an individual basis.

The time embargo imposed by the paywall⁹ also became a nonissue, as we had no plans for including the materials in other large-scale projects during the term covered by our licenses. Once this license expires, we will be free to provide open access to this content through our institutional repository or any other avenue. The paywall was likewise a concern of UMN Libraries administrators, but they agreed that limited digital access was better than no digital access, and they were willing to commit to long-term maintenance of the digital assets created through these projects both during the paywall period and once released to us.

Another issue when using a third-party provider is what happens should that company be sold or go out of business. Fortunately, AMD provides for a dark storage and data recovery service (referenced above) that also ensures continued subscriber services for the life of each existing subscription. When we understood how the recovery process worked, we were satisfied that this did not pose a significant risk.

Once we made the decision to license our content with a third-party publisher, there were staffing and logistical issues to consider. We hired a part-time temporary project manager. This staff person led collection management, document preparation, and communications with the outside providers. As noted, this included working with the digital technician brought on-site by LUNA Imaging as well as liaising with AMD staff. Hosting an outside digital-service provider also necessitated finding adequate space on campus. In order to provide this space, the UMN

required an addendum to our licensing agreement that addressed the use of university real estate by an outside party. In addition to work space, we needed to provide storage for the scanning equipment shipping containers. To accommodate the digitization operation, we had to displace student staff and relocate them elsewhere for the duration of the project. Furthermore, we provided the digital photographer with a UMN e-mail account to support rapid and robust digital content delivery.

Our decision to participate in these particular projects with this particular vendor was based on numerous factors. One of the benefits we identified early in the process was the attractive and easy-to-use public interface and the user support provided with all AMD products. We and our faculty and students were already familiar with these benefits through our own institutional subscriptions to some of their existing products. Also, because some of our content had either minimal or no MARC records, AMD offered improved and original MARC records for the digital surrogates ingested into the database, representing a significant cost savings to us. Additionally, the IHRCA project would result in fully searchable digital versions of increasingly obsolete and degrading audio tapes.

As librarians, we are always interested in access and reach, and user-friendly, high-quality educational publications that are well placed in the global information landscape will help us reach these goals. When they came to us with their proposals, AMD's team had already conceived of the project themes and found institutional partners that ensured a broad array of primary source materials, which would result in multi-institution cross-searchability. They also surveyed our collection content to narrow down the proposed selections for inclusion in the projects. These actions represented significant cost savings to us when compared to undertaking such projects in-house. Perhaps the most important savings was the digitization of our material at no direct cost to us. The resulting 68,000 images and nearly twenty hours of audio recordings would have cost us approximately \$35,000 had we digitized them ourselves. These benefits are in addition to the marketing and user-support services our publication partner will provide for the duration of the license. Moreover, our entire university community—composed of students, staff, and faculty at five campuses around the state—now enjoys access to these publications. Additionally, both of our collections will realize royalties from the licensed materials, income streams that may have a significant impact on our ability to create our own digital assets and curated products in the future.

For all of the benefits and cost savings these projects afforded, we realized that the UMN Libraries would incur some costs. Each curator had to determine if she had the time to take on the project. Meetings with staff and libraries administration,

meetings with the third-party vendors, and hiring and supervising additional project-specific staff would be time consuming. For us, the benefits we foresaw compensated for the time we anticipated spending on the projects. In retrospect, we underestimated that time. Contract reviews with legal counsel, evaluating space and equipment needs, and negotiating with both libraries administration and the vendor was time that was either unanticipated or took longer than expected. Ingesting the metadata created and provided by AMD as well as rights information also absorbed staff time. Going forward, the libraries will invest time and funds in managing the digital surrogates, including providing for dark storage of master files. These were projects that we did not have the internal capacity to undertake ourselves, and therefore, this represents a time and funding commitment not anticipated by the UMN Libraries in its long-term planning.

Overall, we are pleased with how the process unfolded and anticipate reaching all of the project goals. We now have digital assets that we are sharing with the wider community, which both extends our reach and contributes to the research and learning ecosystem. Our experience working with a commercial vendor in this way will, we believe, pave the way for other such partnerships in the future.

CONCLUSION

It is time to begin thinking about outreach and access as part of an investment strategy for building user communities on a global stage. The combination of continuously evolving technology and library administrators' increasing focus on reducing costs makes the inside-out model viable now and a future standard for special collections. Creating digital assets and making them accessible on a broad scale is part of that model. Our experience with a commercial digital publisher demonstrates that partnering with such a third-party vendor can provide a level of access and discoverability that many libraries cannot afford to provide for themselves, as the cost of developing and managing similar projects in-house is beyond their capacity. Third-party publishers often will sustain those costs while at the same time provide a new revenue stream for their library partners that might help ease the tension between preservation and access with respect to the allocation of institutional resources.

Strategic digitization of library assets can be beneficial no matter the size or nature of a particular collection or library, whether academic, public, or private. And these benefits do not have to be realized with a third-party commercial vendor.

Many libraries use digital images of collection materials to promote their collections on social media platforms such as Facebook, Instagram, and Tumblr or as part of online exhibitions, courses, or other projects. New technologies can also assist librarians to innovate teaching and learning and enhance scholarly communication. For most librarians, providing access to our materials is a key motivator, and finding partners for digitization projects, whether commercial vendors or other institutions with similar interests, may help improve access and broaden reach more affordably than attempting to undertake a large digital project on one's own.

The advances in image capture and other types of digital conversion technologies—as well as improvements in data management systems, international standards for output and display such as IIIF (<http://iiif.io>; see chapter 13 for more about this standard), options for long-term digital preservation, and access and discoverability tools—are already changing the ways librarians work and think about their collections. The inside-out model supports the adoption of these technologies and provides a rationale for this type of partnership with third-party publishers. The future of the outward-facing special collections library is entrepreneurial; as we continue to explore improved access and discoverability and extend reach on a global scale, digital publishing technologies offer one option for realizing this future now.

NOTES

1. Lorcan Dempsey, Constance Malpas, and Brian Lavoie, "Collection Directions: The Evolution of Library Collections and Collecting," *Portal: Libraries and the Academy* 14, no. 3 (2014): 420, <https://doi.org/10.1353/pla.2014.0013>.
2. Lorcan Dempsey, "Library Collections in the Life of the User: Two Directions," *LIBER Quarterly* 26, no. 4 (2016), <https://doi.org/10.18352/lq.10170>.
3. Dempsey, Malpas, and Lavoie, "Collection Directions," 394.
4. See, for example, Elizabeth Kirk, "The Entrepreneurial Library: Creating Revenue in the Age of E-commerce," in *Crossing the Divide: Proceedings of the Tenth National Conference of the Association of College and Research Libraries*, ed. Hugh A. Thompson (Chicago: Association of College and Research Libraries, 2001). See also Peter Hirtle, "Archives or Assets?," *American Archivist* 66, no. 2 (Fall–Winter 2003): 236, <https://doi.org/10.17723/aarc.66.2.h0mn427675783n51>: "In short, archives and manuscript repositories control and manage assets worth hundreds of billions, if not trillions, of dollars. . . . The sale of reproductions of archival materials and the licensing of material for commercial use are becoming ever more important as possible sources of income for archives. Archives, in seeking to draw revenue from the content that they own, are following in the footsteps of museums, which have long used licensing programs to augment their income."

5. Allison Zhang and Don Gourley, *Creating Digital Collections: A Practical Guide*, Chandos Information Professional Series (Oxford: Chandos, 2009).
6. See, for example, the variety of results discussed in Ricky Erway, *Rapid Capture: Faster Throughput in Digitization of Special Collections* (Dublin, OH: OCLC Research, 2011), www.oclc.org/research/publications/library/2011/2011-04.pdf.
7. Erway, 12–14.
8. There have been numerous standards used for digital projects. See Iris Xie and Krystyna K. Matusiak, *Discover Digital Libraries: Theory and Practice* (Oxford: Elsevier, 2016), 69–71, for a summary of best practices.
9. This multiyear embargo varies from institution to institution based on a variety of factors. Our contract requires that we not disclose specific terms.