EDITORIAL

Cruel to Be Kind

Reviewing for Library Resources & Technical Services

Michael Fernandez and Rachel E. Scott

A recent call for volunteer peer reviewers for American Library Association Core division journals was incredibly successful—more than 170 answered the call and expressed their willingness to review. As we prepare to onboard and train the volunteers, we thought to share some of the highlights with you. As suggested by countless Reviewer 2 memes, peer review has been the scene of many battles—intellectual, petty, and anywhere in between. A peer reviewer once described a certain *LRTS* editor's writing as "gauche" and "undisciplined," to cite one painful example. On another occasion, peer reviewers failed to draw attention to an omission that continues to haunt this particular editor. But we digress: in most cases, peer reviewers are constructive, helpful, and insightful. Even though it may not feel like it, the point of peer review isn't to traumatize but rather to improve and strengthen manuscripts. It is in this spirit of constructive criticism that the *LRTS* editorial team onboards reviewers.

Working with the ALA Core division journal editorial team, we devised a shared review template. We ask all reviewers to consider these questions and provide feedback to the editors based on these considerations. The template addresses several broad questions dealing with the relevance to the readership; novelty of contribution to the literature; adherence to conventions of academic writing; whether the literature review places the study in perspective and builds on existing research; the clarity, reliability, and appropriateness of the methods; and the implications of the findings for scholarship or practice.

At the head of the reviewer template, there is a link to the Use of Generative AI by Authors and Reviewers Policy. The Core journal editors agree that peer review is uniquely valuable because it is a human process that leverages the reviewer's expertise and experience. Accordingly, reviewers may not use generative artificial intelligence (AI) technologies to assist with any aspect of their review. Other ethical considerations include the responsibility of reviewers to disclose a potential conflict of interest to the editors. The journals use a double-anonymous review process to minimize bias. We ask reviewers to decline review invitations when they suspect that they may know the authors.

Throughout the training, we emphasize the importance of communication. If reviewers have any questions or concerns, they should reach out to the editors. In the course of reviewing an article, a reviewer may have reason to suspect undeclared AI use, knowledge of the authors' identities, or inconsistencies in reporting. There is no disadvantage to reaching out.

Michael Fernandez (fernm@bu.edu) is Head of Technical Services at Boston University Library, https://orcid.org/0000-0001-7912-7036. **Rachel E. Scott** (rescott@illinois.edu) is Professor and Head, Acquisitions and Cataloging Services at University Library, University of Illinois Urbana-Champaign, https://orcid.org/0000-0001-5847-3378.

After the editors receive the reviewer reports, we reconcile what is sometimes contradictory information with our own understanding and assessment. In preparing editorial decisions, we work hard to make sure the reviews are presented in a constructive manner. Reviewers have the opportunity to present certain information to the editors only and some information to both the authors and the editors. We share the information intended for authors with authors—unless it is unprofessional, unproductive, or unkind. In the very rare occasions reviewers include such feedback, the editors are well within their rights to excise it.

By providing guidelines for peer reviews, the Core journal editors establish clear expectations for the review process. Editorial decisions, informed by review reports, enable authors to refine and enrich their work with expert insight. The peer-review process is not unlike the iterative optimization inherent to technical services work. This issue of *LRTS* contains case studies and research articles that demonstrate the value of the systematic assessment and appraisal our peer reviewers lend to the work.

Communications on Practice

In "Making Hidden Collections Visible: A Case Study of Human–AI Collaboration in Archival Metadata Generation," Jeehyun Davis, Alayne Mundt, and Leslie Nellis document work evaluating JSTOR Seeklight, an AI-assisted tool designed to generate descriptive metadata and transcription for digitized archival materials. Five archival collections at American University Library were tested to assess the tool's ability to produce titles, descriptions, subject headings, named entities, and creator fields across a range of formats. The authors found that Seeklight was most successful in generating baseline titles and descriptions for visual materials with textual context and least successful in generating authority-controlled fields. They recommend AI-assisted metadata generation as a workflow enhancement that complements rather than replaces professional expertise.

Features

Christopher A. Palazzolo, William Jacobs, and Matthew J. Jabaily report on a survey of academic and research librarians' experiences with ebook evidence-based acquisition (EBA) models in "A Comparative Study of the Implementation, Management, and Evaluation of Evidence-Based Acquisition Ebook Plans in Academic and Research Libraries." EBA models have emerged as an opportunity to balance broad access to ebook content with librarian autonomy and agency. The results explore librarians' engagement with and management of EBA models and document a variety of motivations for EBA adoption, as well as challenges and preferences related to EBA budgeting, management, and title and program selection.

In "'Unbury' the Book Chapter?: An Investigation of Library and Information Science Book Chapters from Edited Collections in Scholarly Indexes," Juleah Swanson examines the truth of the perception that by publishing a book chapter in an edited collection, the author has "buried" it. Swanson evaluates the discovery of library and information science book chapters in Google Scholar, Web of Science

Book Citation Index, and Dimensions and finds inconsistent indexing across all three indices. Although Swanson confirms concerns about the discovery of book chapters, she helpfully outlines avenues to exhume them by improving discoverability.

Notes on Operations

Jennifer A. Mezick, Louis T. Becker, and Kat Brooks outline a case study on approval plan assessment undertaken at the University of Tennessee, Knoxville, in "Measure, Change, Measure: Assessing Changes to a Traditional Approval Plan." Having updated approval plan parameters considerably between 2019 and 2023, the authors review the purchase and usage data from these periods to investigate the performance of the plan. They find that using this data to update approval plans positively impacts the quantity and usage of approval plans titles.

In "Ebook Call Number Project: Automating Library of Congress Classification Numbers to Ebooks for Reporting," Bonita Pollock, Zhao Liu, Amanda Yesilbas, Laura Pascual, and Skyler Schultz outline their automated process for adding Library of Congress Classification numbers into Alma (Library Management System) call number fields to enhance ebook metadata. The authors developed a Python-based automation tool using OCLC's Connexion API and implemented a workflow within the Ex Libris Alma Library Management System that leveraged existing metadata. This project has significantly improved the accuracy of discipline-specific collection assessment and reporting. The authors offer alternative workflows for librarians who have limited programming expertise or support.

Book Reviews

Reviewed in this issue is *Collection Management Basics,* Eighth Edition, by Margaret Zarnosky Saponaro, John Novak, and G. Edward Evans.

COMMUNICATIONS ON PRACTICE

Making Hidden Collections Visible

A Case Study of Human–AI Collaboration in Archival Metadata Generation

Jeehyun Davis, Alayne Mundt, and Leslie Nellis

Introduction

Archives and special collections increasingly face an imbalance between the rapid expansion of digitized materials and the slower pace of descriptive metadata creation needed to support discovery and access. This article presents a case study of a pilot project conducted at American University Library to evaluate JSTOR Seeklight, an artificial intelligence (AI)-assisted tool designed to generate descriptive metadata and transcription for digitized archival materials. Academic libraries have long served as stewards of unique and irreplaceable primary resources housed in archives and special collections. A substantial portion of these materials remains confined to physical formats, discoverable only through limited or incomplete description and accessible primarily to on-site researchers.[1] Despite decades of professional attention to improving access and discovery of these materials, progress has often been incremental. At the same time, the scope of digital stewardship has expanded well beyond the initial act of resource digitization.

Contemporary digital stewardship encompasses the full life cycle of digital assets, including selection and rights assessment, digitization workflows, descriptive metadata creation, discovery systems integration, preservation storage, and long-term governance and sustainability.[2] The Digital Preservation Coalition's *Digital Preservation Handbook* frames preservation as an ongoing organizational commitment that integrates policy, infrastructure, and workflows rather than a one-time technical task. Crucially, description and access are not secondary outcomes of stewardship; they are central to its purpose.[3]

Simultaneously, archives and special collections staff continue to confront persistent backlogs and hidden collections, a challenge widely documented in the archival literature.[4] Calls to recalibrate labor-intensive descriptive practices in favor of scalable access—most notably articulated in Greene and Meissner's "More Product, Less Process"—underscores how traditional processing norms can inadvertently delay availability to users.[5] In practice, digitization capacity frequently outpaces descriptive capacity. Although scanning can be accelerated through investment in equipment and staffing, metadata creation remains constrained by specialized expertise, time, and uneven staffing levels, producing a critical discovery bottleneck.

Within this context, academic libraries are increasingly experimenting with AI-assisted description—not as a substitute for professional judgment, but as a means of increasing throughput, improving baseline discovery, and creating scalable foundations for human review. Emerging scholarship

Jeehyun Davis (jeehyundavis@american.edu) is University Librarian, American University, https://orcid.org/0009-0006-7719-5696. **Alayne Mundt** (mundt@american.edu) is Director of Technical Services, American University, https://orcid.org/0000-0003-0761-514X. **Leslie Nellis** (lnellis@american.edu) is Head Archivist for Special Collections and Digital Initiatives, American University.

suggests that AI in archival description should not be understood as professional displacement, but instead as a catalyst for evolving competencies. Automation is likely to reshape core functions, requiring archivists and metadata professionals to cultivate new skills in data literacy, algorithmic oversight, ethical decision-making, and quality control.[6] Early experiments with generative AI tools for description, subject extraction, and image recognition demonstrate significant potential to reduce backlogs and enhance discoverability, particularly in large-scale digitization initiatives.[7] At the same time, practitioners emphasize the importance of iterative experimentation—piloting tools on limited collections, benchmarking outputs against human-created records, and establishing local governance frameworks—so that AI adoption becomes a structured learning process rather than a one-time technological shift.[8]

Equally significant is the human dimension of this work. The introduction of AI tools represents not only a technical adjustment but also an organizational change management effort. For archival and technical services professionals, AI may evoke both curiosity and concern—raising questions about professional identity, quality standards, and the future of expertise. Positioning AI exploration as a supported learning opportunity—one with clear guardrails, shared evaluation criteria, and psychological safety—has emerged as a central leadership responsibility in contemporary digital stewardship.

Institutional Background and Leadership Intent

At American University Library, expanding access to primary resources in archives and special collections has been a strategic priority. As digitization capacity increased through sustained investments in staffing and equipment, a structural imbalance became increasingly apparent: the library could scan materials at a growing pace, but the creation of high-quality descriptive metadata—essential for discovery, access, and reuse—remained a persistent bottleneck.

Metadata creation for archival collections is labor intensive, requires specialized expertise, and depends on sustained staffing capacity that is difficult to scale. Even under optimal conditions, traditional metadata workflows can delay access for months or years after digitization. During periods of staff vacancies or competing priorities, that gap widens further. As a result, large portions of digitized content remain scanned but effectively invisible to users. This imbalance between digitization throughput and descriptive capacity prompted the library to explore responsible and scalable approaches to metadata generation that could increase access without compromising professional standards.

The JSTOR Seeklight pilot emerged from this strategic challenge.[9] As an AI-enabled tool designed to generate descriptive metadata and text transcription at scale, Seeklight offered an opportunity to test whether batch metadata generation could meaningfully accelerate discovery for selected archival materials while preserving human oversight, ethical stewardship, and metadata quality. Our participation as a Charter Member positioned the library not simply as a consumer of a new technology but as an active contributor to its development—shaping use cases, workflows, assessment strategies, and ethical guardrails alongside peer institutions.

From a leadership perspective, the intent behind this project was not automation for its own sake. Rather, the goal was to address a structural access problem that disproportionately affects archives and special collections: the accumulation of valuable materials that remain under-described and underutilized due to capacity constraints. By experimenting with AI-assisted metadata generation, the library sought to reduce the time between digitization and access, prioritize service to researchers, and extend the reach of our collections beyond local users to broader research communities. At the same time, the project was designed to assess organizational readiness by examining how new tools might reshape workflows, redistribute labor, require new competencies, and strengthen cross-departmental collaboration.

Equally important, this exploration aligned with the library's broader commitment to human-centered and ethical uses of emerging technologies. The Seeklight project was intentionally framed as a hybrid model in which AI-generated metadata serves as a starting point, rather than an end point. Human review, quality control, and professional judgment remain central to the process. This approach reflects a leadership philosophy that views AI as an augmentation of expertise rather than a replacement—supporting staff productivity while maintaining accountability, transparency, and long-term sustainability.

In this sense, the Seeklight pilot represents more than a technical experiment. It is a strategic response to long-standing access barriers in archives and special collections, a service innovation aimed at accelerating discovery, and an organizational learning opportunity that informs how academic libraries can responsibly integrate AI into core descriptive practices. The outcomes of this project—both its successes and limitations—offer insights for research libraries seeking scalable, ethical solutions to the challenge of making hidden collections visible in an era of expanding digital expectations.

Project

Goals and Rationale

For several years, Technical Services and Archives and Special Collections departments have worked to strengthen cross-department collaboration, and the Seeklight pilot project aligns closely with this work. Under new leadership and evolving workflows, print cataloging volume has decreased while demand for archival descriptions has increased. In response, the director of technical services and head archivist for special collections and digital initiatives have focused on expanding metadata support for archival materials and building capacity for sustained collaboration. These efforts have included developing workflows for print and rare materials, repurposing a vacant position in Technical Services to manage cross-department workflows and metadata production for Archives, and reallocating another position to expand support for digitizing archival collections.

We initiated the Seeklight project to improve the discoverability of the library's digitized archival resources. A primary objective was to accelerate the creation of accurate descriptive metadata so that materials could be made available to users more quickly following digitization. The project also sought

a scalable solution capable of efficiently processing large volumes of archival content and generating consistent descriptive metadata despite limited staff capacity or subject expertise. By incorporating AI-assisted description into existing workflows, we aimed to support sustainable metadata creation across diverse collections and reduce the impact of staffing gaps. An additional goal of the pilot was to establish policies and procedures for integrating AI-assisted description into established descriptive practices.

The head archivist determined which collections and materials were appropriate for inclusion in the Seeklight workflow. Not all library materials are suitable for AI-generated description, and selected collections must comply with American University policies regarding data classification and the use of AI.[10] Additional considerations include federal privacy laws, deed of gift restrictions and donor intent, copyright and ownership status, and broader privacy concerns.

Beyond these requirements, we determined that using a diverse set of collections for initial testing would provide a better understanding of Seeklight's capabilities and limitations. We selected five different collections representing a range of formats, including photographs, handwritten and typed letters, posters, historical real-estate appraisals, blueprints and drawings, and historical business and financial records. These collections enabled us to evaluate Seeklight's performance across multiple-use cases, including handwritten materials, transcription quality, folder-level versus item-level description, and visual materials with minimal contextual information.

All selected collections were intentionally limited in size, to enable rapid review of outputs and comparison across projects. This approach supported our goal of identifying which materials are best suited to Seeklight, which ones may be appropriate for large-scale ingest and metadata generation, and where additional review, remediation, or alternative workflows may be necessary. The initial collections that we loaded were as follows:

- Martin Hupka images—Peace Corps volunteer in Colombia 1963–1965
  - 444 images, partially described (unfinished) collection, previously processed by project staff
  - Selected to evaluate human- vs. AI-generated metadata
- Anne Williams correspondence—Peace Corps volunteer in India 1965–1967
  - 119 letters
  - Selected to evaluate description for handwritten material and test new transcription feature
- Patrick Frazier Political and Social Movements oversize posters—protest posters from the 1960s and 1970s
  - 188 posters
  - Selected to test visual material with little written information
- Shannon & Luchs real-estate collection—real estate firm that developed and redeveloped areas of Washington, D.C., in the 1920s to 1970s
  - 356 folders of mixed materials
  - Selected to test descriptions at the folder level

- Paul Jurmo images—Peace Corps volunteer in the Gambia from 1976 to 1979
  - 191 images
  - Selected to test no-context images

Governance and Staffing Model

The head archivist and the director of technical services jointly lead the integration of the Seeklight platform into local workflows. Together, they designed the project structure, established staff roles, and developed assessment strategies for evaluating Seeklight outputs. The head archivist oversees project setup, including file uploads, entry of contextual information into the platform's Context Box, and generation and evaluation of transcript outputs. The director of technical services oversees sample metadata review across collections, develops staff guidance and oversight procedures, and conducts secondary review of descriptive metadata. Both managers collaborate on field mappings, role assignments, and assessment methodologies.

Two library specialists in Technical Services review item-level descriptions for materials ingested into Seeklight. The acquisitions specialist also performs copy cataloging and contributes descriptive work for digitized collections. The resource description specialist provides traditional cataloging expertise. Both specialists bring extensive experience creating item-level descriptions for digitized materials in special collections.

During implementation, additional staff across the library provided targeted support. The coordinator of original cataloging advised on metadata standards, the library systems coordinator supported initial technical setup, and the library systems and assessment administrator contributed to workflow development, including the creation of a custom application to split transcript exports from Seeklight for ingest into Quartex, the library's digital collection platform.[11] The library assessment coordinator developed the assessment framework based on input from project leadership. As our proficiency with the platform grows and local documentation becomes more robust, these roles are expected to evolve.

Once workflows and local practices are fully established, day-to-day responsibilities will transition to the library's digitization coordinator and metadata coordinator. The digitization coordinator will assume responsibility for creating projects in Seeklight, adding contextual information during ingest, and generating transcripts. The metadata coordinator will oversee quality control, supervise the two specialists, and manage metadata refinement and ingest into Quartex.

The metadata coordinator position is currently vacant, creating a significant gap in fully realizing the potential of the project. This role is responsible for coordinating metadata work across Technical Services and Archives, managing workflows, reviewing descriptive metadata for digitized collections, and loading completed projects into the library's digital asset management system. While this position is vacant, the director of technical services has assumed some of these responsibilities. Although Seeklight does not replace the expertise required for this position, it has helped mitigate workflow delays by enabling progress on descriptive projects that would otherwise have been paused or progressed much more slowly.

Metadata

Seeklight generates metadata using the Dublin Core schema, which provides a relatively minimal set of fields compared with the descriptive, technical, and administrative metadata applied to records in Quartex, which uses a local schema based on MODS.[12] Of the fifteen fields generated by Seeklight, we use seven for all records and apply five additional fields conditionally based on material formats. The remaining three fields are not used (table 1).

Table 1. Metadata Mapping for Seeklight

Field	Field Definition	Using	Mapped to	Notes
Title	The name given to the resource. Typically, a Title will be a name by which the resource is formally known.	Always	Title	
Creator	An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service.	Always	Creator	
Date	A date associated with an event in the life cycle of the resource.	Always	Date Created	Uses ISO 8601. Duplicate field to Date Created EDTF in Quartex formatting.
Type	The nature or genre of the content of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content.	Always	Genre	Remediate with AAT terms outside JSTOR environment.
Description	An account of the content of the resource. Description may include, but is not limited to, an abstract, table of contents, reference to a graphical representation of content, or a free-text account of the content.	Always	Description	
Subject	The topic of the content of the resource. Typically, a Subject will be expressed as keywords, key phrases, or classification codes that describe the topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.	Always	Subject – Topical	Uses appropriate delimiter. Library of Congress Subject Heading (LCSH). Not geographically subdivided. Decision point needed.
Named entities	Proper names associated with the record.	Always	Subjects – Name	Uses appropriate delimiter. Possibly verify against LCNAF. Decision point needed.
Volume		Conditionally	Volume	Only applies to a publication collection.
Issue		Conditionally	Issue	Only applies to a publication collection.

Field	Field Definition	Using	Mapped to	Notes
Language	A language of the intellectual content of the resource.	Conditionally	Language	Only applies to text-based materials.
Coverage	The extent or scope of the content of the resource. Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range), or jurisdiction (such as a named administrative entity). Concatenated field containing values: style, culture, period, and location.	Conditionally	Geographic Location	JSTOR has split the field for newly ingested items. We do not use any of these fields. Fine art terms are not applicable to our purposes. We use location, but this is not always helpfully applied as we follow TGN. May consider using that field dependent on collection.
Publisher	The entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service.	Conditionally	Publisher	Only applies to published materials, which does not fit the profile for the majority of our items.
Resource type	An internal controlled vocabulary that specifies the general type of the original resource.	No		We have our own controlled vocabulary that we will use.
Format	The physical or digital manifestation of the resource. Concatenated field containing values: medium, technique, material, and measurement.	No		JSTOR has split the field for newly ingested items. We do not use any of these fields. Fine art terms are not applicable to our purposes.
Keywords		No		

The two specialists—our project staff—review, edit, and enhance five key metadata fields: titles, Library of Congress Subject Headings (LCSH), descriptions, named entities, and creators. The resource description specialist focuses primarily on controlled vocabulary elements, while the acquisitions specialist reviews free-text fields, such as titles and descriptions, based on their respective expertise and experience working with special collections metadata.

After staff revise these elements according to local standards, the director of technical services enhances and formats the records, adds missing fields, and performs additional cleanup to ensure consistency with established practices. This work includes formatting dates and reconciling AI-generated genre terms with the library's preferred controlled vocabularies.[13] The director of technical services then exports the records from Seeklight, formats them for ingest, and adds locally required fields, including a note, such as "This asset is described with AI-generated metadata and may not be 100% accurate. Please see our AI Statement for more information," along with a local AI statement.[14] These additions provide transparency by informing users that the digital asset includes AI-generated metadata that may not be fully accurate and by directing them to the library's Policy on the Use of AI-Generated Descriptive Metadata.[15]

Discussion

Project Assessment

As described in the earlier project sections, the Seeklight pilot was designed to test the viability of AI-assisted metadata generation across several different archival formats and descriptive contexts. At the time of writing, Technical Services staff had completed a review of two of the five identified pilot collections, and the head archivist had conducted an initial review of the remaining three collections to determine whether further metadata review using the Seeklight tool is viable.

It is important to note that the Seeklight platform remains under active development. Additional functionality has been introduced since the launch of our initial assessments and pilot projects. Although our findings remain relevant for understanding how the tool performs in current archival workflows, features that may have influenced some outcomes are discussed later in the Changes to Seeklight Functionality section.

Martin Hupka Images[16]

Project staff completed their review of the Martin Hupka image collection, focusing on the title, description, subject headings, and named entities fields. This collection was selected for testing because it had been partially processed prior to our participation in the Seeklight pilot. As a result, project staff had developed substantial familiarity with the subject matter. Additionally, the donor provided item-level descriptions for the majority of the images, including place names, named individuals, and activities depicted in the photographs, which reduced uncertainty during assessment and increased staff confidence in reviewing the generated metadata.

At the time of uploading the collection to Seeklight, we were unsure of how to capture existing item-level descriptions within the tool's bulk-applied Context Box. Consequently, reviewers spent considerable time applying their prior knowledge of the collection to metadata remediation and verifying Seeklight-generated metadata against the donor-provided descriptions. Using a specific image from the Hupka Collection as an example (see figure 1), the donor had identified the names of the individual Peace Corps Volunteers depicted in an image, which Seeklight would not know unless prompted. The original Seeklight-generated title was "Three men eating watermelon on a bench," but our project staff was able to input the names of all three individuals, as well as indicate the three were Peace Corps Volunteers. The title updated by our staff was "Leon, John, and Martin Hupka eating watermelons on a bench."[17]

Figure 1. Example of a digital image that required human revision of Seeklight-generated title.

We also noticed that keywords applied in bulk through the Context Box frequently appeared as subject headings, regardless of their relevance to individual images. For example, the following description was applied to all 444 images: "This collection features photographs taken by Martin Hupka during his Peace Corps service in Colombia, working in community development from 1963 to 1965." For figure 1, the generated subject headings were "young men," "watermelons," "outdoor life," "nineteen sixties," and "community development." Project staff subsequently removed extraneous and irrelevant subject headings during metadata cleanup.

Within this collection, the tool performed most successfully in generating titles. The remaining fields required remediation in the majority of records, largely because project staff were able to incorporate more detailed contextual information than Seeklight could generate from the limited context available to the system. Additionally, Seeklight failed to generate subject headings for a substantial number of records due to a system glitch during the ingest of this collection. The platform does not allow metadata generation without deleting and reingesting a batch, which we opted not to do. The lack of subject headings in these records complicated the evaluation of field quality. Based on our experience reviewing Seeklight-generated subject metadata across collections, however, we would have expected a higher degree of remediation in the subject field.

Patrick Frazier Social and Political Movements Oversize Posters[18]

Project staff also completed a full review of the Patrick Frazier poster collection, focusing on the title, description, subject headings, named entities, and creator fields. This collection was selected for testing because the posters represent a visual medium, many of which include textual elements that provide Seeklight with additional context for metadata generation. This format proved to be particularly successful for generating titles and descriptions. Most remediation work focused on fields that required validation against controlled vocabularies or authority files that are not incorporated in Seeklight.

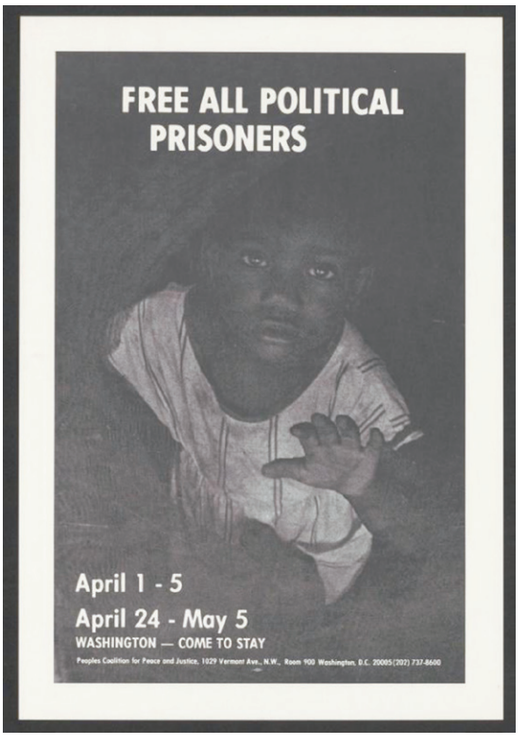


Figure 2. Additional example of a digital image that required human revision of Seeklight-generated title.

Results indicate that the title field performed especially well, with 174 of 180 items acceptable without revision. Titles that required revision were generally not incorrect but required stylistic or structural adjustments. For example, a Seeklight-generated title, "Poster advocating for the release of political prisoners," was revised by project staff to "Free All Political Prisoners protest poster" (figure 2).[19]

The description field showed slightly greater variability in quality but was still largely acceptable without the need for

revision. A total of 152 of 180 descriptions were acceptable without edits. Seeklight-generated descriptions typically captured each poster's visual composition and artistic qualities, alongside relevant contextual information. For example, descriptions often addressed layout, imagery, and stylistic elements, while incorporating contextual interpretation informed by the textual content present on the poster.

Figure 3. Example of a digital image that required human revision of Seeklight-generated description.

Descriptions that required revision generally involved incorrect interpretation of the image context or omission or misidentification of individuals depicted. For example, Seeklight generated the following description for figure 3: "Poster featuring a caricature illustration of three men in military and political attire plotting over a map with missile markers. The text reads 'STOP THEM! they can't stop themselves' and advertises a March on Washington and San Francisco organized by the National Peace Action Coalition, scheduled for April 24, 1971. The artist's signature, 'Edward Sorel '71,' appears at the top right." Project staff revised this description as follows: "A protest poster featuring a political cartoon by Edward Sorel at the top, depicting three men (Richard Nixon as Napoleon, along with Hubert Humphrey and Henry Kissinger) in military and political attire plotting over a map with missile markers. The bold red and black text below reads: 'STOP THEM! they can't stop themselves.' The poster advertises a march on Washington and San Francisco on April 24, 1971, organized by the National Peace Action Coalition."[20]

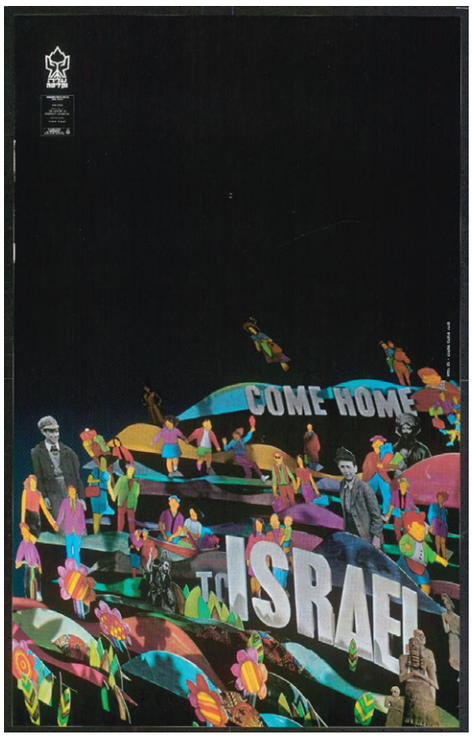


Figure 4. Additional example of a digital image that required human revision of Seeklight-generated description.

Revision was also required when Seeklight-generated descriptions overemphasized visual elements at the expense of contextual meaning. For example, Seeklight generated the following description for figure 4: "A color poster featuring the slogan 'Come Home to Israel' in large, bold letters across a stylized landscape. The lower portion of the poster is filled with illustrated and photographic cut-outs of people, flowers, and abstract shapes, suggesting a diverse and welcoming community. The upper left corner displays a logo and Hebrew text. The overall design uses collage techniques and vibrant colors." Our project staff edited it as: "A promotional poster featuring a collage of illustrated and photographic figures

walking and interacting on a stylized, colorful landscape. The large, bold text reads 'COME HOME TO ISRAEL.' The upper left corner displays a logo and Hebrew text."[21]

Seeklight showed more limitations in generating subject headings, named entities, and creator information. Only twenty of 180 subject headings were acceptable without revision. Although most headings were usable, they often required modification to conform to LCSH conventions, such as the addition of geographic subdivisions. The tool also struggled with repetitive headings or terms that appeared correct but were misapplied according to LCSH definitions.

Additionally, named entities were generated without reference to an authority file, requiring staff to reconcile them with the Library of Congress Name Authority File (LCNAF). Only five of 180 named entity fields required no remediation. Similarly, many creator fields were left blank, necessitating staff intervention; 107 of 180 records required remediation in this field.

Preliminary Review of Remaining Pilot Collections

Shannon & Luchs Real-Estate Collection[22]

The Shannon & Luchs collection proved to be one of the most complicated cases, alongside the Hupka images, for testing Seeklight. This collection was selected to evaluate how effectively the tool could generate metadata for archival materials digitized and presented at the folder level.

It is important to note that JSTOR did not recommend Seeklight for folder-level description. The tool is primarily designed for item-level and compound-object description. Nevertheless, we decided to test the system with this type of material to better understand its limitations.

The collection consists primarily of real-estate appraisal files documenting neighborhoods in Washington, D.C., many of which have since been demolished. These files are organized alphabetically by client name and contain heterogeneous materials that are difficult to parse into discrete items. Any digitized collection made publicly accessible would likely rely heavily on full-text searching for discovery. Attempting to discern relationships among documents in a folder or reliably separating them into items would be so time and resource intensive that it could result in inaccurate interpretation.

Seeklight-generated metadata appeared plausible at first glance and reflected patterns seen in other collections. However, the tool samples content within folders rather than analyzing the entire folder structure. As a result, the generated metadata lacked the depth and contextual completeness that is required for archival description of these materials. An example of a Seeklight-generated title and description is as follows:

- Title: "Appraisal Files and Correspondence on Commercial Properties, Washington, D.C."
- Description: "A compilation of appraisal reports, correspondence, and supporting documentation related to commercial and residential properties in Washington, D.C., primarily generated by Shannon & Luchs Company and associated real estate professionals. The documents include property appraisals, valuation methodologies, legal correspondence, lease agreements, and

> financial calculations. The materials detail property characteristics, neighborhood context, rental income, trust arrangements, and market conditions. The series documents the appraisal process, communication between appraisers, clients, and legal representatives, and includes sample leases and valuation justifications for estate and tax purposes."

This description largely reflects the contextual information provided in the Context Box rather than an actual analysis of the contents of the 111-page digital object. This collection would be worth deleting and reingesting to test more robust prompting using the Context Box, as discussed later in the Changes to Seeklight Functionality section.

Anne Williams Correspondence

The Anne Williams correspondence collection was selected to test Seeklight's transcription tool, which was not available when we first joined the Charter Member program but was later introduced and improved. Preliminary review indicates that the transcription tool performs very well. It successfully handles typescript, handwriting, a combination of the two, and untraditional formatting on the document. However, several workflow limitations remain. Most notably, Seeklight does not currently support batch download of transcripts. Additionally, our repository handles transcripts on a page-by-page basis, whereas Seeklight generates transcripts per asset. To accommodate this difference, we developed a local process to split transcripts before ingesting them into our digital collections platform.

At the time we used the transcription tool for this collection, transcripts were generated independently from metadata. JSTOR has since released enhancements allowing transcripts to add more nuance to descriptive metadata, potentially improving metadata richness.

Paul Jurmo Images

The Paul Jurmo image collection presents a different testing scenario. The donor digitized the images prior to donation but did not provide item-level contextual information. Only collection-level description was available.

Prior to using Seeklight, this collection would likely have been considered a lower priority for digitization and public access due to the lack of contextual information. However, the pilot offered an opportunity to test the tool's ability to generate baseline descriptions for images with minimal context. The preliminary review is promising: the generated descriptions appear comparable with what a human cataloger might produce without insider knowledge of the collection.

Changes to Seeklight Functionality

As noted earlier, JSTOR Seeklight remains under active development. JSTOR has made improvements based on feedback from Charter Member institutions, as well as JSTOR's internal development roadmap. Several limitations we encountered during the pilot have since been addressed, and our understanding of the tool has evolved in ways that would inform how we configure and ingest collections in future projects.

Our experimentation with the Context Box has produced promising results, and JSTOR demonstrations have indicated that it may be possible to point the tool to external data dictionaries and potentially parse larger mixed-manuscript PDFs. Incidentally, this capability would have been particularly valuable for the Shannon & Luchs collection.

We have also experimented with using the Context Box to better leverage existing item-level descriptions during batch image uploads. When images include individual descriptions, we instruct the Context Box to match file names to a corresponding list of descriptions copied directly from an Excel spreadsheet. The results have been extremely promising, with generated metadata being more specific and accurate, and requiring significantly less staff time for editing. However, this approach is constrained by the Context Box's character limit, which restricts uploads to approximately fifteen to twenty images at a time when including existing metadata. Despite this limitation, the approach would have been particularly useful in the scenario described earlier in the Martin Hupka Images section.

Seeklight was initially presented as a no-prompting tool, and significant engineering effort is evident in its design to remain simple and produce reliable results. The Context Box was intended to capture mostly a narrative description from sources such as finding aids or other existing descriptive records and apply across a batch of uploaded files.

JSTOR is now exploring the use of the Context Box as a one-shot prompt tool. This experimental approach offers notable benefits for our project, and we have started uploading collections with simple prompts. It is important to note that Seeklight does not allow for prompt refinement and iteration. Users would have to analyze the metadata results and reupload the collection with a different prompt if refinement were desired. Additionally, Seeklight does not store prompts, nor does it learn from them.

Further improvements to Seeklight have expanded batch metadata editing functionality. Although the Seeklight interface is simple and designed for large batch ingest—allowing users to upload thousands of images per batch, trigger transcript generation by checking a single box, and input metadata and prompts into the Context Box—the back end initially offered limited support for batch metadata editing. Recent enhancements now allow users to create image sets and perform bulk changes by using metadata field filters. For example, format fields can now be batch-replaced with preferred entries from external data dictionaries. These features will likely be used in future projects.

Workflow Adaptation and Human–AI Collaboration

Overall results from the pilot are encouraging. Across collections, the title and description fields proved to be the most reliable outputs generated by Seeklight. Our experimentation suggests that the tool performs best with:

- visual media containing textual context
- textual documents uploaded at the item level
- low- to no-context images.

The tool was less effective for folder-level materials and image collections where existing item-level description could not be easily incorporated into Seeklight.

Our goal during our Charter Membership term is to identify which metadata fields can be reliably generated with minimal manual editing. One area of ongoing evaluation involves the remediation required for LCSH. Our digital collections site currently contains approximately 60,000 assets using LCSH, making interoperability between manually created and AI-generated records an important consideration.

Internally, we consider Seeklight as a workflow enhancement rather than a workflow replacement. The tool integrates into our existing local practices of metadata creation, rather than creating a separate AI-driven process. Adaption of Seeklight has not changed our descriptive standards or workflow sequence. Instead, the project has reinforced that human oversight remains essential for quality control, contextual accuracy, and maintenance of metadata interoperability across collections.

Leadership, Change Management, and Organizational Readiness

The Seeklight pilot has also reinforced broader strategic priorities within the library. Library leadership was not only the impetus for adapting Seeklight, but two members of the Library Leadership Team also designed the framework and workflow used for our launch. The use of Seeklight was framed as a strategic response to the imbalance between digitization output and metadata creation capacity, rather than simply a technology experiment. Project leadership, the director of technical services and the head archivist, intentionally positioned the project as a learning and exploration initiative, emphasizing that AI would augment rather than replace professional expertise.

This framing proved essential for effective organizational change management. Project staff responsible for manual metadata creation were invited to evaluate the tool critically and provide feedback based on their professional expertise. This approach strengthened the project by creating space for staff to experiment with the new tool, evaluate the efficiency of the workflow, and identify both strengths and limitations of AI-generated metadata, while maintaining professional standards and reducing anxiety about AI adoption.

Leadership's willingness to listen to staff and trust in their expertise created a positive environment for experimentation and collaborative problem-solving. Early feedback from staff suggests that working with Seeklight increased comfort with AI tools and encouraged openness to new approaches, highlighting the role of pilot projects in building organizational readiness. Treating the pilot as an iterative learning process allowed the organization to move gradually toward informed decisions about AI-assisted workflows and future technology adoption.

Community Engagement and Collaborative Development

Participation as a Seeklight Charter Member has allowed the library to actively contribute to the development of the platform through monthly meetings and working groups. This collaborative

approach reflects an emerging model in which libraries participate in shaping AI technologies rather than simply adopting them. JSTOR has significantly developed the project since we joined the program in July 2025 and actively seeks input from participating Charter Members.

To date, we have participated in several focus groups centered around Personally Identifiable Information (PII) detection, the transcription tool, folder-level description, and Context Box prompting. Whether any of these ideas will develop further remains uncertain, but the community of practice has shown a strong commitment to both maximizing the potential of the tool in its current form and contributing to its continued development.

Organizational Constraints, Staffing, and Innovation

Our use of Seeklight came at a time of institutional instability, including vacancies in two key positions, metadata coordinator and digitization specialist, which required staff to redistribute responsibilities and cross-train across departments. Instead of delaying experimentation, the team designed workflows that allowed the project to proceed while maintaining oversight and quality control.

Because we positioned Seeklight as a workflow enhancement rather than a replacement, we were able to integrate the tool into our existing infrastructure and training and continue moving collections through production. Most of the work involved in adapting the tool centered on establishing our usage philosophy and developing assessment strategies. Learning to use the Seeklight upload interface and editing metadata within the platform required relatively little effort from staff and virtually no retraining because the process relied primarily on existing expertise. Early observations indicate that AI-assisted workflows increase descriptive productivity. To date, the use of Seeklight has increased metadata production by approximately three times its previous rate.

This project demonstrates that innovation can continue even during periods of staffing and budget constraints, particularly when supported by flexible workflows and collaborative leadership. Investments in exploratory initiatives such as Seeklight can generate organizational learning, cross-department collaboration, and new professional skills, extending benefits beyond the immediate technological outcomes of the project.

Conclusion

The Seeklight pilot project demonstrates how academic libraries can responsibly explore AI as part of broader efforts to expand access to archival and special collections. As many institutions experience the growing imbalance between digitization output and descriptive capacity, AI-assisted metadata generation offers a promising approach for accelerating baseline description and improving discoverability. Our experience with Seeklight suggests that, when thoughtfully integrated into existing workflows, AI tools can function as meaningful workflow enhancements that support, rather than replace, professional expertise.

Several key insights emerged from this case study. First, the pilot confirmed that AI-assisted metadata creation is most effective when implemented as a hybrid human–AI workflow. Across the collections tested, Seeklight generated useful baseline titles and descriptions, particularly for visual materials containing textual elements and for collections with limited contextual information. Authority control, contextual interpretation, and metadata standardization, however, continued to require professional intervention. These findings reinforce the importance of maintaining human oversight and professional judgment within AI-assisted descriptive practices.

Second, the project highlighted the critical role of leadership and change management in the successful adoption of emerging technologies. Leadership framing was essential in shaping how the project was received by staff. By positioning Seeklight as a pilot and learning initiative rather than a mandated technological shift, library leadership created space for experimentation while maintaining confidence in existing professional standards. This approach helped energize project participants—including catalogers, archivists, and managers—while also reducing anxiety about AI among team members. The project further demonstrated that meaningful technological change benefits from actively involving staff whose expertise and daily work are most directly affected. Looping human expertise into the evaluation and refinement of AI-generated metadata proved essential both for maintaining metadata quality and for fostering organizational engagement with the technology.

Third, participation in the Seeklight Charter Member program underscored the value of collaborative development and professional community engagement. By contributing feedback through focus groups and working groups, participating institutions helped shape the evolving capabilities of the tool. This model reflects a broader shift in which libraries play an active role in guiding the development of emerging technologies rather than simply adopting them after implementation.

The project also demonstrated that innovation can continue under conditions of organizational constraint, including staffing vacancies and resource limitations. Because Seeklight could be incorporated into existing workflows with relatively little technical overhead, the pilot proceeded despite institutional instability. In doing so, the project fostered new forms of collaboration between Archives and Technical Services and created opportunities for staff to develop familiarity with emerging AI tools.

More broadly, the Seeklight pilot suggests that the integration of AI into archival workflows should be understood as an organizational and professional transformation process, rather than solely as a technical implementation. Effective adoption depends on leadership support, collaborative experimentation, and the continued importance of human expertise in interpreting and validating machine-generated outputs. As AI tools continue to evolve, libraries will need to develop governance frameworks, evaluation practices, and professional competencies that allow them to incorporate these technologies responsibly.

Future research is needed to better understand the long-term impact of AI-assisted metadata generation on archival discovery, metadata quality, and researcher use of digital collections.

Additional work should also examine how AI tools perform across a wider range of archival formats and descriptive contexts, particularly for complex or hierarchical materials that remain difficult for automated systems to interpret.

Ultimately, the Seeklight pilot illustrates that responsible experimentation with AI can help academic libraries address long-standing barriers to archival access while reinforcing the professional values that guide digital stewardship. When approached as a collaborative, human-centered process, AI-assisted description has the potential to expand discovery, support scalable metadata creation, and enable libraries to make more of their unique collections visible to the communities they serve.

Notes

1. Elizabeth Yakel, "Hidden Collections in Archives and Libraries," *OCLC Systems & Services: International Digital Library Perspectives* 21, no. 2 (2005): 95–9, https://doi.org/10.1108/10650750510598675.
2. The Digital Stewardship Lifecycle model is composed of five phases: Prepare, Gather, Enhance, Save, and Share (https://www.webjunction.org/news/webjunction/Explore-digital-stewardship-lifecycle.html). The Digital Curation Cycle outlines nine key phases, including Conceptualize, Create or Receive, Appraise and Select, Ingest, Preservation Action, Store, Access/Use/Reuse, Transformation, and Dispose/Destroy (https://www.dcc.ac.uk/guidance/curation-lifecycle-model). Discovery and access is an important phase of managing digital resources.
3. The Digital Preservation Coalition, "Digital Preservation Handbook," accessed January 15, 2026, https://www.dpconline.org/docs/digital-preservation/handbook/superseded/299-digital-preservation-handbook/file.
4. Mary Burns, "Hidden Collections and Rare Boks Cataloging: A Review of the Literature," *Library Resources & Technical* Services, 68, no. 3 (2024), https://doi.org/10.5860/lrts.68n3.7863.
5. Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (Fall/Winter 2005): 208–63, https://www2.archivists.org/sites/all/files/MPLP-AmericanArchivist-2005.pdf.
6. Andrew Cox, "How Artificial Intelligence Might Change Academic Library Work: Applying the Competencies Literature and the Theory of the Professions," *Journal of the Association for Information Science and Technology*, 74, no. 3 (2023): 367–80, https://doi.org/10.1002/asi.24635; Abdulhalik Pinar and Andrew Cox, "An Analysis of Artificial Intelligence (AI) Capability in Libraries and Archives," *Cataloging & Classification Quarterly* 63, no. 6–7 (2025): 566–99, https://doi.org/10.1080/01639374.2025.2539790.
7. Anna Gibson Hollow, Lindsay Cline, Mohsen Kardar, Olesya Komarnytska, and Laurel Warkentin, "The AI-Powered Archivist: Harnessing Generative Artificial Intelligence for Streamlined Archival Description," *Journal of Digital Media Management* 14, no. 1 (2025): 6–20, https://doi.org/10.69554/ZMQC4019; Ellen Prokop, X. Y. Han, Vardan Papyan, David L. Donoho, C. Richard Johnson Jr., "AI and the Digitized Photoarchive: Promoting Access and Discoverability," *Art Documentation: Journal of the Art Libraries Society of North America* 44, no. 1 (2021), https://www.journals.uchicago.edu/doi/abs/10.1086/714604?journalCode=adx.
8. Suzhen Chen and Mingyan Li, "AI for Cataloging and Metadata Creation: Perspectives and Future Opportunities from Cataloging and Metadata Professionals," *Technical Services Quarterly* 41, no. 4 (2024),

317–32, https://doi.org/10.1080/07317131.2024.2394919; Edward M. Corrado, "Artificial Intelligence: The Possibilities for Metadata Creation," *Technical Services Quarterly* 38, no. 4 (2021): 395–405, https://doi.org/10.1080/07317131.2021.1973797.

9. JSTOR, "Accelerate Impact with Purpose-Built AI for Stewardship," https://about.jstor.org/products/digital-stewardship/seeklight/.
10. American University approved and published the University Guidance on the Responsible Use of Artificial Intelligence (AI), developed by a cross-campus working group. The guidance outlines principles and expectations for the responsible use of AI in teaching, research, AI literacy, data usage, operational activities—including areas such as procurement—and applies to faculty, staff, and students.
11. American University Library Archives & Special Collections: https://digitalcollections.american.edu/.
12. Dublin Core: https://www.dublincore.org/specifications/dublin-core/.
13. The library primarily uses AAT, TGN, LCSH, and LCNAF controlled vocabularies in our Digital Asset Management system.
14. An example of a metadata record from the Martin Hupka collection that shows our two AI-generated metadata control fields. The first is an internal controlled vocabulary that is meant to tell both the user and internal staff what type of AI-generated metadata is present. The second field is an accuracy warning and will ask the user to read our policy for more information on our philosophy and process.
15. "American University Library Policy on the Use of AI-Generated Descriptive Metadata," American University Library Archives & Special Collections, August 2025, https://digitalcollections.american.edu/use-of-ai-generated-descriptive-metadata.
16. Martin A. Hupka served as a Peace Corps Volunteer in Colombia between 1963 and 1965 working in Community Development. This archival collection includes Hupka's personal papers, including correspondence, sketches, photographs, publications, and Peace Corps training materials.
17. The full description page for this item: https://digitalcollections.american.edu/Documents/Detail/leon-john-and-martin-hupka-eating-watermelons-on-a-bench/217223.
18. Patrick Frazier Political and Social Movements Collection contains digitized broadsides, flyers, handbills, photographs, and posters and covers all of the major political and social movements of the 1960s and 1970s, including civil rights, gay and lesbian rights, and feminism, as well the antinuclear and the Vietnam War protests.
19. The full description page for this item: https://digitalcollections.american.edu/Documents/Detail/free-all-political-prisoners-protest-poster/242527.
20. The full description page for this item: https://digitalcollections.american.edu/Documents/Detail/stop-them-they-cant-stop-themselves-protest-poster/242504.
21. The full description page for this item: https://digitalcollections.american.edu/Documents/Detail/come-home-to-israel-poster/242425.
22. Shannon and Luchs archive finding aid: https://www.american.edu/library/archives/finding_aids/shannonandluchs_fa.cfm.

FEATURE

A Comparative Study of the Implementation Management, and Evaluation of Evidence-Based Acquisition Ebook Plans in Academic and Research Libraries

Christopher A. Palazzolo, William Jacobs, and Matthew J. Jabaily

Ebook evidence-based acquisition (EBA) models have emerged as an important strategy for providing broad access to ebook content while preserving librarian autonomy and agency. As opposed to demand-driven acquisition programs, which have received significant study within the collection management literature, EBA programs are less researched. A survey of academic and research librarians' experiences with EBA was conducted to explore adoption decisions, motivations, management, selection, and overall impressions of EBA programs. Results suggest that EBA programs are becoming more popular and that motivations for their adoption are varied. The results underscore notable challenges and preferences in terms of budgeting, management, and title and program selection. Further work is suggested to deepen the profession's understanding of how EBA programs fit within multiple ebook acquisition approaches.

Introduction

Ebooks have become increasingly common in academic libraries over the last two decades, with the uptake accelerating because of COVID lockdowns.[1] During this time, new models of ebook monograph acquisition—made possible by the new format—came into existence, including demand-driven acquisition (DDA) and evidence-based acquisition (EBA) alongside more established and "traditional" models, such as packages and firm orders.[2] DDA programs have reached a level of maturity, while at the same time evolving considering changes in the marketplace and publisher offerings (e.g., changes in short-term loan allowances). The DDA model is well studied, but EBA is a newer collection method, so the literature regarding it is limited.[3]

EBA models, sometimes referred to as evidence-based selection or user-based collection management models, are monograph-purchasing models in which a library licenses access to a large corpus of ebooks, usually from a single publisher, for a set period, typically a year. At the end of the term, the library selects titles for which to acquire permanent access, using the funds pledged in advance. EBA has emerged as a compromise between pure demand-driven approaches to ebook acquisitions and individual selection.[4] These models allow for some cost containment while also providing a wide range of content and flexibility in selection, along with more publisher revenue stability. These advantages may be coupled with potential negative aspects, including low return on investment (ROI), time- and

Christopher A. Palazzolo (cpalazz@emory.edu) is Director of Collections and Open Strategies, Emory University, https://orcid.org/0000-0003-2707-0685. **William Jacobs** (wjacobs@sfsu.edu) is Associate Librarian for STEM, San Francisco State University, https://orcid.org/0000-0002-0377-7275. **Matthew J. Jabaily** (mjabaily2024@coloradocollege.edu) is Associate College Librarian, Technical Services, Colorado College, https://orcid.org/0000-0001-8069-0555.

resource-consuming management of multiple publisher plans, and content limitations on the part of publishers that exclude most popular titles.

The National Information Standards Organization (NISO), in partnership with vendors, publishers, and librarians, has developed best practices relating to DDA of monographs.[5] The standard addresses a wide array of germane issues, including the management of discovery or candidate records at both the institutional and consortial levels, service implementation (to more transaction-based services for vendors, including the introduction of new features such as usage purchase triggers), and better measurement of usage. The NISO standard makes some reference to EBA, as a variant or alternative of the DDA model, but there were few EBA implementations, aside from some experiments, at the time of the NISO publication (e.g., University of Denver, Brigham-Young and the Orbis Cascade Alliance are notable examples in the 2014–2015 period). Our findings suggest that a set of more widely shared best practices or standards for EBA would benefit the library community, particularly standardized usage data calculation and reporting by vendors.

Literature Review

Early experiments with EBA frequently refer to this model in different terms, for example, as a sunk cost model, rather than specifically as "EBA."[6] Although the term was used earlier to denote general data-driven collections strategies for collection development,[7] the first mentions of EBA as we know it occurred at conferences in 2014,[8] journals in 2015,[9] and collection development references and guides in 2016.[10] Almost all of the literature since has consisted of case studies primarily focusing on the technical issues of implementing EBA or overall performance of the plan.[11]

The focus of this study is the issues surrounding selection of EBA programs, management of said programs, identification of individual titles for purchase from those programs, and renewal or cancellation of programs. Although many of these articles offer glimpses of the libraries' methods, only three recent articles report substantially on them.[12]

EBA Program Selection and Strategy

Carrico et al. provide a good overall set of considerations when beginning an EBA plan, including:

- Is the primary goal access, ownership, or a combination of both?
- Does the library's primary goal include a greater ROI for purchased ebooks and/or streaming videos?
- Do the librarians prefer unmediated user-driven acquisition models or some mediation in purchasing decisions?
- Are current ebook subscription packages providing value to users through repeat use, justifying annual expenditures?
- Are librarian-selected individual ebook titles seeing enough usage to justify their purchase?
- Are there specific publishers whose content is highly valued and consistently used?[13]

Although some studies note that they chose their specific EBA programs using general criteria, such as number of titles or affordability, a couple describe using formal rubrics to select their EBA programs. The majority of articles on EBA that comment on how the programs were selected mention a preference for the inclusion of a large number of titles and an affordable price point. Pilgrim and Dolabaille, reporting on their EBA work at the Alma Jordan Library (AJL) at the University of the West Indies, St. Augustine Campus in 2019, describe the use of a more sophisticated rubric:

- Subject scope: coverage of subject areas was important because the St. Augustine Campus offered a wide range of courses. It was suggested that the broadest range of titles would be the best option when selecting an EBA program.
- Quantity and currency: the number of titles and currency of the collection. The collection chosen needed to be current and not include too many dated publications.
- Patron access: the duration of the program during which time patrons would have access to the titles.
- Vendor relationship: the existing relationship with the vendor.[14]

Pilgrim and Dolabaille note that they were aware of a different rubric created by Robbeloth, Ragucci, and DeShazo to choose EBA programs for the Orbis Cascade Alliance of libraries. However, Pilgrim and Dolabaille chose not to use it. That previously published rubric included the following criteria: "diversity of publishers, DRM [Digital Rights Management]-free, ILL [Interlibrary Loan] rights, price and title stability, ADA [American Disability Act] compliance, interface quality, record management, usage report quality, deposit amount, (title) pricing, and total cost."[15]

Title Selection

Title selection was not mentioned or only briefly touched on in almost all of the twenty-eight EBA-related publications reviewed for this article. Most authors reported selecting in an automated fashion, beginning with the most used titles and continuing downward until the commitment is exhausted. However, Jacobs and Hellman document the development of more involved title selection criteria and process rubric for selection from their EBA programs at San Francisco State University.[16] The criteria combine usage, including usage patterns over time, multiple alternate usage counts, cost-per-use (CPU) calculations, and subject selector picks. General collections staff purchased high-use titles preferentially and then provided more sophisticated usage indicators for the selectors to aid in their selection among mid-use titles. Pilgrim and Dolabaille developed a similar process at their library; acquisitions staff create lists of unowned titles sorted by main subject disciplines for selection by the liaison librarian. Pilgrim and Dolabaille state:

> Although there was no evidence in the literature of the liaison librarian's involvement in the selection process, [Jacobs & Hellman, 2023 having not yet seen print], at the AJL, it was important to include these stakeholders because of their thorough knowledge of the subject collections and their core responsibility of collection development.[17]

Before focusing on the EBA program renewal decision process for their consortium, which was the main thrust of their article, Birchhall, Devarenne, and Wegmann briefly describe their title selection process:

> For selecting titles to add to our permanent collection at the end of the EBA period, we collect data on title usage, overlap with print and electronic holdings, and number of Orbis Cascade Alliance libraries holding the title to indicate ILL availability. This data is used on a title level to select items and not used to evaluate the EBA program as whole.[18]

Whether this analysis is done by the subject librarian or collection manager varies by library but is nearly omnipresent.

EBA Program Evaluation

To evaluate EBA programs, Birchhall, Devarenne, and Wegmann focused on comparability not only between their two EBA programs, but between EBA programs and other subscription services. To that end, they analyzed what data they could extract from EBA programs, what they could learn from it, and how it would affect the renewal decision-making. In the end, they produced reports that listed unique titles used, percentage of collection used, usage distribution, and number of titles by usage (under fifty) that they provided to their decision-makers.[19] Jacobs and Hellman[20] considered some of the same metrics, but hit difficulty with the opacity of COUNTER usage reports, particularly because they are affected by the EBA platform and individual vendor decisions on which clicks count. These considerations are echoed in many of the responses we received to our survey, as each library with EBA programs discovers the complexities of making best use of them. However, there is little evidence of research conducted to learn what other libraries are doing, through the literature or informally.

Methods

This study presents the results of a survey of librarians at a wide range of academic libraries. The researchers created and distributed the survey using Qualtrics in the summer of 2024. Survey results and codebook are available via Mendeley.[21] The survey and study were submitted to the institutional review board (IRB) of Emory University, and it was determined that the study was exempt from IRB approval. The researchers distributed the survey via several professional forums and email lists.[22] The survey was open from July 10, 2024, to August 26, 2024. Participants were given the option of including their names and email addresses for possible follow-up interviews, and institution names were collected to ensure that only one response per institution was counted for quantitative results. Participant and institutional identities were not used in the analysis of responses, although relevant characteristics of the institutions, such as student full-time equivalent (FTE), were considered when reviewing results. If more than one response was given from an institution, the numerical responses from the higher-ranking position, typically the department head, were used, but free text responses from all surveys were included in the analysis. Survey results were cleaned in Microsoft Excel to remove uncompleted surveys and to facilitate analysis. Examples of cleanup steps taken include normalizing

spellings, removing extraneous text from responses that asked for numbers, and converting numbers from words to numerals. Calculations and visualizations were made using Excel.

RESULTS

Institutional Characteristics and Ebook Practices

The survey yielded eighty-four completed surveys with results from eighty-one unique institutions. As shown in table 1, participants (n = 81) represented a variety of Carnegie Classification institution types, but the largest number of participants was from those with R1 classifications (44.4 percent). R2 institutions were also strongly represented (27.2 percent), but relatively few participants were from master's-only granting institutions (7.4 percent), baccalaureate-only granting institutions (9.8 percent), or community colleges (1.2 percent). A final 9.9 percent were from outside of the United States and did not fall neatly into these categories.

Table 1. Breakdown of Research Level of Survey Respondents

Classification	Count	Percentage
Community college	1	1.2%
Baccalaureate-only granting institution	8	9.9%
Master's-only granting institution	6	7.4%
R2 doctoral granting institution	22	27.2%
R1 doctoral granting institution	36	44.4%
Other (if outside of United States)	8	9.9%

Figure 1 shows the student FTE distributions. The largest number of participants was from institutions with a student FTE greater than 20,000 (42 percent). Medium-size institutions with a student FTE

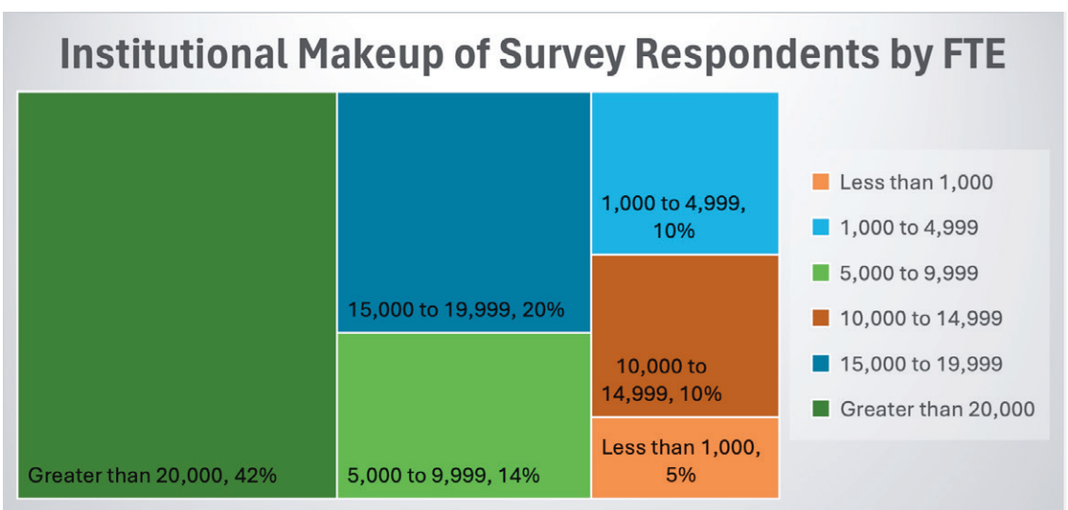


Figure 1. Institutional makeup of survey respondents in terms of FTE.

of 15,000 to 20,000 were also well represented (20 percent), with the remaining smaller institutions. Collection budgets at those institutions varied widely, but the largest portion had collections budgets in the $1 to $5 million range (39.5 percent) or $5 to $10 million range (19.8 percent). Table 2 shows the full collection budget distributions.

Table 2. Breakdown of Overall Collections Budgets of Survey Respondents

Collections Budget	No. of Respondents	Percentage
$1 million or less	10	12.3%
$1 to $5 million	32	39.5%
$5 to 10 million	16	19.8%
$10 to $15 million	11	13.6%
$15 to $20 million	7	8.6%
Greater than $20 million	2	2.5%
Other (non-US currency)	3	3.7%

Although the survey focused primarily on EBA programs, only sixty-nine of the eighty-one participants (85.2 percent) said their institution was currently participating in an EBA. A majority of participants who responded to the question (n = 80) reported having discontinued at least one EBA in the past five years (55.5 percent). Five participants (6.2 percent) reported that their institution did not currently have an EBA but had discontinued at least one in the past five years. Seven participants (8.8 percent) reported their institution had neither participated in nor discontinued an EBA in the past five years. Because our survey was distributed preferentially to librarians managing EBA programs, the actual percentage of libraries without EBA programs is likely significantly higher. Survey participants (n = 81) were asked a series of questions about their ebook collection practices. When asked to estimate the percentage of their collections budget that went toward ebooks, most reported spending less than 20 percent on ebooks (82.7 percent); the most common amount selected was 5 to 10 percent of the collection budget (33.3 percent). Participants were asked to estimate how their ebook spending was split across four common acquisition methods: firm orders, EBA programs, DDA programs, and packages. Practices varied widely; on average, packages claimed the biggest portion of ebook funds, but some institutions spent more on firm orders or EBA programs (see table 3). EBA programs typically claimed between 5 and 35 percent of funds, but some spent much more and, as noted earlier, some did not participate in any EBA programs.

Table 3. Ebook Acquisition Models Reported By Survey Respondents

Ebook Acquisition Model	Minimum/Maximum	Mean	Standard Deviation	Median
Firm orders (n = 81)	0%/100%	23%	22.8	16%
Evidence-based acquisition (n = 81)	20%/80%	22%	19.7	20%
Demand-driven acquisition	0%/75%	7.6%	11.5	2%
Ebook subscriptions/packages (n = 81)	0%/100%	37.6%	29.4	32%
Other	0%/64%	4.2%	11.5	0%

Note: Although not specifically delineated or suggested, "other" might include ebook approvals.

Individual respondents provided estimates of proportions of ebook budgets dedicated to various available acquisition strategies. Surveyed libraries still dedicate substantial percentages of their budgets to firm orders. In terms of the balance between DDA and EBA, EBA seems to be preferred. As one R1 respondent noted, EBA is preferred over DDA from several positions: "(1) predictability of cost, (2) entire ebook frontlists from subject-diverse publishers, (3) no DRM, [and] (4) helping selectors free their time to focus on selections from smaller/more diverse publishers" (R1, >20,000 FTE).

Most respondents currently participating in at least one EBA (n = 66) indicated they participate in more than one (84.8 percent). Figure 2 shows the distribution of number of EBA programs, with the mode (31.8 percent) participating in two EBA programs and some participating in many more. Several participants commented on the number of EBA programs at their institutions, with one participant saying, "We would like to expand our EBA programs in coming years but are limited by budget cuts" (R1, 15,000–20,000 FTE). The commenter who noted a desire to grow these plans was motivated by a "diminishing level of engagement in collection development by subject librarians" with the "modest firm order allocation . . . not getting spent by selectors" (R1, 15,000–20,000 FTE).

Participants at institutions currently enrolled in an EBA program (n = 69) reported many variations of EBA programs, with almost half participating in at least one consortia EBA (46.4 percent) and slightly more than half reporting participating in a multipublisher EBA (50.7 percent). Regarding consortial EBA programs, one participant wrote, "In addition to local plans with locally allocated budget, we benefit from additional EBA programs through a consortium without contributing any local funds" (R1, >20,000 FTE).

Figure 3 displays responses from participants (n = 66) when asked when their institution began its first EBA. The earliest reported was 2010. Most likely, this EBA was an early experiment with an

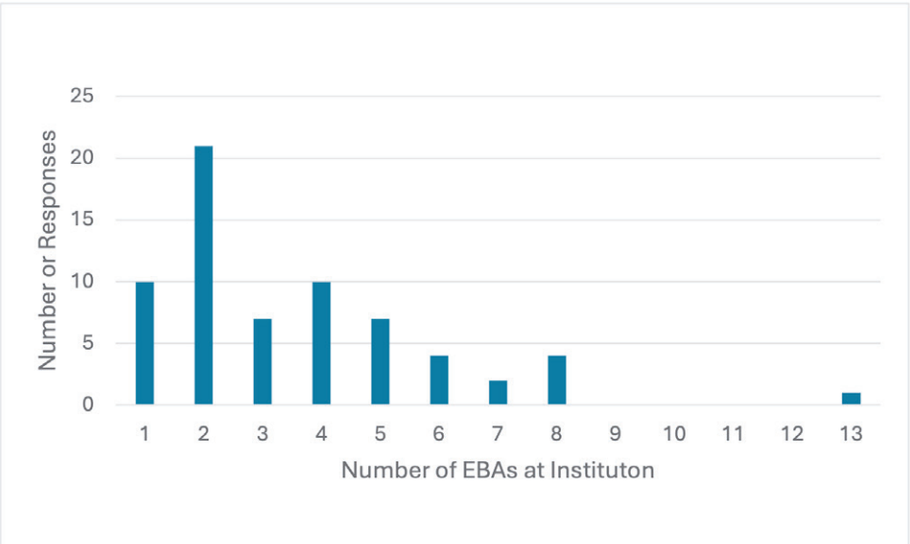


Figure 2. Number of EBA programs currently in use (at time of survey).

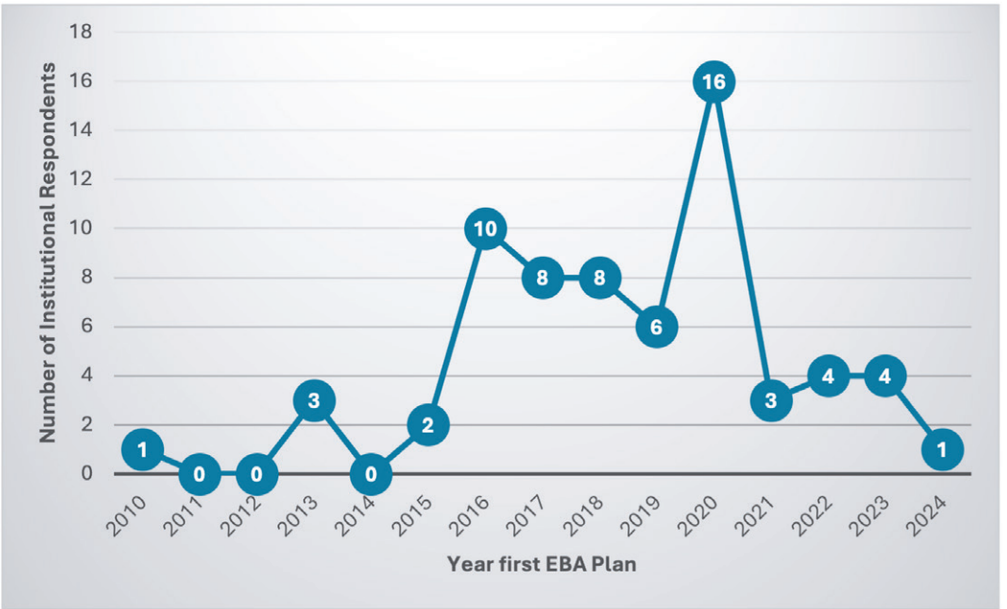


Figure 3. Date of library's first EBA.

upfront deposit and not referred to as an EBA, except in retrospect. There was a bump in 2016, with ten (15.2 percent) reporting starting their first EBA that year. The most cited year of starting a first EBA was 2020 (24.2 percent). This start date could be attributed to the pandemic, although when asked regarding motivations, the "response to closure" scored last. Full details are displayed in figure 3.

Respondents were asked about their motivations for adopting an EBA. Responses (n = 69) are represented in figure 4. A desire to "expand access" was the most common response (84.1 percent), followed by wanting to "align to user needs" (71.0 percent), and "save selection time" (50.7 percent). A desire to "increase eBook use" (23.2 percent) or "diversify collections" (24.6 percent) was reported by fewer participants, although one did comment, "EBA programs are a way to provide a diverse array of books to users to cover subject areas that might not have seen recent attention by librarians" (R1, 15,000–20,000 FTE). Only ten institutions (14.5 percent) reported starting their first EBA in "response to library closures (e.g., COVID)," despite more institutions reporting they started their first EBA in 2020, the first year of COVID-related library closures. Some did comment that COVID closures encouraged them to expand use of EBA programs. One reported, "I was initially skeptical of EBAs, and I admit that the first one we did was experimental, but the pandemic really pushed us into more of these, and I think to the good" (R1, >20,000 FTE).

EBA Plan Evaluation and Motivations

Participants who reported currently participating in an EBA were asked about their motivations for starting their initial EBA, selecting all applicable motivations from a list of choices. Participants were also asked about factors considered when evaluating whether to begin an EBA and rated them on a scale

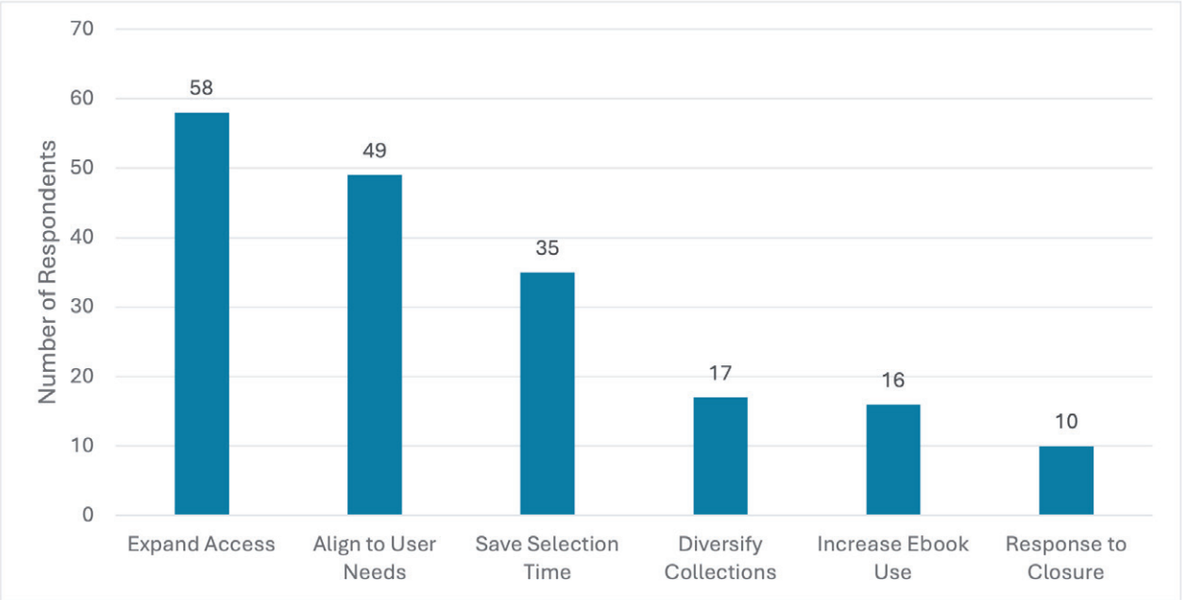


Figure 4. Motivating factors for adopting an EBA.

of 1 (not that important) to 5 (very important). Results are available in figure 5. The most important factors, with mean scores between 4 and 5, were deposit cost and alignment with the library's current collection areas. Other important factors, with average ratings between 3 and 4, included:

- copyright years included
- availability of MARC records
- filling of collection gaps
- completeness of corpus
- ease of selection process
- price of individual ebooks
- access to backfile content

Less important factors for EBA evaluation, with means between 2 and 3, were EBA management through a third-party vendor (nonpublisher direct), customization, and duplication with the print collection. Some participants reported the importance of balancing several factors related to cost, collection development, and management. As one wrote: "It is very important when I consider an EBA whether it is accessible through my knowledge base; how much it costs; and how much content (and copyright years) is included" (baccalaureate only, 1,000–5,000 FTE).

Management criteria were also emphasized by another respondent as a principal factor for evaluation:

> EBAs are most difficult to manage when the publisher does NOT work with our primary book vendor, as it is difficult for our selectors to know/remember what titles are included in the EBA, if the indicators are not in the book selection/ordering system. (R1, 15,000–20,000 FTE)

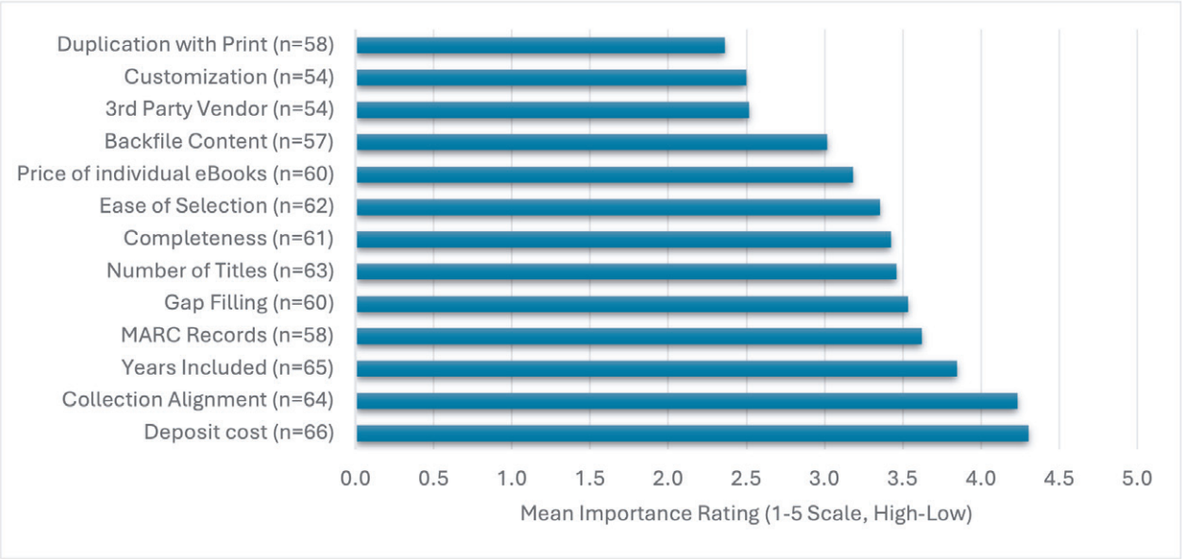


Figure 5. Factors that influence the choice of an EBA (means, 1–5 scale).

EBA Title Selection and Renewal or Discontinuation

Participants were asked about practices for selecting titles for perpetual access at the end of the EBA evaluation period. Participants (n = 68) were asked who was responsible for title selection and were allowed to choose more than one option (table 4). Most reported selection was done by a collections librarian (86.8 percent). A third reported that subject or liaison librarians were involved (35.3 percent). Fewer reported involving other technical service staff (22.1 percent) or selected other (20.6 percent). Only one reported having nobody assigned (1.5 percent).

Participants were asked about what factors they used for selecting titles and were asked to score the importance of criteria on a scale of 1 (not considered at all) to 5 (very important). These results are available in figure 6. Current year usage was the most important factor, with a mean rating of 4.5. Relatively important criteria, with mean ratings between 3 and 4, were whether titles had been used or adopted for courses, whether there was access to the title through another database, whether there was evidence for use over time, whether the titles aligned to collection needs, and how use was distributed throughout the year. The least important factors—with mean scores between 2 and 3—were CPU, subject/liaison librarian feedback, format/type of book, and duplication with print ownership.

When participants (n = 68) were asked what statistics were used to guide decision-making, almost all reported using the vendor-provided reports (95.6 percent). Some calculated CPU (38.2 percent) and fewer pulled their own COUNTER statistics (22.1 percent). Only five reported using another form of locally harvested usage data (7.4 percent). Although most respondents reported use of vendor data such as COUNTER reports in evaluating selections and/or renewals, other metrics such as historical denial (turnaway) data, more granular usage data (e.g., unique title requests, not just total title requests), and longitudinal data (e.g., data over multiple EBA periods or cycles) were identified as potentially

important in overall EBA assessment. Some librarians have developed complex approaches to title selection, but cost-effectiveness and ROI remain paramount factors for overall plan assessment.

Participants who prioritized the speed and ease of the selection focused solely on use. A typical response noted, "We pretty much go by the total usage as reported by the vendor when deciding which ebooks to purchase at the end of the year." Although saving selection time was given as a reason for many libraries to begin an EBA, several comments about title selection noted that the selection process could be difficult. One noted, "I think it was overall a good deal, but the selection process was more time consuming than I expected." Another reported that for one of their EBA programs, "as the years go on, it is harder to find ebooks to purchase in the EBA as high use ones are already owned" (R1, >20,000 FTE).

Table 4. Responsibility for Selecting Titles at End of EBA Period

Role	No. of Respondents*
Collections librarian	59
Subject librarian/liaison librarian	24
Technical service staff	15
Other (multiple groups for review, continuing resources groups, acquisitions, graduate student for collection management, etc.)	14
No one assigned	1
Total	113

*Number of responses is greater than the number of survey respondents (n = 81) due to the ability to choose multiple responses.

Participants were asked about their decisions to renew or cancel their EBA programs (figure 7). They were asked to rate factors in these decisions on a scale of 1 (not that important) to 5 (very important). The two most important factors, with mean scores greater than 4, were the overall CPU for the EBA and whether there was a deposit increase. The proportion of titles used, difficulty in management, and whether there was a change in available content were somewhat important, with mean scores in the 3 to 4 range.

As noted earlier, a majority of participants (n = 44) reported that they had discontinued an EBA in the past. They were asked for their reasons for doing so and were allowed to select more than one option. The most common reason, as shown in figure 8, for cancellation was low overall use (63.6 percent), with budget cuts given as the second most common reason (38.6 percent). Difficulty with management (25.0 percent) and large cost increases (18.2 percent) were selected less frequently. Many participants volunteered other reasons for discontinuing the programs. Several reported changes to programs' contents or costs, but others noted that the program had "run its course" or that they had otherwise selected the titles they wanted to collect and "a second year was not necessary."

DISCUSSION

EBA has become a substantial and legitimate component of ebook acquisition strategy at many institutions because it balances wide access to ebook content while allowing for individual selection and

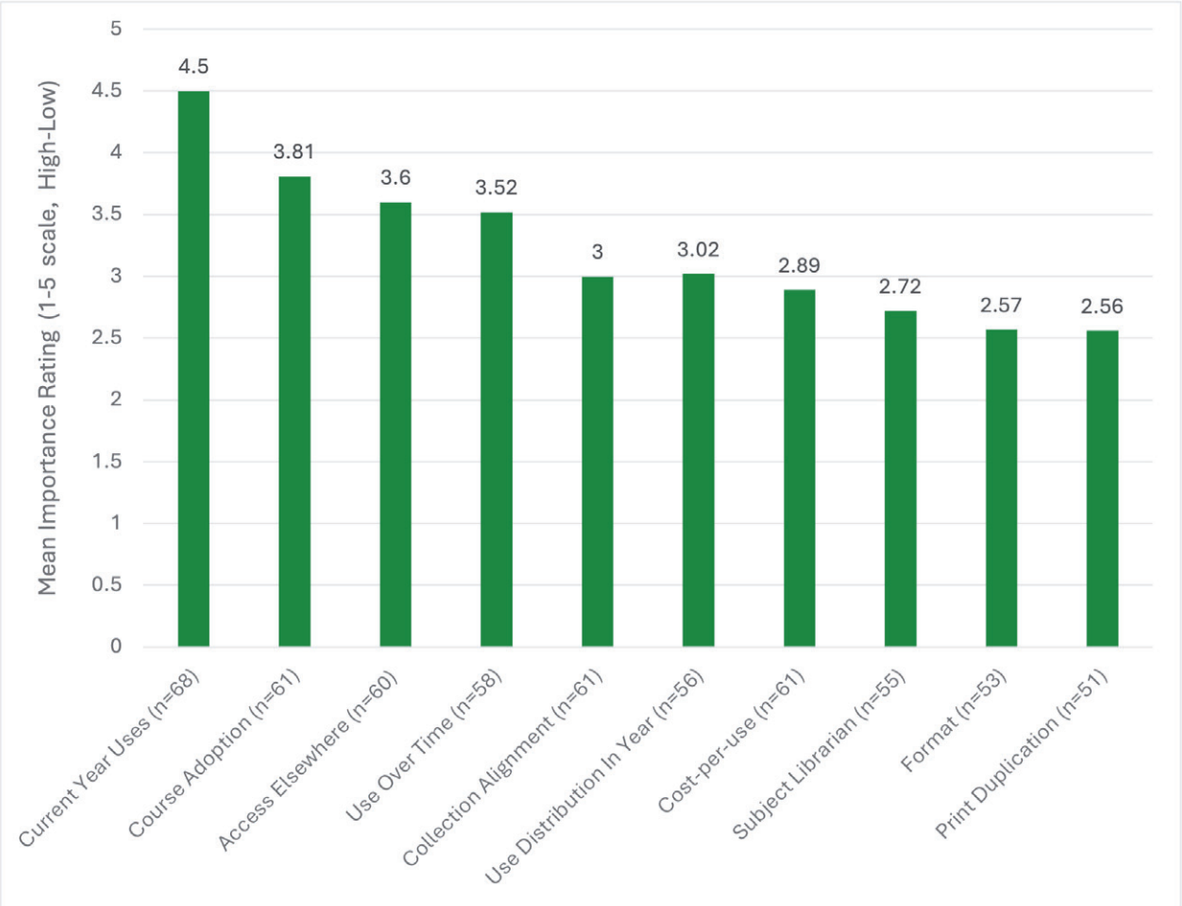


Figure 6. Factors for making end-of-term EBA selections (means, 1–5 scale).

perpetual access.[23] One respondent concluded: "As a large, diverse, research-intensive university, EBAs offer us excellent ROI as well as flexibility and purchasing based on evidence of use" (R1, >20,000 FTE). However, motivations for implementation of EBA programs alongside other models, including packages, firm orders, and demand-driven models, vary.

Although issues of content, duplication, use, access, research and curricular alignment, and data needs for evaluation all factor into adoption and evaluation of EBA programs, perennial budget considerations remain of central importance. As many respondents noted, budget control and budget predictability make EBA programs attractive and allow for greater access at a lower cost. Yet, these plans are managed as recurring subscriptions (frequently with regular market increases) that must be evaluated and assessed for their ROI and balanced against other collection needs. In several cases, EBA programs were abandoned because of being no longer fruitful as all relevant titles were chosen.

An important caveat is that our data is limited to a single point in time. The prevalence of EBA adoption we saw may be an artifact of our distribution methods and respondents. Future studies might inquire

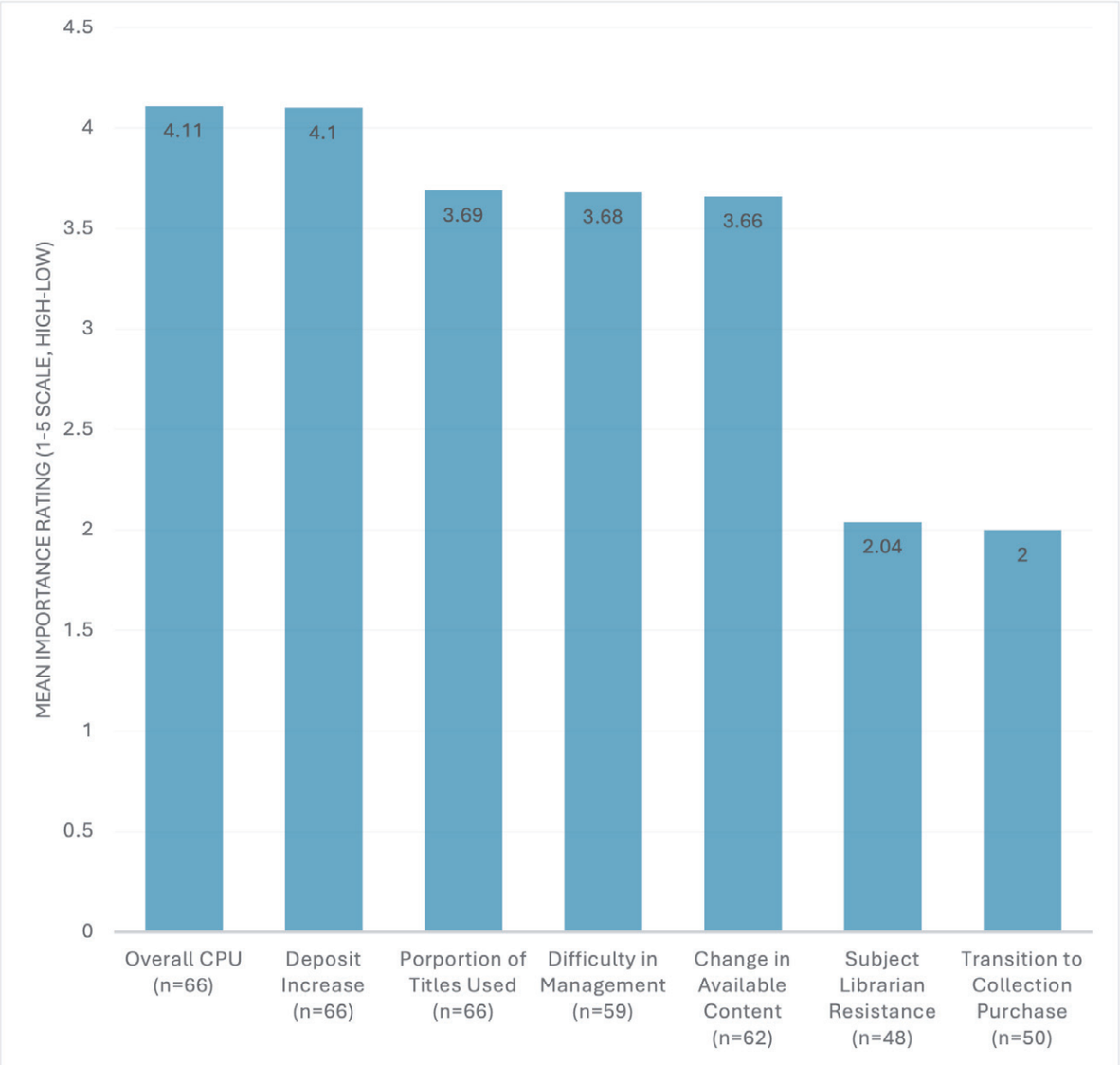


Figure 7. Factors that influence renewal of an EBA (means, 1–5 scale).

whether these proportions have significantly changed over time and expand the survey distribution further. Interestingly, packages continue to be the primary means of access to ebook content. However, the survey did not ask for a distinction between purchased or subscribed. As one respondent commented:

> EBA plans with specific publishers as both an individual institution and a member of a consortium have been successful from several positions: 1. predictability of cost; 2. entire ebook frontlists from large, subject-diverse publishers; 3. no DRM; 4. helping selectors free their time to focus on selections from smaller/more diverse publishers. (R1, >20,000 FTE)

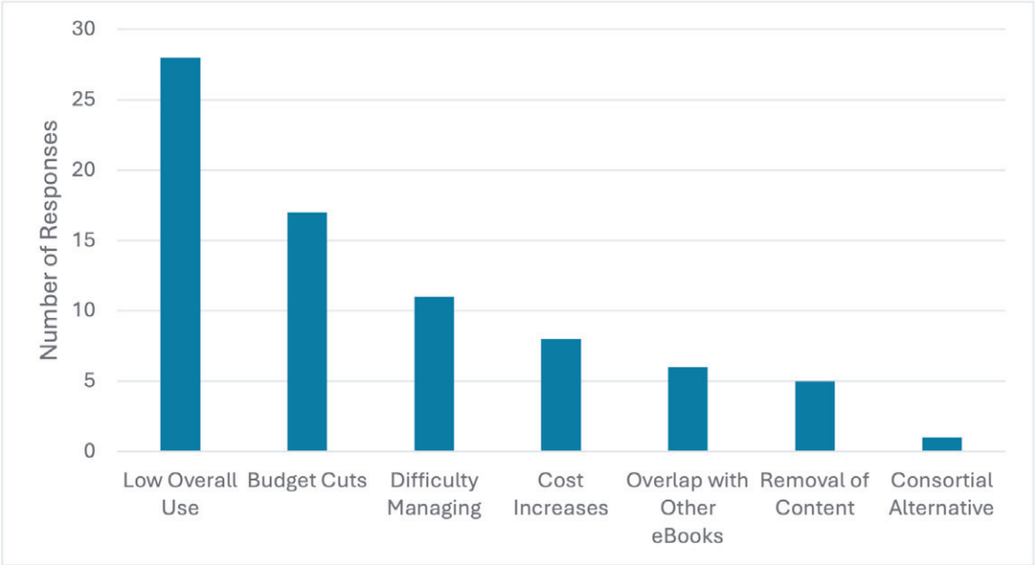


Figure 8. Reasons for discontinuing an EBA.

Another respondent emphasized that DDA programs have run their course:

> We have over the past two years expanded our EBA plans to new vendors and cancelled some due to cost increases and low usage. Overall EBAs are a better form of collection tool, than say DDAs because they provide more control over the spend. I do wish University Presses would be get [sic] more onboard with providing new content through EBAs. (R1, >20,000 FTE)

Although seeming to supplant the dominance of DDA, it appears that individual libraries may struggle with workflows and assessment as part of the EBA life cycle. As one respondent noted:

> Overall, we value just-in-time collecting like EBA, and are trying to do more of it, but due to the staffing challenges and vendor-driven aspects of these particular plans, our approach is not as evidence-based or efficient as it probably could be. (R2, 10,000–15,000 FTE)

Another noted that DDA had been preferred over EBA due to efficiency issues:

> We have generally preferred DDA plans to EBA plans because of the challenges in reviewing title and usage lists, particularly if they are bunched up at the end of the fiscal year. As we've gotten better at it (including streaming media EBA), this has lessened but is still a bandwidth challenge. (R2, 5,000–10,000 FTE)

With many libraries balancing multiple plans and managing multiple data reports and workflows for EBA programs, capacity and infrastructure become central issues (one respondent referred to these considerations as "bandwidth" issues). Management of EBA programs can be challenging and time-consuming (including title selection, duplication, records management). For example, initial EBA setups have multiple workflow decision nodes and multiple stakeholders and decision-makers.

Considerations include budget and cost, content and overlap, process and source of record retrieval (initial and ongoing content loads), and end-of-term purchases.

More specifically, the following emerged as key issues:

- Challenges for staff (e.g., technical services staff) due to variations in individual plans (publication dates included, MARC records delivery, term dates), selection deadlines, etc., all require local best practices and discussion.
- Adjustments to existing programs and assessment of EBA programs (e.g., content changes, ROI, adjustments to deposit) take time to perform.
- Renewal periods could be extended or terms extended to allow for more deliberate and careful selection of perpetual access titles and content adjustments, along with more standardized usage reporting.

Both our literature review and our survey found many libraries bringing subject librarians back into the ebook selection process with EBA after their exclusion from DDA programs. In many other cases, however, central collections staff handle selections.

> As noted, evidence-based acquisitions (EBA) plans work similarly to DDA plans, but the titles are vetted by the subject librarian before determining whether the eBook accessed via short-term loan for the user should become a piece of the permanent collection. If DDA causes loss of identity for subject librarians as "selector," EBA increases the selection value of the subject librarian . . . Subject librarians must be able to discern return on investment to ensure their budget allocation is used judiciously.[24]

Our results suggest that individual libraries and consortia should consider best practices around EBA programs, especially when managing multiple plans. Libraries should outline processes, workflows, and roles and responsibilities of staff across the EBA life cycle.[25] In addition, clear criteria for the selection of titles and defined roles and responsibilities in management and selection would benefit library staff. One such model for end-of-term selection is by Hellman and Jacobs; the authors outline criteria and a decision tree for such decisions, which includes collection managers and subject teams.[26] There could be a substantial opportunity for joint, collective publisher–library initiatives to develop standards, in line with the NISO DDA standards.

Birchhall, Devarenne, and Wegmann developed a set of prioritized data points for the annual comprehensive evaluation of their EBA that goes beyond traditional simple CPU or overlap with print and electronic holdings (both locally and for consortia).[27] They used newer variables such as curricular importance, CPU book (not just CPU for the collection), ILL availability, and consortial availability, and prioritized these against traditional measures so that their two plans could be assessed in concert with each other.[28] In support of this approach, several respondents argued that better technical support and more involved usage reports would certainly be welcome (e.g., more longitudinal data to assess long-term use, better duplication and deletion control, extended terms, fixed increases). Vendors and publishers should take these data points into consideration when providing data for selection.

Conclusions and Future Research

This study focused on uptake of EBA programs by (a subset of) libraries in isolation. Further research—both qualitative and quantitative—at a wider scale of participation may provide further insights and different patterns of EBA adoption and evaluation. In addition, our research did not delve into patterns of EBA adoption by library size and/or research level or classification (although we did receive a good number of responses across size and research level). Future investigations could potentially involve publisher aggregate data on EBA programs and trends to compare against librarian self-reporting of EBA adoption.

Considering the normalization of EBA as an ebook acquisitions strategy, another avenue of possible investigation is the effect of EBA on overall collection development. Usage patterns may be analyzed to investigate patron interests and new areas of concentration on campus. In addition, the increased popularity of EBA may have implications for resource-sharing. With a combination of candidate or access-only records and purchased records, there may be more complexity in fulfilling user requests across libraries. Learning how early adopters of EBA have developed long-term approaches to selection and renewal decisions would also be of interest. For instance, multiple respondents mentioned EBA programs losing value over time as the most used items are selected out. Is there an optimal length of time for subscriptions or length of pause before renewing to a refreshed corpus?

Vendor and publisher perspectives could be an area of future research as well. Best practices for managing EBA programs, reporting usage, and calculating optimal commitments as library budgets shrink have yet to be determined. The long-term profitability of EBA programs is also unknown.

Librarians have had to balance ebook access (and demand) with ownership in the face of constrained budgets. EBA has served as a model to mitigate some of these concerns, but it can be complicated due to the multiple issues of budget, selection, records management, evaluation, and stability. With that said, EBA could potentially be a more complex and sophisticated collection development tool than it currently is. Collaboration among publishers, vendors, and librarians should continue to develop to further facilitate management and effective assessment.

CRediT Author Contribution Statement

Christopher A. Palazzolo: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, and Visualization.

William Jacobs: Conceptualization, Methodology, Investigation, Writing – Original Draft, and Writing – Review & Editing.

Matthew J. Jabaily: Methodology, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, and Visualization.

Notes

1. David Wells and Anita Sallenbach, "Print Books and Ebooks: The New Equilibrium in an Academic Library," *Journal of the Australian Library and Information Association* 72, no. 2 (2023): 166–77, https://doi.org/10.1080/24750158.2023.2183560.
2. Rachel E. Scott and Michael A. Arthur, "Electronic Ebook Acquisition Strategies in Academic Libraries: A Review of the Literature," *portal: Libraries and the Academy* 26, no. 2 (2026): 277–301, https://doi.org/10.1353/pla.2026.a986768.
3. Amalia Monroe-Gulick, Andi Back, Gwen Geiger Wolfe, Sara Outhier, and Sara E. Morris, "Demand Driven Acquisitions in Academic Libraries: A Scoping Review," *The Journal of Academic Librarianship* 50, no. 3 (2024): 102862, https://doi.org/10.1016/j.acalib.2024.102862.
4. Steven Carrico, Michelle Leonard, Erin Gallagher, and Trey Shelton, "Evidence-Based Acquisitions," in *Implementing and Assessing Use-Driven Acquisitions: A Practical Guide for Librarians* (Rowman & Littlefield Publishers, 2016).
5. NISO DDA Working Group, *Demand Driven Acquisition of Monographs: A Recommended Practice of the National Information Standards Organization* (National Information Standards Organization, 2014).
6. Kari Paulson, "The Story of Patron-Driven Acquisition," in *Patron-Driven Acquisitions: History and Best Practices*, ed. David A. Swords (De Gruyter Saur, 2011), 63–78, https://doi.org/10.1515/9783110253030.63.
7. Colleen Cleary and Bruce Heterick, "Getting the Balance Right with EBooks," in *VALA 2018: Libraries/Technology and the Future*, 2018, https://api.semanticscholar.org/CorpusID:108375376; Leila Salisbury, "From a University Press: Opening Pandora's (Cable) Box," *Against the Grain* 26, no. 6 (2014), 33, https://doi.org/10.7771/2380-176X.6968.
8. Michael Levine-Clark, "Evidence-Based Selection at the University of Denver," *Against the Grain* 27, no. 5 (2015), 10, https://doi.org/10.7771/2380-176X.7176.
9. Julia Proctor, "Demand-Driven Acquisition and the Sunk Cost Model," *Collection Building* 34, no. 1 (2015): 2–5, https://doi.org/10.1108/CB-06-2014-0033.
10. Carrico et al., "Evidence-Based Acquisitions."
11. Hillary Robbeloth, Matthew Ragucci, and Kristina DeShazo, "Evidence-Based Acquisition: A Real Life Account of Managing the Program within the Orbis Cascade Alliance," *The Serials Librarian* 73, no. 3–4 (2017): 240–47, https://doi.org/10.1080/0361526X.2017.1388331; Stephanie J. Spratt, Gabrielle Wiersma, Rhonda Glazier, and Denise Pan, "Exploring the Evidence in Evidence-Based Acquisition," *The Serials Librarian* 72, no. 1–4 (2017): 183–9, https://doi.org/10.1080/0361526X.2017.1321901; John Abresch, Laura Pascual, and Andrea Langhurst Eickholt, "EBA in Practice: Facilitating Evidence-Driven Ebook Programs in Both Consortium and Individual Library Settings," in *What's Past Is Prologue: Proceedings of the 2017 Charleston Library Conference*, eds. Beth R. Bernhardt et al. (Purdue University Press, 2018), 116–21, https://doi.org/10.5703/1288284316693; Mariella Pilgrim and Arlene Dolabaille, "Crossing Frontiers: Implementing an EBA Model at a Caribbean Academic Library," *The Serials Librarian* 83, no. 2 (2022): 176–85. https://doi.org/10.1080/0361526X.2022.2121798; Marion Favreand Manon Velasco, "Patron-Driven Acquisition et Evidence-Based Acquisition: Comment implémenter ces modèles afin d'étendre l'offre documentaire dans une bibliothèque," in *Beyond the Library Collections: Proceedings of the 2022 Erasmus Staff Training Week at the University de Liège Library*, eds. François Renaville and Fabienne Prosmans (Université de Liège, 2023).

12. Braith Birchhall, Sidonie Devarenne, and Mary Wegmann, "Evidence Based Assessment of Evidence Based Acquisitions," *Western Libraries Faculty and Staff Publications*, October 2024, https://cedar.wwu.edu/library_facpubs/78/; William Jacobs and David Hellman, "Beyond Usage – Selection Strategies for EBA Programs," *The Journal of Academic Librarianship* 49, no. 3 (2023): 102662, https://doi.org/10.1016/j.acalib.2023.102662; Pilgrim and Dolabaille, "Crossing Frontiers."
13. Carrico et al., "Evidence-Based Acquisition," 78.
14. Pilgrim and Dolabaille, "Crossing Frontiers," 177–8.
15. Robbeloth et al., "Evidence-Based Acquisition," 241.
16. Jacobs and Hellman, "Beyond Usage," 3.
17. Pilgrim and Dolabaille, "Crossing Frontiers."
18. Birchhall et al., "Evidence Based Assessment of Evidence Based Acquisitions," 3.
19. Ibid.
20. Jacobs and Hellman, "Beyond Usage."
21. Christopher A. Palazzolo, William Jacobs, and Matthew Jabaily, "Survey Data from 'A Comparative Study of the Implementation, Management and Evaluation of Evidence-Based Acquisition Ebook Plans in Academic and Research Libraries,'" *Mendeley Data* V1 (2026), https://doi.org/10.17632/2h8j6jpdx2.1.
22. The following listservs were used to distribute the survey: Association of Southeast Research Libraries (ASERL), Collections Group, Chief Collection Development Officers (CCDO) of the Association of Research Libraries (ARL), LibLicense, Electronic Resources in Libraries (ERIL), SERIALST, Colorado Alliance of Research Libraries Collection Development, STEM Librarians Collective—Collections Channel, ACRL Sci-Tech Connect, and Cal State Collections.
23. Rebecca Schroeder and Rebecca Boughan, "Doing More with Less: Adoption of a Comprehensive Ebook Acquisition Strategy to Increase Return on Investment While Containing Costs," *Library Resources & Technical Services* 62, no. 1 (2018): 28, https://doi.org/10.5860/lrts.62n1.28.
24. Jennifer Farthing, *Exploring the Subject Librarian Role in eBook Purchase Processes: A Quality Improvement Study* (Master's Thesis, Vanderbilt University, 2021), 59–60, https://irbe.library.vanderbilt.edu/server/api/core/bitstreams/328d7e52-f736-4aa2-9d6a-9f869070a82b/content.
25. Kat Hart, Chris Palazzolo, and Sofia Slutskaya, "Cross-departmental collaboration for successful DDA and EBA management, *Journal of Electronic Resources Librarianship* 34, no. 4 (2022): 336–41, https://doi.org/10.1080/1941126X.2022.2133924; Sofia Slutskaya and Alexis Linoski, "Ebooks: Access vs. Ownership," *The Serials Librarian* 78, no. 1–4 (2021): 79–85, https://doi.org/10.1080/0361526X.2020.1716927.
26. Hellman and Jacobs, "Beyond Usage," 4.
27. Birchhall et al., "Evidence Based Assessment of Evidence Based Acquisitions," 2–3.
28. Ibid.

FEATURE

"Unbury" the Book Chapter?

An Investigation of Library and Information Science Book Chapters from Edited Collections in Scholarly Indexes

Juleah Swanson

Publishing a book chapter in an edited collection carries a perception that doing so is akin to "burying" one's work. This research investigates whether book chapters from edited collections within the field of library and information science suffer from discovery challenges in three indexing databases: Google Scholar, Web of Science Book Citation Index (BKCI), and Dimensions. Findings show a consistent lack of comprehensive indexing across all three indices. This research also sought to answer questions about open access and book chapters; due to the lack of discoverability of half of the book chapters, however, these questions could not be addressed. The results highlight concerns about the discoverability of book chapters within edited collections, particularly in the field of library and information science, while also examining potential avenues to increase discovery.

Introduction

A persistent metaphor of "burying one's work" in relation to authoring a book chapter exists. This metaphor stems from an informal investigation that Dorothy Bishop, a professor of Developmental Neuropsychology at the University of Oxford, performed on her own research. In her exploration, Bishop found that when comparing citation counts of her journal articles versus book chapters, her book chapters received significantly fewer citations than her journal articles.[1] Based on the outcomes of her investigation, Bishop jokingly writes, "If you write a chapter for an edited book, you might as well write the paper and then bury it in a hole in the ground." In many disciplines, particularly in the social sciences and humanities, edited collections remain an important mode of scholarly communication. This article grapples with Bishop's observations about book chapters and looks specifically at book chapters within edited collections from the field of library and information science (LIS).

An edited collection is a distinctive form of scholarly communication that includes chapters written by different authors and compiled and organized by an editor(s) into a cohesive collection.[2] Edited collections can enable different authors to reflect on, further develop, and communicate with each other around a central topic or theme.[3] Chapters are intended to convey ideas, and the collection—read as a whole—conveys connections and intersections among the chapters. Some edited collections serve as important resources for students or practitioners who want to gain an overview of a field, and may be included in instruction or on syllabi due to their usefulness to students.[4] In addition, when comparing edited collections versus monographs,[5] edited collections see a higher citation average than monographs.[6] Edited collections remain an important and distinctive form of scholarship for researchers and within academic library collections. In the field of LIS, from 2020 to 2024, the database Dimensions identified 1,420 edited books, defined as books with multiple authors and editor(s).[7]

Juleah Swanson (juleah.swanson@colorado.edu) is Associate Professor, Director of Collection Sustainability at University of Colorado Boulder, https://orcid.org/0000-0002-0095-1119.

This averaged 284 per year. During that same period, Dimensions identified only 1,158 LIS monographs or, on average, 232 books per year defined as distinctive from edited collections and generally on a single subject and often a single author.[8] The higher average publication of edited collections over traditional monographs suggests the continued importance of edited collections within the field of LIS.

For scholars and authors, there is much debate over the advantages and disadvantages of authoring a book chapter for an edited collection. Advantages of writing a book chapter may include a higher acceptance rate than journal articles, the opportunity to publish alongside experts in one's subfield, and the opportunity to engage with scholars on a shared topic and build relationships or future work together.[9] Furthermore, if an edited collection is adopted into the curriculum, a chapter may find more readership than a journal article.[10] However, as Bishop suggests, there are drawbacks and challenges in disseminating and discovering book chapters. Drawbacks of publishing scholarship in an edited collection relative to journal articles include lower citation counts, lower discoverability and readership at the book chapter level, less perceived prestige, a longer time to publication, and the perception of a less rigorous review process.[11] In addition, the quality of many edited collections can suffer as a result of disparate chapters, lack of cohesion, and lack of cross-chapter communication, resulting in edited collections that fail to advance the field meaningfully.[12]

In a follow-up to Bishop's piece, an article from *The Scholarly Kitchen* stated, "Journals long ago moved into the 'article economy,' shedding the volume/issue shell for searchable articles, in every way except citation construction and print delivery. Books remain in a packaging shell, with chapters trapped therein. There is no 'chapter economy.'"[13] Edited collections are designed to provide a unified collection of chapters that complement each other and are assembled in a single book, setting book chapters apart from journal articles. This structural requirement also poses challenges for book chapters in the digital environment that do not apply to journal articles, impacting discoverability. One significant example is the difference in use of identifiers for articles (digital object identifiers, or DOIs) versus books (ISBNs). The DOI is a persistent identifier that in part is designed to be interoperable among systems,[14] and many of the major DOI registries such as Crossref allow open and freely available metadata for use and reuse. In contrast, the ISBN identification system is not based on free and open metadata.[15] For example, Bowker is the official ISBN agency in the United States and maintains a proprietary database of metadata for publications with ISBNs and repackages and sells this metadata as commercial products such as the Books in Print database.[16] Another drawback of the ISBN, as many librarians are aware, is a single title may be subject to multiple ISBNs depending on format, such as whether it is a hardcopy, paperback, or ebook.[17]

The development of the Book Citation Index (BKCI), launched in 2011 by Web of Science, aspired to make bibliographic data on books and book chapters more available to scholars.[18] Early research on the BKCI found some significant limitations in what was indexed. For example, one study found that 83 percent of the books included were published by commercial publishers, only 14 percent were published by university presses, and three commercial publishers represented half of the publications indexed.[19] A 2018 study found that although Google Scholar is particularly strong at providing journal and

conference proceedings bibliographic data, books and edited collections bibliographic data are prone to incorrect, incomplete, duplicated, and distributed data.[20] Variations on title alone—from shortened titles to miscitations—further exacerbate issues in indexing book and book chapter citations.[21] Book chapters continue to face ongoing challenges related to indexing, metadata, and discoverability—issues that are less common among journal articles.

Open access (OA), however, may increase the readership, discoverability, and attention metrics of book chapters. One study from the field of conservation biology found that there was a statistically significant citation advantage for book chapters available as OA versus non-OA book chapters.[22] Another study by Michael Taylor theorized that one reason for these findings could be the disproportionate assignment of DOIs to book chapters among OA publishers.[23] Taylor also found that OA chapters consistently received more altmetric attention than non-OA chapters.[24]

When considering OA status for book chapters, however, the environment is often more complicated than for journal articles or even monographs. A report from the British Academy on the current state of OA for book chapters highlights that although OA for journal articles has become widespread, OA for book chapters is still relatively rare.[25] In the United Kingdom, three-fourths of book chapters submitted in the 2014 Research Excellence Framework survey were found to be outputs from publicly funded research projects.[26] As funder mandates for OA expand, there is a need to look closer at forms of scholarship beyond journal articles, as book chapters remain a consistent form of research output, some of which are results of publicly funded research.

Authors face numerous challenges to publishing their book chapters as OA. Publisher policies on book chapters are more difficult to track down than journal articles, and although a small number of publishers are responsible for issuing large portions of all book chapters, the long tail of publishers is substantial.[27] In addition, it is common for prominent scholars to be invited to contribute to an edited collection in hopes of bolstering the interest and marketability of a book. Yet, prevalent author-pay models supporting OA publication contradict the common invited-author practice.[28] An OA model that supports edited collections would need to grapple with the question of whether an invited author would be expected to fund the publication costs of the book chapter they were asked to write. Once published, authors may face challenges with tracking usage metrics of an OA book or book chapter across platforms.[29] Promisingly, though, among authors who have previously published an OA monograph, the majority have also published at least one OA chapter in an otherwise non-OA book, suggesting that efforts related to OA monographs could also be beneficial to authors publishing book chapters in edited collections.[30]

Librarians face financial conundrums with OA book chapters. Libraries and institutions have already faced the issue of "double dipping" within hybrid journals, where publishers receive payment for both a journal subscription and an article processing charge to publish articles within a paywalled journal as OA.[31] Libraries may be resistant to pay for the full cost of a monograph in which some of the chapters are made available as OA. Publishers may need to devise complicated pricing models for each titled depending on the amount of OA chapters in each work.[32]

Many of the initiatives libraries engage in to support OA are geared toward the journal article, whether it is libraries providing financial support to faculty and students to pay for article processing charges (APCs) or through transformative agreements that negotiate provisions for OA publication of articles by the institution's authors within license agreements for journal subscriptions. There are some initiatives to address OA monographs through library subvention funds that pay for the publication of an author's monograph as OA, or through collective purchasing initiatives, such as Knowledge Unlatched, in which libraries contribute funds to support the publication of a collection of ebooks as OA. Yet there has been little attention to book chapters, particularly those in edited collections, as a unit to consider for OA initiatives.

Research Questions

This research was inspired by Michael C. Calver and J. Stuart Bradley's study from conservation biology that found a statistically significant citation advantage for book chapters available as OA compared with non-OA book chapters.[33] Their study found that the mean number of citations for OA book chapters was at least twice as high as non-OA book chapters.[34]

The author applied Calver and Bradley's work to the field of LIS to investigate whether LIS book chapters also demonstrated an OA citation advantage.

The author additionally sought to explore questions about book chapters in the field of LIS, namely:

- To what extent are book chapters from edited collections in the field of LIS indexed?
- To what extent are book chapters from edited collections in the field of LIS available as OA?

Methodology

Although the author was inspired by the research questions and findings of Calver and Bradley, the methodology used to reach their conclusions on book chapters in the field of conservation biology were limited in some respects. Calver and Bradley selected for inclusion in their study four titles from the field of conservation biology, published in different years, and verified as available as ebooks. None of the titles were fully OA; titles ranged from 5 to 17 percent available as OA based on chapter count. All titles were edited collections with multiple authors at the chapter level.

In order to explore a similar research question as Calver and Bradley—plus the additional research questions—the author designed this study differently. This research was scoped to examine edited collections published within LIS within a specific copyright year. To identify publications within the LIS, the author used the Library of Congress Classification (LCC) system Class Z, which covers bibliography, library science, and information resources.[35] The author selected the copyright year 2015 to examine, to have at least five years from publication at the time of data collection.

The author used the ProQuest Oasis purchasing database to identify an initial set of edited collections to examine.[36] To create a search query in Oasis, the following parameters were selected: titles falling under the LCC Class Z* and a copyright year of 2015. The author specified book type "edited compilation" and

excluded edited compilations that were reprints of previously published journal articles. The search query was designed to filter out unwanted results and to ensure that only relevant titles meeting the specified criteria were included in the search results. This generated a list of books with a copyright year of 2015, LCC Class Z, and defined as an "edited compilation." Despite the query specifications intended to filter out such results, some titles in the dataset were compilations of previously published journal articles. The author manually identified and excluded these from analysis. The results from Oasis included title-level and not chapter-level data. The author compiled and exported date as an Excel file.

To compile the dataset of book chapters, the author used the book titles found in the Oasis dataset and manually searched each book individually in OCLC FirstSearch to obtain a table of contents of the book. The table of contents data—including book chapter titles and authors—was then compiled into a comprehensive dataset. If table of contents data was not available in FirstSearch, the author conducted an online search to locate the table of contents for each book. The final dataset—the manually compiled dataset of book chapters from books published in 2015 and classed as LCC Class Z—has been deposited and is available in CU Scholar.[37]

The author first searched Google Scholar to identify whether the chapter was indexed, and if so, whether the chapter was available as OA. This required manual, chapter-by-chapter searching. Book chapter titles were searched in Google Scholar using the bibliographic data identified in WorldCat First Search table of contents. In the majority of cases, this bibliographic information included author names. Not all book chapter titles included authors, however, and some titles included extraneous bibliographic information, such as chapter numbers or book subsection headings. For the purpose of this research, the author defined a chapter as OA in Google Scholar if the chapter was accessible via a repository, publisher site, or OA ebook aggregator without a paywall. The author manually verified access to establish whether the content was available under the earlier conditions. The author collected data from October 21, 2021, through April 22, 2022.

After analyzing the Google Scholar results, the author identified additional indexing sources. This comparison would address the author's curiosity as to whether other sources provide superior or different results compared with Google Scholar. The indexing sources—Web of Science BKCI and Dimensions—were selected because of the author's access to these sources through their institutional subscriptions and both sources' inclusion of and ability to identify edited book chapters as a format. The author acknowledges that other scholars may use different sources than the BKCI and Dimensions to identify scholarship in the field of LIS.

In both Dimensions and the BKCI, the author used the advanced search functionality to search by ISBN to retrieve all book- and book chapter-level results for the ISBN. The author searched both indices on February 7, 2025. OA chapters were identified through OA indicators in both indices' datasets. Dimensions categorizes OA articles and book chapters as Gold, Hybrid, Bronze, or Green.[38] Web of Science categorizes OA articles and book chapters as Gold, Green, and Free to Read with additional subcategories for each.[39] Results for book-level bibliographic data were removed, and only results at the chapter level were analyzed. Basic statistical methods were used to analyze the data.

Results

The author's findings and ability to address the initial research questions were ultimately impacted by the discoverability of the book chapters in the indices searched. To investigate the extent to which book chapters from edited collections in the field of LIS are indexed, the author searched 1,500 eligible book chapters from 90 edited collections across Google Scholar, BKCI, and Dimensions. Of 1,500 book chapters, 750—only 50 percent—book chapters were found across the three sources. Conversely, 750 (50 percent) book chapters were not found to be indexed in any of the three sources (figure 1). Both the BKCI and Dimensions had low percentages of book chapters indexed, 7 percent and 9 percent, respectively. Google Scholar had the highest number of book chapters indexed in total (737) and uniquely (521). BKCI had twelve chapters uniquely found in its index, whereas Dimensions had only one unique title. Nearly all of the titles found in both citation indices were also found in Google Scholar. Only one chapter was found to be indexed in all three indices, and there was no overlap in titles in BKCI and Dimensions and not Google Scholar (figure 2).

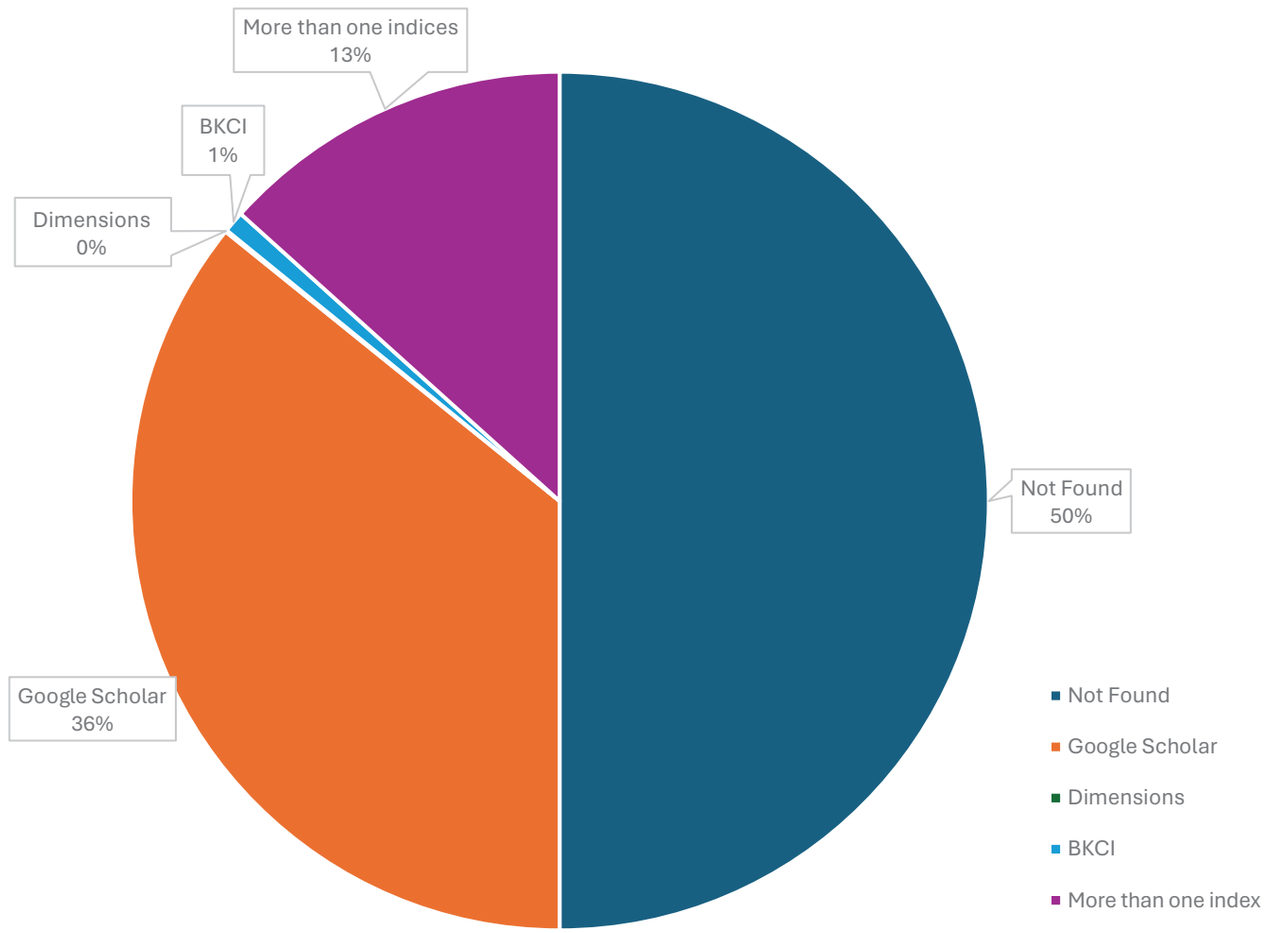


Figure 1. Comparison of book chapters searched and indexed.

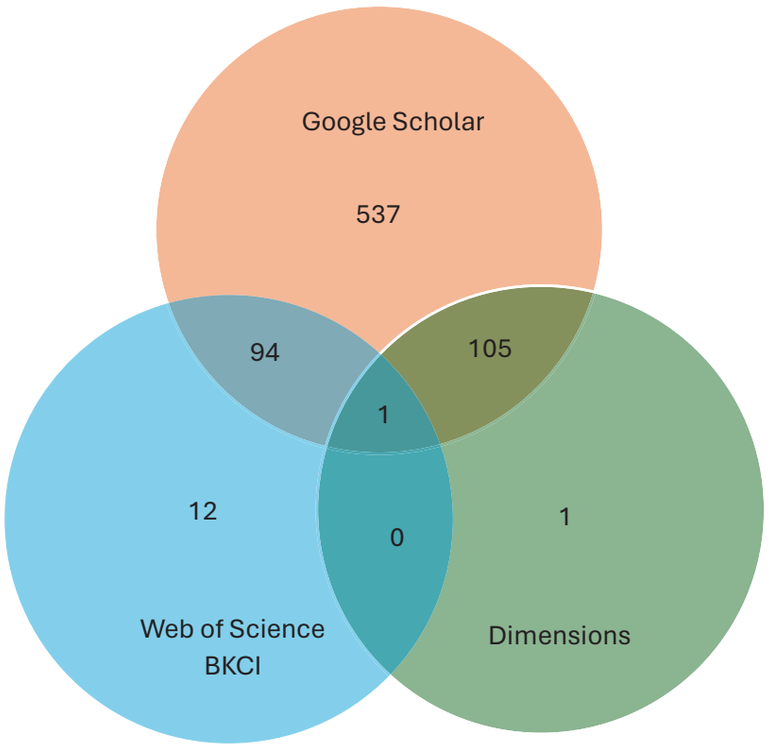


Figure 2. Comparison of book chapters indexed in Google Scholar, Web of Science Book Citation Index (BKCI), and Dimensions.

Because of the significant lack of discoverability of the book chapters in all of the indexing sources, investigation of the additional research questions was hindered. The author had intended to explore the extent to which book chapters from edited collections are available as OA and whether book chapters available as OA had a statistically significant citation advantage in comparison with non-OA chapters, as Calver and Bradley[40] had found. Because half of the dataset could not be identified in the indexing sources, however, the findings for these questions are inconclusive and cannot be adequately analyzed.

Discussion

The author's inability to answer the research questions on OA availability and citation count advantage based on the lack of indexing and discoverability of book chapters within the three major indexing sources is worth discussing. The lack of access to book chapters in the digital environment is what Bishop originally quipped about, "However good your chapter is, if readers don't have access to the book, they won't find it."[41] Scholars rely on online indexing sources to find and discover scholarship, and when half the relevant sources are not indexed, it is problematic. This may be particularly true for authors from fields not well served by indexing databases, or authors who publish with noncommercial or smaller publishers. The challenge of indexing book chapters comprehensively is surprising given that Dimensions has indexed more than 16 million book chapters, and the BKCI has indexed all the chapters from more than 160,000 titles, adding more than 10,000 new titles every year.[42]

Compared with the BKCI and Dimensions, Google Scholar provides more discoverability of both book chapters and book titles. An examination at the book title level found that seventy-eight of ninety titles (85 percent) were represented with one or more book chapters in the Google Scholar results. In comparison, Dimensions and the BKCI each had nine titles (10 percent) represented with one or more book chapters in the results. Only thirteen books were not found in Google Scholar, both at the title and chapter level.

The representation of publishers, particularly smaller and noncommercial publishers, appears to be limited across all indices for the field of LIS. When looking at the publishers for the LIS titles found in Dimensions and BKCI, only eight publishers were represented, all of whom are either based in the United Kingdom or Germany. No titles from the American Library Association or any associated imprints were indexed in either database, and only one title from Rowman & Littlefield, a prolific publisher in LIS, was found in Dimensions, although the publisher listed was Nomos Verlag rather than Rowman & Littlefield, which appears to be incorrect metadata stemming from the inclusion of the Rowman & Littlefield title as an ebook on the Nomos eLibrary, an academic ebook platform in Germany. In Google Scholar, titles published by the American Library Association and its imprints also lacked discoverability. Of the book titles not found in Google Scholar, 30 percent were published by the American Library Association and its imprints. The remaining publishers with titles not found in Google Scholar primarily came from smaller, independent presses but also included titles from larger publishers such as Bloomsbury and De Gruyter. Previous research found that in the BKCI, only 14 percent of books indexed were from university presses, only 3 percent were from noncommercial, nonuniversity academic presses, and the remaining 83 percent of titles indexed came from commercial publishers.[43] Additional findings from this previous research on the BKCI include a bias toward publishers from the United Kingdom and United States, as well as a concentration of titles indexed from three major commercial publishers: Springer, Palgrave, and Routledge represented 50 percent of the titles indexed.[44] When smaller academic presses or even university presses books are not well represented across indices, this can be particularly problematic for fields such as LIS where noncommercial publishers produce important work. It is also an important bias for librarians to recognize as they work with researchers using these indices to understand the limitations and bias of these tools and how this impacts what is and is not discoverable.

Across Google Scholar, BKCI, and Dimensions there appears to be significant inconsistencies in what gets indexed and what does not. Titles from De Gruyter, for example, were indexed in all three sources, yet not one index managed to index all of the De Gruyter titles published in 2015 in the field of LIS and searched from the dataset. It is clear, however, that the discoverability of book chapters is greater via Google Scholar than the BKCI and Dimensions. Meho and Yang found that Google Scholar was able to identify 53 percent more publications authored by LIS faculty than Web of Science and Scopus combined, but Google Scholar still lacks comprehensiveness for the field of LIS.[45] Factors that have prevented the indexing of books and book chapters in these databases could be explored in future research. For example, future research could explore whether book chapters with assigned DOIs are

discoverable in various indices. The author asserts that it would benefit both publishers and authors of edited collections published—for example, by the American Library Association and associated imprints—to include DOIs at the book chapter level for titles that are available as ebooks for these titles to become more discoverable by scholars.

The author's inability to address the research questions related to OA reflects the challenges with OA and book chapters noted in the introduction. To make meaningful advancements in models for OA at the book chapter level, more research on this topic is needed. If fundamental questions about OA and book chapters cannot be answered, such as determining how many book chapters are available as OA, then it will be challenging to develop a nuanced understanding of OA for book chapters.

Limitations

There are limitations in searching these indices due to the lack of a consistent, reliable identifier across all databases. Google Scholar was searched without using any identifier and relied on chapter title and author metadata, which may impact discovery due to inconsistent or differing metadata in Google Scholar. Both BKCI and Dimensions relied on the ISBN as the identifier to locate books and book chapters; however, as noted in the introduction, there are variations on the use of ISBN by format and which ISBN appears in bibliographic records.

Conclusion

This research highlights the continued importance of both edited collections and book chapters in fields such as LIS, while also demonstrating challenges in making book chapters discoverable. Although the author was not able to answer some of the research questions, particularly those related to OA and citation impact, the results nonetheless provide insight into book chapters as a form of scholarship and open the door for further discussions among librarians, publishers, and authors about the indexing, discoverability, and availability of book chapters within edited collections. The highly selective indexing of LIS book chapters within BKCI and Dimensions is notable and worth further conversation with vendors to better understand why most book chapters from the field of LIS are not represented in their databases. Similarly, at the publisher level, it is notable that American Library Association publications and imprints are not indexed in either BKCI or Dimensions, and they also lack comprehensive coverage in Google Scholar. Collections, acquisitions, and e-resources librarians working directly with publishers and vendors are well positioned to engage in conversations about indexing at the book chapter level, use of open metadata, registration of DOIs for book chapters, and other methods that might enhance discoverability. Although book chapters in edited collections face many discoverability and indexing challenges, the results of this research suggest promising actions library and information professionals can take to help "unbury" the book chapter.

Notes

1. Dorothy Bishop, "How to Bury Your Academic Writing," *Impact of Social Sciences*, August 29, 2012, https://blogs.lse.ac.uk/impactofsocialsciences/2012/08/29/how-to-bury-your-academic-writing/; Dorothy Bishop, "BishopBlog: How to Bury Your Academic Writing," *BishopBlog*, August 26, 2012, https://deevybee.blogspot.com/2012/08/how-to-bury-your-academic-writing.html.
2. Lindy Woodrow, "Publishing Research: Book Chapters and Books," in *Writing About Quantitative Research in Applied Linguistics*, ed. Lindy Woodrow (Palgrave Macmillan UK, 2014), https://doi.org/10.1057/9780230369955_14.
3. Debbie Epstein, Jane Kenway, and Rebecca Boden, "Publishing Books and in Books," in *Writing for Publication* (SAGE Publications Ltd, 2011), https://dx.doi.org/10.4135/9780857020314.
4. Pat Thomson, "Is Writing a Book Chapter a Waste of Time?" *Impact of Social Sciences*, August 31, 2012, https://blogs.lse.ac.uk/impactofsocialsciences/2012/08/31/is-writing-a-book-chapter-a-waste-of-time/.
5. This article defines *monograph* as a book-length scholarly work that is typically solo-authored but can be multiauthored and is distinctive from edited collections as discussed in Peter Williams, Iain Stevenson, David Nicholas, Anthony Watkinson, and Ian Rowlands, "The Role and Future of the Monograph in Arts and Humanities Research," *Aslib Proceedings* 61, no. 1 (2009): 67–82, https://doi.org/10.1108/00012530910932294.
6. Daniel Torres-Salinas, Nicolás Robinson-García, Álvaro Cabezas-Clavijo, and Evaristo Jiménez-Contreras, "Analyzing the Citation Characteristics of Books: Edited Books, Book Series and Publisher Types in the Book Citation Index," *Scientometrics* 98, no. 3 (2014): 2113–27, https://doi.org/10.1007/s11192-013-1168-4.
7. Digital Science, "Publication Year Is 2024 or 2023 or 2022 or 2021 or 2020, Fields of Research (ANZSRC 2020) Is 4610 Library and Information Studies, Publication Type Is Edited Book," September 19, 2025, Dataset, https://app.dimensions.ai.
8. Digital Science, "Publication Year Is 2024 or 2023 or 2022 or 2021 or 2020, Fields of Research (ANZSRC 2020) Is 4610 Library and Information Studies, Publication Type Is Monograph," September 19, 2025, Dataset, https://app.dimensions.ai/.
9. Woodrow, "Publishing Research."
10. Ibid.
11. Ibid.
12. Peter Webster, *The Edited Collection: Pasts, Present and Futures*, Elements in Publishing and Book Culture Series (Cambridge University Press, 2020).
13. Kent Anderson, "Bury Your Writing—Why Do Academic Book Chapters Fail to Generate Citations?" The Scholarly Kitchen, August 28, 2012, https://scholarlykitchen.sspnet.org/2012/08/28/bury-your-writing-why-do-academic-book-chapters-fail-to-generate-citations/.
14. DOI Foundation, "What Is a DOI?" accessed January 30, 2026, https://www.doi.org/the-identifier/what-is-a-doi/.
15. Michael Taylor, "An Altmetric Attention Advantage for Open Access Books in the Humanities and Social Sciences," *Scientometrics* 125, no. 3 (2020): 2523–43, https://doi.org/10.1007/s11192-020-03735-8.

16. "Bowker Statement of Use | ISBN.Org," accessed January 30, 2026, https://www.isbn.org/statement_of_use.
17. "What Is an ISBN? | International ISBN Agency," accessed March 5, 2025, https://www.isbn-international.org/content/what-isbn/10.
18. "Book Citation Index," Clarivate, n.d., accessed June 17, 2026, https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/book-citation-index/.
19. Torres-Salinas et al., "Analyzing the Citation Characteristics of Books."
20. David Zeitlyn and Megan Beardmore-Herd, "Testing Google Scholar Bibliographic Data: Estimating Error Rates for Google Scholar Citation Parsing," *First Monday*, 23 no. 11 (2018), https://doi.org/10.5210/fm.v23i11.8658.
21. Judit Bar-Ilan, "Citations to the 'Introduction to Informetrics' Indexed by WOS, Scopus and Google Scholar," *Scientometrics* 82, no. 3 (2010): 495–506, https://doi.org/10.1007/s11192-010-0185-9.
22. Michael C. Calver and J. Stuart Bradley, "Patterns of Citations of Open Access and Non-Open Access Conservation Biology Journal Papers and Book Chapters," *Conservation Biology* 24, no. 3 (2010): 872–80, https://doi.org/10.1111/j.1523-1739.2010.01509.x.
23. Taylor, "An Altmetric Attention Advantage for Open Access Books in the Humanities and Social Sciences."
24. Ibid.
25. The British Academy, "Open Access and Book Chapters," The British Academy, October 2019, https://www.thebritishacademy.ac.uk/publications/open-access-book-chapters-report/.
26. Ibid.
27. Ibid.
28. Ibid.
29. Ronald Snijder, "Measured in a Context: Making Sense of Open Access Book Data," *Insights* 36, no. 1 (2023), https://doi.org/10.1629/uksg.627.
30. Ros Pyne, Christina Emery, Mithu Lucraft and Anna Sophia Pinck, "The Future of Open Access Books: Findings from a Global Survey of Academic Book Authors" (Springer Nature), June 2019, https://figshare.com/s/a11075d04d333dba4ee0.
31. Bernhard Mittermaier, "Double Dipping in Hybrid Open Access—Chimera or Reality?" *ScienceOpen Research*, ahead of print, May 25, 2015, https://doi.org/10.14293/S2199-1006.1.SOR-SOCSCI.AOWNTU.v1.
32. The British Academy, "Open Access and Book Chapters."
33. Calver and Bradley, "Patterns of Citations of Open Access."
34. Ibid.
35. "Library of Congress Classification Outline. Class Z – Bibliography. Library Science. Information Resources (General)," Library of Congress, n.d., accessed June 23, 2023, https://www.loc.gov/catdir/cpso/lcco/lcco_z.pdf.
36. "OASIS," accessed July 7, 2023, https://about.proquest.com/en/products-services/OASIS/.

37. Juleah Swanson, "2015 Library and Information Science Book Chapters from Edited Collections," CU Scholar, n.d., accessed February 3, 2026, https://scholar.colorado.edu/concern/datasets/37720f60q.

38. "Where Does the Definition of 'Open Access' Come from in Dimensions? What Does It Include?" Dimensions, accessed February 2, 2026, https://help.dimensions.ai/en/articles/9783681.

39. "Open Access," Web of Science, August 6, 2025, https://webofscience.zendesk.com/hc/en-us/articles/20016332411281-Open-Access.

40. Calver and Bradley, "Patterns of Citations of Open Access."

41. Bishop, "How to Bury Your Academic Writing"; Bishop, "BishopBlog."

42. Digital Science, "Publication Year Is 2024 or 2023 or 2022 or 2021 or 2020, Fields of Research (ANZSRC 2020) Is 4610 Library and Information Studies, Publication Type Is Monograph;" "Book Citation Index," Clarivate accessed May 8, 2023, http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/.

43. Torres-Salinas et al., "Analyzing the Citation Characteristics of Books."

44. Ibid.

45. Lokman I. Meho and Kiduk Yang, "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar," *Journal of the American Society for Information Science and Technology* 58, no. 13 (2007): 2105–25, https://doi.org/10.1002/asi.20677.

NOTES ON OPERATIONS

Measure, Change, Measure

Assessing Changes to a Traditional Approval Plan

Jennifer A. Mezick, Louis T. Becker, and Kat Brooks

This article details the impact and outcomes of a case study designed to improve the book approval plan at a Carnegie R1 academic library, with implications for assessment of approval plans at large. The project analyzes data from two different time periods: 2019 and 2023. Between these periods, a large update was made to the parameters that control the approval plan as a way to improve the performance of the plan. The authors report on the compilation of purchase and use data and detail their assessment. Findings indicate that using these types of data to guide changes can positively impact the quantity and use of titles arriving on approval.

Introduction

Approval plans are tools intended to save librarians time on collection development without spending hours reading book reviews and placing individual orders. They are an automated, just-in-case[1] method for collection development that relies on various parameters and criteria established by the library and vendor to match books that are then sent to the library. Like many large libraries, the University of Tennessee Libraries maintain approval plans that require no review or approval step. Despite the benefit of time savings, if left unchecked, approval plans can run the risk of overspending limited budgets and supplying titles that do not align with institutional research or curricula.

The 2019 publication, "Using Data to Rightsize Approval Spending," details the University of Tennessee Libraries' effort to assess the primary approval plan and update criteria with the goal to reduce spending and increase the number of titles used.[2] The approval plan was reevaluated in 2023 to determine whether the 2019 updates met the intended goals, and if so, which adjustments led to meaningful changes. This article compares the 2019 and the 2023 data to draw conclusions about the changes made between these periods. For clarity throughout this article, purchase and use activity are referred to by the years analyses were conducted: 2019 and 2023. The primary indicators considered are the number of titles purchased through the plan and the number of those titles that showed evidence of use within two years of their purchase. Areas of the plan that had no changes applied to Library of Congress (LC) classification parameters are used as a control group and provide a baseline to understand the impacts of changes made to LC areas. Overall, changes made to the approval plan were successful in reducing the number of titles purchased and broadly increasing the percentage of

Jennifer A. Mezick (jamezick@utk.edu) is Associate Professor, Associate Dean for Research, Collections, and Assessment at University of Tennessee, Knoxville, https://orcid.org/0000-0003-4505-9316. **Louis T. Becker** (lbecker@utk.edu) is Assistant Professor, Assessment Programs Librarian at University of Tennessee, Knoxville, https://orcid.org/0000-0002-8283-2593. **Kat Brooks** (kbrooks@utk.edu) is Assistant Professor, Collections Strategist at University of Tennessee, Knoxville, https://orcid.org/0000-0003-0568-5179.

automatic purchases that were used. Libraries interested in improving approval plans could adopt the method described in this article for their own analyses.

Literature Review

As approval plans have become ingrained in library collection development practices, many researchers have added to the decades-long stream of literature and discourse with case studies assessing the effectiveness and potential uses of approval plans.[3] Much of this work focuses on approval plans as new additions to a library's ecosystem or assessments to determine whether an approval plan is successful or justified to continue.[4] For many libraries, budget pressure and the staff time required to manage plans are the impetus for assessing approval plans.[5]

Circulation statistics are the leading metric used to assess approval plans, although the definition of a successful circulation varies for each environment.[6] Comparing the usage of print books and ebooks to assess the performance of an approval plan is common. Although a theme of higher use in ebooks than print books purchased via approval plan is generally found, other findings vary widely in different environments and contexts.[7]

Some researchers suggest that defining success by the number of circulations points back to historical literature that predates our modern library systems. Fry called this conventional wisdom into question ten years ago, pointing out the lack of data to support these claims.[8] This critique encourages librarians to reconsider what they frame as success and whether circulation is the most valuable metric. With approval plans capable of making sweeping, high-level changes, a study of circulation changes over time and in step with plan updates can help examine that question in the modern acquisitions landscape.

Even with the long history of approval plans in libraries, very little research on assessing approval plans to improve their performance has been added to scholarly discourse. Studies that investigate the success of changes made to approval plans are even more difficult to find. Despite the widespread adoption of approval plans and a longstanding history of assessments, a distinctive gap still exists in the literature. Because approval plans are treated as standard practice in collection development, evaluation and ongoing improvement of performance should be standard practice as well. This article intends to move the discussion forward and begin to address that gap.

Method

The authors utilize print and ebook data from the Libraries' main approval plan provider, EBSCO Information Services' GOBI Library Solutions (GOBI), and integrated library system, Alma, to build the 2019 and 2023 datasets.[9] Titles purchased via the approval plan are identified using the financial records of a fund dedicated to the approval plan. Although the Libraries' approval plan does provide notifications for books of possible interest (termed "slips" by GOBI), this analysis includes only purchases made automatically (termed "shipment" by GOBI). Books eventually purchased based on slip

recommendations are outside the scope of this analysis because they do not provide significant time savings for the collections or acquisitions teams and therefore do not meet the automation benefits of approval plans. Title-level use and bibliographic data were extracted from Alma for analysis. To make the datasets as comparable as possible, we selected an equivalent period of data (thirty-one months) from the available data. For this comparative analysis, the 2019 set includes 27,177 purchases made between May 19, 2016, and December 21, 2018, and the 2023 set includes 18,758 purchases made between November 18, 2019, and June 30, 2022.

Classification Data

GOBI's approval plan is structured largely around LC classifications with additional parameters for publisher, format, and interdisciplinary considerations. Each time the approval plan is updated or changed by the library, those changes are noted in a table of more than 3,000 ranges. For this study, purchases are considered part of the LC class in which they are cataloged locally. Occasionally local classification practices present a limitation to the study by placing a title in a different LC class than the class GOBI assigns. These titles are generally interdisciplinary treatments of a subject. Titles that could not be assigned to a classification range are excluded from the comparative analysis.

GOBI's classification range system includes hierarchically nesting classes. For example, it is possible to set purchasing instructions for the parent range class BL (Religion), to set different criteria for the child range BL175-299 (Natural Theology), and yet to revert BL238 (Fundamentalism) to the overarching criteria for BL. Changes made to parent ranges are coded onto each of their child ranges; however, if a parameter already applied to a child range was newly applied to the entire parent range, that child range was coded as unchanged. A wide variety of changes were flagged and ultimately categorized based on their primary action:

- No change/unchanged: classification ranges to which no change was made between the 2019 and 2023 datasets. This includes sections of the humanities, social sciences, and natural sciences, and serves as a control group for the analysis.
- Removed: classification ranges that purchased titles automatically in the 2019 dataset and were excluded from automatic purchases in the 2023 set. A variety of subjects are represented in this category and are often subranges rather than wide ranges. Exceptions and other nonclassification parameters can still trigger purchases in these ranges.
- Added: the group of classification ranges that were blocked from automatic purchases in the 2019 dataset and allowed automatic purchases in the 2023 set. Changes in this set usually affect smaller subtopics in select areas of the humanities and social sciences.
- E-preferred: classification ranges purchasing print books instead of the ebook equivalent in the 2019 dataset and updated in the 2023 dataset to purchase ebooks instead of the print equivalent, if available. Ranges across several disciplines were included in this change category. Titles in e-preferred ranges sometimes arrive in print because they are not available in electronic format within a set time after the initial print publication.

- Added exception: classification ranges updated with specifications in the 2023 dataset that differ from the entire approval plan or the parent range. Exceptions include inclusion or exclusion of a specific publisher, type of material, or interdisciplinary aspect of a topic. For example, the approval plan may be scoped to purchase titles published only in English, but a subset of ranges could have an exception to allow for titles published in German. For this study, exceptions related to format are coded as e-preferred or print preferred.
- Removed exception: the group of classification ranges in the 2023 dataset updated with the removal of specifications such as those described earlier. This causes a reversion to the parent range instructions or to the plan parameters at large if no parent range exists.
- Print preferred: classification ranges updated to purchase print books instead of the ebook equivalent. For this study, this change was documented only for a single child range (HT388-400). Very few books were purchased in either study period in this area, and none were used. Due to the small number of titles impacted, the change is not discussed in depth in this study.

This comparative analysis assesses the use of each title individually. To allow for fair comparison between print and electronic titles across multiple generations of COUNTER electronic resource usage reporting schemes in the two analyses, titles are counted as used if there is evidence of use within twenty-four months of the purchase date. Print usage is restricted to external loans only and excludes any in-house or processing-type interactions. Electronic use is compiled using title or chapter downloads from COUNTER Release 4 (R4) Book Report 1 (BR1) or Book Report 2 (BR2) for use before June 30, 2020, and using unique title requests from COUNTER Release 5 (R5) title reports for use after January 1, 2020. In the COUNTER R4 period, some vendors provided only BR1 and others only BR2, and usage is derived from whichever report is available. Similarly, reporting transitioned to the COUNTER R5 code of practice during the spring of 2020, and usage on either version of report is counted.

To help account for the natural variation in the number of new books appearing in different subject areas over time, this study uses a standard error calculation.[10] The geometric mean of the standard error for the change group and the standard error for the no change group (the control group) were used to make comparisons between different segments of the plan. A z score that compares the change's magnitude with the standard error is used to assess the significance of the change. Changes greater than 1.96 times the standard error are considered significant ($p < 0.05$). Similar statistical calculations are used to assess the shifts in the share of titles used and the share of titles purchased in electronic format.

Limitations to the Data

Titles are matched to electronic usage reports by ISBN and to print circulation records at the bibliographic record level. If multiple items of a title exist in the collection, the use of any one of those items is included in this evaluation. Multivolume titles are counted as used if any volume has been used during the appropriate time frame. Some electronic titles purchased on approval are also available in our catalog through aggregator subscriptions. Matching at the ISBN level means that use through alternate access points, such as subscription ebook packages, counted as use of the title. The purchase

date used for all titles is the date that the order was paid for in Alma. Some titles are received and used prior to this purchase date, and that use is accounted for in the data.

Each LC area is coded with only one category of change. In cases where there is more than one change, the category of change that is associated with greater impact is coded to a range. For example, a change in format is presumed to be impactful to more titles in a range than an exception to a topical aspect. Further study at a more granular level may uncover nuances in the effect of changes to the approval plan.

The data considers use of the purchased titles within a relatively short evaluation period (two years from the date of payment) and does not account for long tails of use often observed in some disciplines.[11] In addition to the short periods of use, the 2023 dataset includes five months when the university was closed and an additional five months when most students were distance learners due to the COVID-19 pandemic. These disruptions are a limitation to the usage data for both print and ebooks because overall use of the Libraries' electronic resources experienced a longer than normal summer dip during 2020. If business had been normal during this period, higher use might have been reported. Despite these limitations, observations in the data still provide evidence that some approval plan changes positively impact the use of purchased titles.

Results and Insights

Changes Outside the Approval Plan

Continual development in library acquisitions and substantial differences in the publishing landscape are threaded throughout the two review periods. Three big changes outside the control of the approval plan contribute to a difference in book purchasing between datasets. First, the COVID-19 pandemic reduced the number of titles produced for reasons such as delays at printing factories due to closures or decreased staff and disruptions in supply chains.[12] Based on GOBI reports, the number of total titles profiled for all customers, regardless of matching with the Libraries' approval plan, declined by 1.1 percent between the two periods, from 180,961 titles profiled between May 19, 2016, and December 23, 2018, to 178,962 titles profiled between November 1, 2019 and June 30, 2022.[13] At the same time, the share of titles profiled in electronic form increased from 79 percent of titles in the 2019 period to 82 percent of titles in the 2023 period.

Second, as part of the changes that impact the 2023 data, the Libraries began moving all eligible electronic titles selected for the approval plan to a demand-driven acquisition (DDA) plan, reducing the number of electronic titles purchased on approval while still providing access to those titles. This process was intentionally created to help keep approval plan costs under control. Because DDA purchases are not accounted for in this analysis, their absence reduces the overall number of purchased electronic titles in the datasets. Between November 2019 and June 2022, GOBI reported that 1,013 electronic titles were manually blocked from arriving on the Libraries' approval plan. Most blocked electronic titles are those that were moved to DDA.

Third, an evidence-based acquisition (EBA) plan for Taylor and Francis (T&F) titles was adopted in July 2021, and all T&F imprints were blocked from being purchased via approval. In the 2019 dataset, T&F imprints account for 16 percent of titles purchased. In the 2023 dataset, T&F imprints account for 21 percent of purchases before July 2021 when the EBA was implemented. Average quarterly approval purchases declined by 18 percent after July 2021. This decline is most notable in the newly added or unchanged ranges of the 2023 dataset.

Considerable alterations impact the overall number of titles purchased on approval and present the reality that even segments of the plan that are unchanged show significant shifts in the number of titles purchased, the share of those titles purchased in electronic format, and the share of purchased items used during the study period. Alterations outside the approval plan have a significant impact on the total number of titles purchased, with 8.4 percent (1,048) fewer titles purchased in the unchanged category of the plan. Usage and purchase comparisons are made with the unchanged ranges as a control group to partially account for these external factors.

Change in Number of Titles Purchased

One of the original goals of assessing and updating the approval plan was to reduce the overall amount spent to make sure that the budget pattern is sustainable and predictable. Historically, there was little attempt to predict how much the approval plan would cost, and it subsequently and consistently overran its allocated budget. When this happened, the plan was often paused and titles forfeited or shipped and billed later. The inconsistency this creates is visible in the odd distributions of the pre-2020 data in figure 1. The fourth quarter of 2019 remains an outlier in the data because the approval

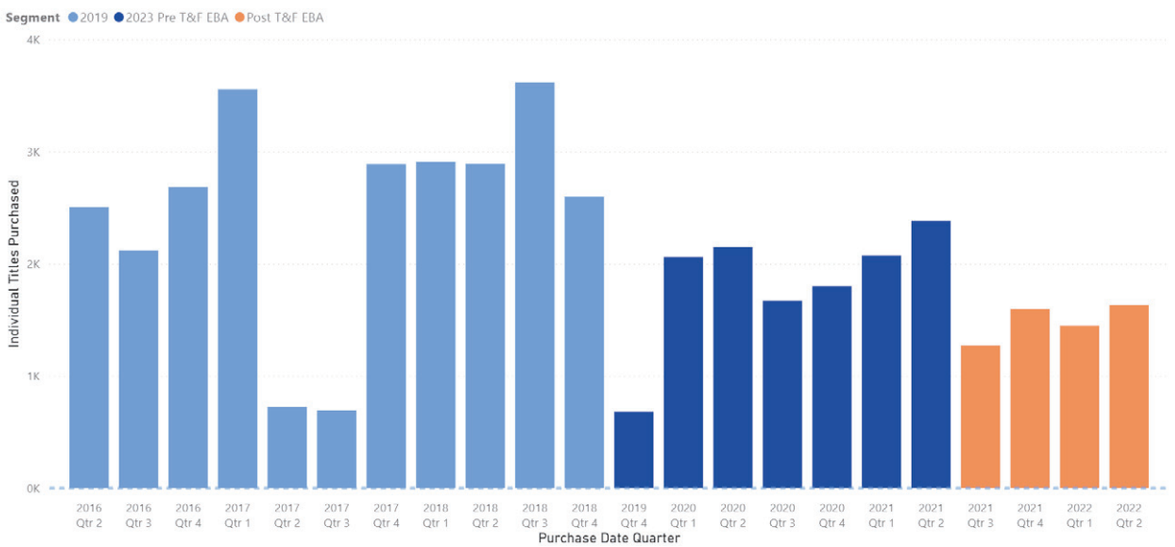


Figure 1. Number of titles purchased quarterly by dataset with purchases after implementation of T&F EBA differentiated within the 2023 data.

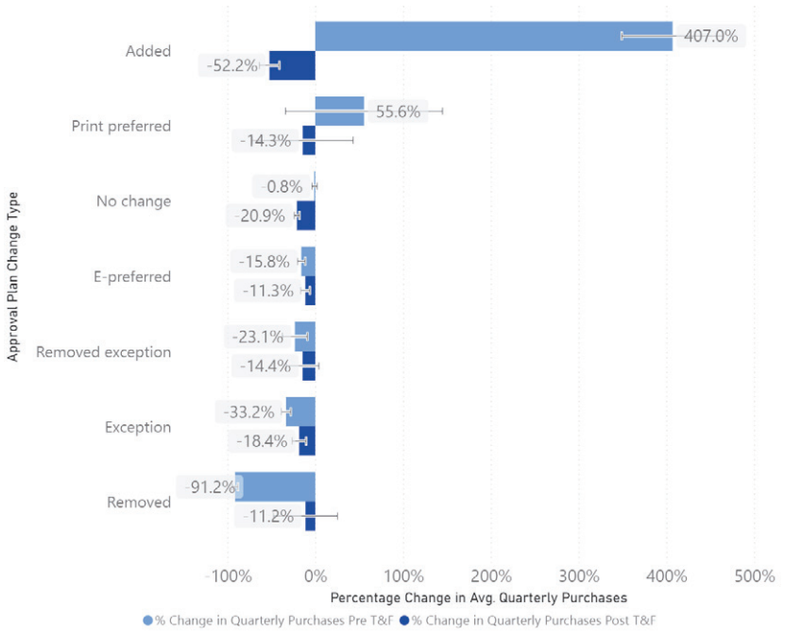


Figure 2. Percent change in average quarterly purchases before and after implementation of the T&F EBA.

plan was largely paused while being updated. To reduce the overall cost, changes were made with the intention of purchasing fewer titles in LC ranges with poor use.

Overall, 30.98 percent (8,419) fewer titles were purchased in the 2023 set compared with the 2019 set. Unchanged areas show a 14.87 percent increase in electronic purchases and a 19.49 percent decrease in print purchases for an overall 8.4 percent decline (1,048 fewer titles) in the control group. As demonstrated in figure 2, quarterly purchases in the areas without change decrease by more than 20 percent after the implementation of the T&F EBA, leading to the conclusion that the mid-2021 removal of T&F publications is likely a major factor in the decrease in purchases overall.

As shown in table 1, removing or adding classifications from the approval plan is the change type that most directly affects the number of titles purchased. Changing format preferences also has a clear effect, even against the backdrop of increasing electronic availability of monographs. Ebook purchases increased by 39 percent overall in the 2023 dataset. Moving ranges from print to e-preferred resulted in 3,062 additional ebook purchases from a baseline of twenty-one purchases in the 2019 dataset (14,581 percent increase). This is a substantial shift in comparison with other areas of the plan. Purchase of ebooks in ranges that did not change to e-preferred increased from 31.8 percent of the 2019 set to 46 percent of the 2023 set (15 percent increase). At the same time, purchases in the e-preferred areas decreased overall relative to the unchanged classifications. How this decrease relates to the change in format preference is unclear.

Table 1. Changes in Purchases (2023 over 2019)

Category of Change	Format	Purchases 2019	Purchases 2023	Difference in Purchases	% Difference in Purchases	% Difference in Purchases vs. No Change Group	Standard Error: (% +/−) Difference in Purchases vs. No Change Group	z Score: Difference in Purchases vs. No Change Group
Added	E	42	189	147	350.00%	335.1%	36.26%	9.24
	P	42	156	114	271.43%	290.9%	33.53%	8.68
	Subtotal	**84**	**345**	**261**	**310.71%**	**319.1%**	**24.69%**	**12.93**
E-preferred	E	21	3,083	3,062	14,580.95%	14,566.1%	265.31%	54.9
	P	5,150	1,094	(4,056)	−78.76%	−59.3%	2.12%	27.98
	Subtotal	**5,171**	**4,177**	**(994)**	**−19.22%**	**−10.8%**	**2.24%**	**4.83**
Exception	E	1,502	1,369	(133)	−8.85%	−23.7%	4.25%	5.58
	P	1,631	583	(1,048)	−64.26%	−44.8%	3.23%	13.84
	Subtotal	**3,133**	**1,952**	**(1,181)**	**−37.70%**	**−29.3%**	**2.59%**	**11.31**
No change	E	*4,035*	*4,635*	*600*	*14.87%*	*0.0%*	*3.26%*	*0*
	P	*8,457*	*6,809*	*(1,648)*	−19.49%	0.0%	2.07%	0
	Subtotal	***12,492***	***11,444***	***(1,048)***	***−8.39%***	***0.0%***	***1.75%***	***0***
Print preferred	E	9	–	(9)	−100.00%	−114.9%	33.41%	3.44
	P	–	*9*	*9*	–	–	–	–
	Subtotal	***9***	***9***	**–**	***0.00%***	***8.4%***	***47.16%***	***0.18***
Removed	E	1,157	271	(886)	−76.58%	−91.4%	4.00%	22.87
	P	4,663	219	(4,444)	−95.30%	−75.8%	2.09%	36.23
	Subtotal	**5,820**	**490**	**(5,330)**	**−91.58%**	**−83.2%**	**1.84%**	**45.14**
Removed exception	E	265	250	(15)	−5.66%	−20.5%	8.87%	2.31
	P	203	91	(112)	−55.17%	−35.7%	8.57%	4.16
	Subtotal	**468**	**341**	**(127)**	**−30.98%**	**−18.70%**	**6.20%**	**3.02**

Note: Rows in italics indicate no significant difference from the no change group.

Areas with added or removed exceptions show fewer purchases relative to the unchanged areas of the plan, with print purchases declining to a greater extent than electronic purchases. It is unknown whether individual exceptions meet their intended purposes without further analysis, but this finding suggests that targeted changes to approval criteria can be effective in decreasing unwanted purchases, without removing ranges altogether.

Use of Purchased Titles

Overall, print use decreased significantly across the study period. For the entire Libraries collection, not limited to titles purchased in the approval analysis, Alma recorded 158,542 loans of physical books during the time frame of the 2019 dataset and only 72,384 loans between September 2019 and June 2022 (time frame covered by the 2023 dataset). Titles purchased on approval in print decline in usage across the study, with nearly 22 percent of print titles in the 2019 dataset loaned within two years of purchase (4,350 titles used of 20,146 purchased). In the 2023 dataset, despite the significant decrease in print purchases, only 14.8 percent of print titles were loaned within two years of purchase (1,329 titles used of 8,961 purchased). Major external factors, such as the five-month closure of the library because of COVID-19 in 2020, negatively impacted an already decreasing trend in print circulation. Unchanged ranges in the approval plan serve as a control group to account for these factors and others outside the approval plan changes. This allows for determination of whether increases or decreases in use are significant across the categories of change.

Although print usage decreased overall, areas changed to e-preferred or removed had less of a decrease compared with the unchanged areas. Sections changed to e-preferred or removed also purchased substantially less print. In contrast, newly added ranges purchased more print in the 2023 set than in the 2019 set and show more of a decrease in print usage compared with the unchanged areas. This analysis demonstrates that increasing print purchases, even in a small area, is ineffective at increasing print usage. Adding or removing exceptions did not significantly affect print usage rates. Given the broader environmental shift away from print, changes to the approval plan that do not reduce print purchases have minimal impact on countering declining print usage.

Electronic titles, however, show substantial increases in use. In the 2019 dataset, 28.5 percent of electronic purchases were used within two years of their purchase date (2,005 titles used of 7,031 purchased). In the 2023 set, this percentage increased to nearly 40 percent (3,908 titles used of 9,797 purchased). Because electronic title use can be associated with a patron accessing only a single chapter, it is easier to track electronic usage compared with circulation of a print item, which is recorded only when a patron takes an item out of the library. As print circulation has declined overall, patrons are also increasingly using electronic monographs.[14]

Figure 3 shows that areas switched to e-preferred see the biggest increase in usage rates across formats, with overall usage increasing by 16.72 percent (±1.08 percent) compared with areas that experienced no change. These areas purchased substantially less print in the 2023 dataset compared with the 2019 dataset, and although print use in these areas decreased, it decreased less than the unchanged control group. At the same time, the overall increase in ebook usage helped boost the use of purchased titles in areas changed to e-preferred and suggests that purchasing ebooks in these areas better meets patron needs.

Two categories of changes show significant improvement in electronic usage compared with the areas that went unchanged. The small number of books purchased in ranges removed from the approval plan have the highest usage rate (112 e-titles used of 271 purchased, or 41.33 percent) with 12.32

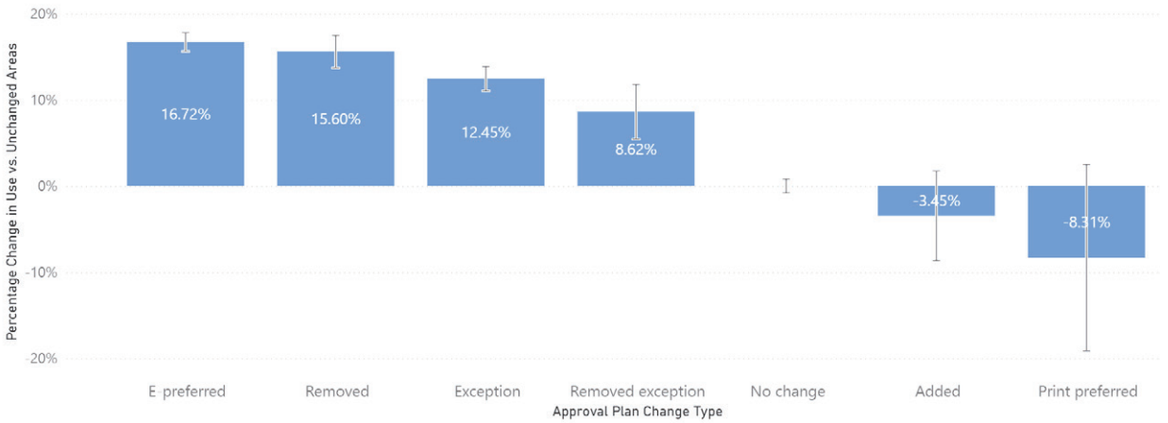


Figure 3. Changes to usage shown by coded categories of adjustment.

percent (±3.05 percent) increase in use above unchanged areas. Titles purchased in the removed ranges are far fewer than in other ranges and would have been acquired through an exception to the range or differences in GOBI and local cataloging practices. Both scenarios suggest that the titles are interdisciplinary in nature. Use of these titles demonstrates value to our users where single-discipline titles in these areas that match parent range parameters or defaults for the plan do not. In contrast with areas removed, areas newly added declined in usage, but the decline is not significant when compared with the areas that went unchanged. Areas in which exceptions were added also have substantial increases in electronic usage with a 7.37 percent (±2.06 percent) improvement above the areas that went unchanged. Table 2 provides an overall summary of changes in use. Driven largely by the shift to electronic purchases, the combined changes resulted in a 7.33 percent (±0.70 percent) increase in the number of titles used in any format within two years of purchase compared with the unchanged areas.

Table 2. Changes to Usage Broken Down By Coded Categories of Adjustment

Category of Change	Format	% Used 2019	% Used 2023	Difference in % Used	Standard Error: Difference in % Used	Difference in % Used vs. No Change Group	Standard Error: Difference in % Used vs. No Change Group	z Score: % Used vs. No Change Group
Added	E	*19.05 (8/42)*	*30.69 (58/189)*	*11.64%*	*7.71%*	*4.03%*	*7.77%*	*0.52*
	P	38.1 (16/42)	12.18 (19/156)	−25.92%	6.63%	−15.25%	6.66%	2.29
	Subtotal	***28.57 (24/84)***	***22.32 (77/345)***	***−6.25%***	***5.16%***	***−3.45%***	***5.19%***	***0.67***
E-preferred	E	*19.05 (4/21)*	*41.03 (1,265/3,083)*	*21.98%*	*10.76%*	*14.38%*	*10.81%*	*1.33*
	P	20.58 (1,060/5,150)	16.09 (176/1,094)	−4.49%	1.33%	6.17%	1.48%	4.17

Category of Change	Format	% Used 2019	% Used 2023	Difference in % Used	Standard Error: Difference in % Used	Difference in % Used vs. No Change Group	Standard Error: Difference in % Used vs. No Change Group	z Score: % Used vs. No Change Group
	Subtotal	**20.58 (1,064/5,171)**	**34.5 (1,441/4,177)**	**13.92%**	**0.92%**	**16.72%**	**1.08%**	**15.45**
Exception	E	27.83 (418/1,502)	42.8 (586/1,369)	14.98%	1.78%	7.37%	2.06%	3.58
	P	*20.97 (342/1,631)*	*13.04 (76/583)*	*−7.93%*	1.89%	2.73%	2.00%	1.37
	Subtotal	**24.26 (760/3,133)**	**33.91 (662/1,952)**	**13.92%**	**0.92%**	**12.45%**	**1.41%**	**8.81**
No change	*E*	*31.23 (1,260/4,035)*	*38.83 (1,800/4,635)*	*7.61%*	*1.03%*	*0.00%*	*1.46%*	*0.00*
	P	*25.53 (2,159/8,457)*	*14.86 (1,012/6,809)*	*−10.67%*	*0.66%*	*0.00%*	*0.93%*	*0*
	Subtotal	***27.37 (3,419/12,492)***	***24.57 (2,812/11,444)***	***−2.80%***	***0.57%***	***0.00%***	***0.80%***	***0***
Print preferred	E	*11.11 (1/9)*	*0 (0/0)*	*−11.11%*	*10.48%*	*−18.72%*	*10.53%*	*1.78*
	P	0 (0/0)	0 (0/9)	0.00%	0.00%	10.67%	0.66%	16.15
	Subtotal	***11.11 (1/9)***	***0 (0/9)***	***−11.11%***	***10.80%***	***−8.31%***	***10.81%***	***0.77***
Removed	E	21.52 (249/1,157)	41.33 (112/271)	20.26%	2.88%	12.32%	3.05%	4.03
	P	15.61 (728/4,663)	15.07 (33/219)	−0.54%	2.51%	10.12%	2.59%	3.90
	Subtotal	**16.79 (977/5,820)**	**29.59 (145/490)**	**12.80%**	**1.80%**	**15.60%**	**1.89%**	**8.27**
Removed exception	E	*24.53 (65/265)*	*34.8 (87/250)*	*10.27%*	*4.02%*	*2.66%*	*4.15%*	*0.64*
	P	*22.17 (45/203)*	*14.29 (13/91)*	*−7.88%*	*5.02%*	*2.78%*	*5.06%*	*0.55*
	Subtotal	**23.5 (110/468)**	**29.33 (100/341)**	**5.82%**	**3.12%**	**8.62%**	**3.17%**	**2.32**

Note: Rows in italics indicate no significant difference from the no change group.

Conclusion

The changes assessed in this study met the goal of reducing overall expenditures and increasing overall usage on the Libraries' main approval plan. Approval plan spending decreased by 15 percent in the 2023 dataset compared with the 2019 set, whereas the reduction in titles purchased was almost 31 percent. The average cost per title for each format did not change significantly between the two

datasets; however, electronic books cost an average of twice their print counterpart, and the significant increase in electronic books purchased in the 2023 dataset offsets a more significant cost savings. Still, this reduced spending is enough to provide a more stable structure for budgeting and staff time, while the overall increase in use demonstrates better alignment with patron needs. The study demonstrates that quantitative analysis of purchases and their subsequent usage can inform changes to improve automated approval plans. Gathering and analyzing this data is time consuming but provides a big benefit in allowing libraries to successfully manage approval plans as tools to save the time of individual selectors. The most impactful categories of changes are those made to exclude areas, vastly reducing the number of titles automatically purchased; changing areas to e-preferred, resulting in a large number of ebooks that correlate to more use without impacting the titles purchased; and adding exceptions to refine the topics that trigger automatic purchases, consequently increasing use of the titles purchased.

This study includes significant variables that impact comparison of the two datasets and should be followed up with an additional analysis to continue assessing changes. Even in the absence of those variables, however, research interests, university curricula and programs, and library acquisition methods are always evolving, which requires that any approval plan be reviewed and updated periodically to ensure relevance. Possibilities for assessment of and changes to approval plans depend on the categories and options provided by the vendor.

The rise of DDA and EBA plans that prioritize electronic access and patron choice challenges the traditional approval plan. The increase in the percentage of approval titles from the T&F group in the first part of the 2023 dataset provides additional justification for the Libraries' adoption of the EBA. Moving these imprints from the approval plan to an EBA reduces costs associated with invoicing, processing, and physical space in exchange for patron access to T&F's entire catalog. Future comparisons of the use of titles provided through the EBA program with the use of those titles that would have been selected by the approval plan might provide a better understanding of the role of the traditional approval plan in the modern acquisitions' environment.

The increasing availability of ebooks and the increasing comfort of the academic community with electronic formats is a big part of the success of these changes. In the 2023 dataset, ebook purchases increased by 26 percent, including areas in which there was no change to the approval plan and areas in which the change did not directly involve a format preference. Publishers seem to be providing more books in electronic format and providing those titles more simultaneously with print releases. This publishing shift, combined with the increased usage of ebooks, points to a larger external trend and suggests that review of format preferences with shifts toward ebooks is essential for libraries seeking to improve automatic purchasing strategies of any kind.

Acknowledgments

We thank Gabby Cochran, former graduate research assistant, for compiling early-stage data and contributing to the methodology section during the initial phase of this project. We thank Tyler

Martindale, former collections data analyst, for help assembling the preliminary 2023 dataset and helping to conceptualize the comparison.

CRediT Author Contribution Statement

Jennifer A. Mezick: Conceptualization, Investigation, Project Administration, Writing – Original Draft, and Writing – Review & Editing.

Louis T. Becker: Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft, and Writing – Review & Editing.

Kat Brooks: Writing – Original Draft and Writing – Review & Editing.

Notes

1. Just-in-case is a collection development model where materials are acquired in anticipation of meeting a potential need. Described in Marilyn Deegan and Simon Tanner, "Collection Development: Just in Case, or Just in Time?" in *Digital Futures: Strategies for the Information Age* (Facet Publishing, 2013), 63–4.
2. Jennifer Mezick and Louis Becker, "Using Data to Rightsize Approval Spending," *Against the Grain* 32, no. 1 (2020): 16–18, https://against-the-grain.com/2020/04/v321-using-data-to-rightsize-approval-spending/.
3. Sarah Tudesco, Julie Linden, and Daniel Dollar, "Do We Approve? New Models for Assessing Approval Plans," *Proceedings of the Charleston Library Conference* (2016), http://dx.doi.org/10.5703/1288284316445; Victoria Koger and Virginia Kay Williams, "Evaluating Purchase Plans for Niche Collecting Areas," *Library Resources & Technical Services* 67, no. 4 (2023): 113, https://doi.org/10.5860/lrts.67n4.113; Kelly A. McCusker and Molly W. Rainard, "Too Broad and Too Narrow: One Library's Experience with Approval Plans," *Library Resources & Technical Services* 69, no. 2 (2025), https://doi.org/10.5860/lrts.69n2.8434; Wenli Gao, Cherie Turner, and Irene Ke, "Programming Plus Subject Expertise: A Combined Approach for Approval Profile Modification," *Collection Management* 43, no. 2 (2018): 89–100, https://doi.org/10.1080/01462679.2018.1438950.
4. Ashlynn Kogut, Laura D'Aveta, and Simona Tabacaru, "Assessment of Acquisition Methods for a Juvenile Literature Collection at a Research University," *Collection Management*, 47 (2022): 87–100, https://doi.org/10.1080/01462679.2021.1919270; Molly Strothmann and Karen Rupp-Serrano, "A Comparative Analysis of Evidence-Based Selection, Professional Selection, and Selection," *Library Resources & Technical Services* 64, no. 1 (2020), https://doi.org/10.5860/lrts.64n1.15; Sarah Tudesco, Julie Linden, and Daniel Dollar, "Do We Approve? New Models for Assessing Approval Plans," *Proceedings of the Charleston Library Conference* (2016), http://dx.doi.org/10.5703/1288284316445.
5. Ramirose Ilene Attebury, "Are Patrons Using the Research Books? A New Approach to Approval Plan Assessment," *Collection and Curation* 41, no. 2 (2022): 47–9, https://doi.org/10.1108/CC-01-2021-0002; Andrei Mihailovic, "The State of the Approval Plan in the Wake of the COVID-19 Pandemic," *Technical Services Quarterly* 39, no. 4 (2022): 412–27, https://doi.org/10.1080/07317131.2022.2125678.
6. Ramirose Ilene Attebury, "Are Patrons Using the Research Books? A New Approach to Approval Plan Assessment," *Collection and Curation* 41, no. 2 (2022): 47–49, https://doi.org/10.1108/CC-01-2021-0002; Denise Brush, "Circulation Analysis of an Engineering Monograph Approval Plan," *Collection Building* 26,

no. 2 (2007): 59–62, https://doi.org/10.1108/01604950710742095; David C. Tyler, Christina Falci, Joyce C. Melvin, MaryLou Epp, and Anita M. Kreps, "Patron-Driven Acquisition and Circulation at an Academic Library: Interaction Effects and Circulation Performance of Print Books Acquired via Librarians' Orders, Approval Plans, and Patrons' Interlibrary Loan Requests," *Collection Management* 38, no. 1 (2013): 3–32, https://doi.org/10.1080/01462679.2012.730494.

7. Jan C. Horner, "E-Preferred Approval Books at the University of Manitoba: A Comparison of Print and Ebook Usage," *Evidence Based Library and Information Practice* 12, no. 2 (2017): 90–105, https://doi.org/10.18438/B8BT04; Simona Tabacaru, "E-Preferred Approval Plan at a Large Academic Library: Assessment and Collection Development Implications," *Collection Management* 43, no. 3 (2018): 155–176, https://doi.org/10.1080/01462679.2018.1459328.
8. Amy Fry, "Conventional Wisdom or Faulty Logic? The Recent Literature on Monograph Use and E-book Acquisition," *Library Philosophy and Practice* (2015), http://digitalcommons.unl.edu/libphilprac/1307.
9. "Collection Development for Academic Libraries," EBSCO Information Services, accessed May 6, 2024, www.ebsco.com/academic-libraries/products/gobi-library-solutions/collection-development.
10. T. D. V. Swinscow, "Differences between Percentages and Paired Alternatives," in *Statistics at Square One*, 9th ed. (BMJ Publishing Group, 1997), www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/6-differences-between-percentages-and.
11. Marilyn Christianson, "Patterns of Use of Electronic Books," *Library Collections, Acquisitions, and Technical Services* 29, no. 4 (2005): 351–63, https://doi.org/10.1080/14649055.2005.10766084.
12. Rachel Done, Rylee Warner, and Rachel Noorda, "Publishing Distribution Practices: New Insights About Eco-Friendly Publishing, Sustainable Printing and Returns, and Cost-Effective Delivery in the U.S.," *Publishing Research Quarterly* 38, no. 2 (2022): 364–381, https://doi.org/10.1007/s12109-022-09882-5.
13. Because the study generally considers books actually purchased during the two time periods, and there is naturally a delay between the profiling work done by GOBI and the final registration of a purchase in Alma, these cannot be seen as definite counts of the population of books from which our selections were made.
14. Sandra De Groote and Jung Mi Scoulas, "Impact of COVID-19 on the Use of the Academic Library," *Reference Services Review* 49, no. 3/4 (2021): 281–301, https://doi.org/10.1108/RSR-07-2021-0043; Erin Owens, Soo-yeon Hwang, Dianna Kim, Tyler Manolovitz, and Lisa Shen, "Do You Love Them Now? Use and Non-Use of Academic Ebooks a Decade Later," *The Journal of Academic Librarianship* 49, no. 3 (2023), https://doi.org/10.1016/j.acalib.2023.102703. Groote and Scoulas (2021) found increasing reports of ebook usage through the pandemic, whereas Owens et al. (2023) noted that increased use of ebooks did not necessarily indicate an increasing preference for the format in the academic context.

NOTES ON OPERATIONS

Ebook Call Number Project

Automating Library of Congress Classification Numbers to Ebooks for Reporting

Bonita Pollock, Zhao Liu, Amanda Yesilbas, Laura Pascual, and Skyler Schultz

The Ebook Call Number Project at the University of South Florida Libraries automated the process of adding Library of Congress (LC) Classification numbers into call number fields, thereby enhancing ebook metadata while minimizing manual work. This initiative aimed to improve the accuracy of analytics reporting for discipline-specific collection assessment and reporting. The project involved analyzing existing metadata, developing a Python-based automation tool using OCLC's Connexion API, and implementing a systematic workflow within the Ex Libris Alma Library Management System for cataloging and quality control. The effort increased the number of ebook records with LC call numbers from 91 percent to 99 percent, significantly enhancing reporting accuracy and efficiency. This article details the project's rationale, methodology, and outcomes, demonstrating how automation and artificial intelligence can streamline metadata enrichment in academic libraries. Finally, the article presents an alternative method for smaller institutions that lack access to programmers or programming solutions and discusses future applications.

Introduction

The University of South Florida (USF) Libraries completed an extensive project that enhanced ebook metadata by inserting Library of Congress (LC) Classification (LCC) into the call number field, allowing staff to produce more accurate and detailed analytical reports on the USF Libraries' collections by subject area. The Metadata/Cataloging Unit of the Collections & Discovery (C&D) Department needed to analyze the electronic resource records and determine how many vendor records contained LC call numbers. The unit devised a plan to incorporate LC call numbers into as many electronic records as possible to improve the results of Alma Analytics reports. This article discusses why USF Libraries chose to undertake this enhancement project and details the steps involved, including planning, implementation, and outcomes.

Bonita Pollock (pollockb1@usf.edu) is Associate Director, Collections & Discovery, University of South Florida Libraries, https://orcid.org/0000-0001-5500-6862. **Zhao Liu** (zhaoliu@usf.edu) is Electronic Resources Catalog Librarian, University of South Florida Libraries, https://orcid.org/0009-0000-0685-7193. **Amanda Yesilbas** (ayesilbas@usf.edu) is E-Resources Librarian, University of South Florida Libraries, https://orcid.org/0009-0007-2348-2312. **Laura Pascual** (lcpascua@usf.edu) is Director Collections & Discovery, University of South Florida Libraries, https://orcid.org/0000-0002-9565-4466. **Skyler Schultz** (skyler21@usf.edu) is Complex Copy Cataloger, University of South Florida Libraries, https://orcid.org/0009-0009-0075-0227.

Background

The USF is a large public R1 research institution with nearly 50,000 students. Under a single accreditation, the USF has locations in Tampa, St. Petersburg, and Sarasota. USF is a metropolitan institution, integrated with the local community, and diverse. In 2023, the USF joined the prestigious Association of American Universities (AAU). The USF Libraries are composed of the Tampa Campus Library, the Nelson Poynter Library in St. Petersburg, and the Library Services and Information Commons on the Sarasota campus. The University Health Libraries consists of two libraries: the Shimberg Health Sciences Library on the Tampa campus and the Florida Blue Health Knowledge Exchange in downtown Tampa.

In 2021, the twelve universities in the State University System (SUS) and the twenty-eight state colleges migrated to the Ex Libris Alma library management system and the Primo VE discovery tool. On a consortial level, the library system and discovery tool are managed by the consortium; however, the USF Libraries C&D Department manages the bibliographic content and the discovery interface for all USF Libraries. Cataloging practices are coordinated across all campuses.

The Tampa Campus Library has adopted an electronic-first acquisition strategy since 2009. It directly manages five evidence-based ebook acquisition programs and one very large ebook data-driven acquisitions program. Most electronic content is available across all USF campuses and colleges through a single shared authentication process with common licensing. The Tampa Library is working to make better use of Alma functionality, using it as a repository for all collections-related data to provide a cohesive source for reporting. If Alma has good data, good reporting can be developed from it. This includes using Alma Acquisitions for purchase information and SUSHI harvesting to Alma for COUNTER R5 usage report data.

It is essential for all academic libraries to report and analyze their content, whether for developing and managing the collection, demonstrating the value of the library's collections, providing subject-area support information, preparing accreditation reports, developing new program proposals, securing grant support, or benchmarking, among other purposes. Developing an automated, standardized reporting process requires a consistent, reliable data source with key metrics. This case study specifically addresses the need to report USF Libraries' electronic collections metrics on inventory and usage by subject area, using the LCC in the MARC bibliographic call number fields.

Aside from reporting for collection management, USF Libraries need to report externally for a variety of purposes, including those that relate to demonstrating the value of the library's collections, for letting colleges and departments know what content the library has that supports a given discipline, for college program accreditations, for grant support, and for benchmarking. Reporting required at the USF Libraries included the Integrated Postsecondary Education Data System (IPEDS) and the Association of College and Research Libraries (ACRL), as well as the library support sections of program accreditation applications and new degree programs. Because the basis for most of this reporting is the subject area or discipline, the LCC code data fields—using standardized alphanumeric coding—are an excellent way

to report content by subject. The USF Libraries prepare annual reports that include specific assessments of collections and provide data to meet the USF's accreditation requirements. One such document, the "Report Card," was devised to report library collection content and usage data by discipline.

Literature Review

The LCC system functions as a foundational organizational and analytical structure for academic library collections, shaping collection development practices for decades. Historically, LCC enabled libraries to describe, measure, and evaluate their print holdings through standardized subject ranges and subclass alignments, thereby providing the basis for quantitative and qualitative assessment. Early methodologies such as the RLG Conspectus and the North American Collections Inventory Project relied on LCC's disciplinary structure to determine collecting strengths, define peer comparisons, and assess institutional priorities.[1] Contemporary iterations, including OCLC's Conspectus-based WorldShare Collection Evaluation tools, continue to employ LCC ranges to support benchmarking and historical analysis.[2]

LCC has also been central to quantitative measurement approaches that use call number boundaries to define countable segments of the collection. Shelf-list counts and comparative subject inventories are widely used to determine relative subject strengths, collecting intensity, and collection gaps.[3] Targeted evaluation practices, including the Brief Test methodology, similarly rely on LCC as a validating mechanism to define the scope of a subject study and exclude materials that fall outside the designated subclass range.[4] Together, these approaches demonstrate LCC's long-standing role as a shared analytical language for academic library assessment.

Beyond counting and benchmarking, LCC supports the analysis of usage and curricular content. Circulation statistics aggregated by call number ranges allow researchers to evaluate subject demand, identify fields that are overused or underused, and inform purchasing decisions.[5] This granular usage mapping clarifies how collections interact with institutional teaching and research needs, particularly when paired with curricular mapping based on LCC subclasses.[6] These methods provide evidence that call number segmentation produces focused, discipline-specific insights that broad subject labels alone cannot supply.

Researchers acknowledge, however, that the structure of LCC also introduces persistent assessment barriers. Because LCC organizes knowledge primarily by discipline, interdisciplinary subjects can become fragmented across multiple classification ranges, thereby complicating systematic evaluation. Research on Women's Studies collections, for example, illustrates how LCC distributes content across subclasses, hindering accurate analysis and complicating comparative evaluation.[7] This core structural constraint continues to limit the precision of LCC-based assessment even in the digital era.

The growth of electronic monograph collections has expanded LCC's operational scope. Although ebooks are not physically shelved, librarians increasingly rely on call numbers as metadata to manage digital content, map subject overlap, and support the development of user interfaces.[8] Tools such as the Hierarchical Interface to Library of Congress Classification (HILCC) demonstrate how call number data can support browsable classification systems for digital collections.[9] Large-scale ebook mapping

projects, such as those at the University of North Texas, have implemented LCC ranges to reflect overlapping and interdisciplinary relationships across digital packages.[10] These developments suggest that LCC retains value not only as a structural classification scheme for print formats but also as a metadata framework for digital resource discovery and assessment.

At the same time, LCC-based ebook assessment introduces notable challenges. Many electronic records lack call numbers because cataloging standards historically did not require classification for nonphysical formats.[11] Retrospective call number assignment and metadata enhancement require substantial labor investment, and topic-mapping techniques that connect LCC subclasses to subject categories demand continuous refinement.[12] Moreover, the disciplinary bias that fragments interdisciplinary subjects in print collections remains equally problematic in the digital environment.[13] Nevertheless, ebook usage studies show the growing value of LCC for aggregating platform-specific data—such as section requests and page views—into meaningful subject-level analytics.[14]

Several studies have examined the automation of the creation of LCC in MARC records. In the White Paper "Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper," authors Suzanne Kemperman et al. indicated that automation in bibliographic data management plays an important role in enhancing efficiency and accuracy.[15] According to a study conducted by Isidoro Gil-Leiva et al. on South American libraries, almost 59 percent of the respondents surveyed felt that using automation to assign classification codes was appropriate.[16] This research provided the team with strong motivation to develop an automated process for creating LC call numbers in catalog records.

Taken together, the literature indicates that LCC continues to function as a relevant and adaptable framework for academic collection assessment. Although its disciplinary structure and metadata gaps pose methodological limitations, LCC remains a critical tool for subject-based analysis, curricular alignment, usage interpretation, and digital collection mapping. As electronic monographs grow in scale and complexity, LCC's metadata utility positions the system not as a relic of print organization but as an evolving data structure capable of supporting evidence-based collection development in hybrid print-digital environments.

Report Cards

In the spring of 2023, the C&D Department at the USF Libraries was asked by library administration to create collection report cards for the USF Libraries' annual report. These report cards allowed USF Libraries to communicate which library resources are available in specific disciplines to departments across the campuses. Library administration wanted the report cards to be a one-page snapshot of subject-level collections and usage data, aligned with academic disciplines and colleges/departments. Each report card focused on a specific discipline at the USF and highlighted the libraries' resources in print and electronic formats for that subject area, according to the LCC. Therefore, inventory by LC call number and usage/circulation by LCC for print and electronic books and journals needed to be in Alma. Initially, no working reports were found, and it did not appear that LC call numbers were present on a substantial portion of the records in critical discipline areas.

The decision to use the LCC to categorize library resources was twofold. First, LCC is more streamlined and easier to manage than LC Subject Headings. LC Subject Headings exhibit greater variation and combinations, making subject heading classification and searching more difficult. In addition, in her book *Managing E-Book Metadata in Academic Libraries: Taming the Tiger*, Donna E. Frederick stated that classification should be included in the ebook metadata process to assist patrons with selection by topic when using electronic shelf listing.[17] Second, Alma Analytics already includes LCC in its reporting, making it more convenient to use when creating custom reports. The Ebook Call Number Project Team decided to focus on LCC for ebooks first and then branch out to e-journals for a future project.

In the first year, report cards were created manually. This manual process was developed using vendor title lists or entitlements reports, vendor MARC records, Alma electronic collection exports, COUNTER reports, Alma circulation data, and physical item counts. It was difficult to determine the LCCs from the vendor inventory title lists because many appeared to be missing. It was also difficult to then manually match those titles to the COUNTER reports to get COUNTER usage by LCC. Exported Alma electronic collections for book and journal platforms also appeared to be missing many LC call numbers. With this manual process for ebooks and e-journals, the data was limited to major vendors with COUNTER reports pertinent to the subject area. This informed the goal of improving the data in Alma to enable more automated report card production. Achieving this goal and facilitating future report card creation required enhancing all electronic records in Alma with SUSHI COUNTER statistics and LCC numbers.

The top section of figure 1 contains Program Area data that is broader and created from keywords in subject searches using publication date as a limiter. This yields a year-over-year comparison of inventory only. The middle and bottom left sections of the Snapshot of Subject Area Holdings portion of figure 1 are created from inventories by LCC, however, and include usage data. Figure 2 is a close-up of the lower left section of the report card highlighting the Print Inventory by Subject Area and the Electronic Inventory by Subject Area with numbers of titles with usage and usage for books and journals for the selected LC subclasses. This project concerns the Electronic Inventory by Subject Area section. The pie charts in figure 1 were derived from this data.

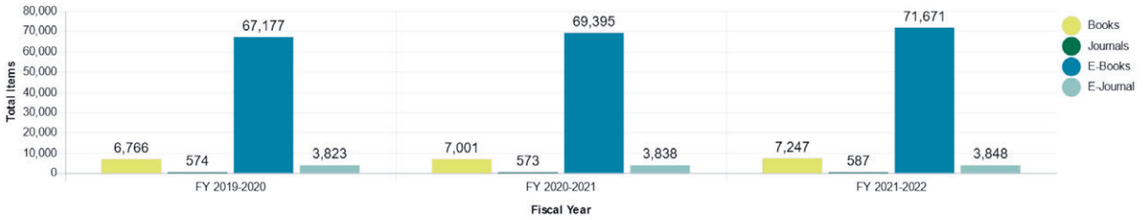
Project Overview

In the summer of 2023, the Metadata/Cataloging Unit created an Ebook Call Number Project Team and tasked the group with improving the process of creating reports via enhancing the electronic records. Although many members of the C&D Department provided input, the core team consisted of nine members: the head of metadata/cataloging, an e-resource catalog librarian, an electronic resources librarian, an applications developer programmer, a metadata specialist, a metadata quality-control manager, and three complex catalogers. The bulk of the work was completed by four key members who organized the project, ran the workflows, and recorded results. This extensive project started in the summer of 2023 and concluded in the spring of 2024. The team spent three months planning, longer than six months in implementation, and two months building reports.

ENGINEERING

TRENDS OVER 3 FISCAL YEARS (FY 2019-2022)

Trend of Format Holdings by Program Area Per FY



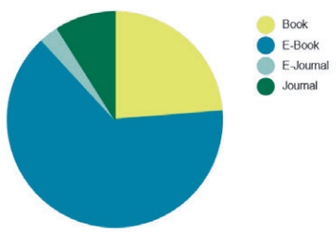
Program Area	Books	E-Books	Journals	E-Journal
FY 2019-2020	**6,766**	**67,177**	**574**	**3,823**
Engineering	5,446	46,768	393	2,703
Civil Engineering	193	1,575	29	180
Mechanical Engineering	230	3,757	42	223
Electrical Engineering	142	6,642	50	289
Materials Engineering	369	4,034	26	213
Chemical Engineering	260	2,059	33	180
Management Systems Engineering	126	2,342	1	35

Program Area	Books	E-Books	Journals	E-Journal
FY 2020-2021	**7,001**	**69,395**	**573**	**3,838**
Engineering	5,575	48,300	393	2,738
Civil Engineering	188	1,669	29	181
Mechanical Engineering	260	3,914	42	224
Electrical Engineering	142	6,747	42	262
Materials Engineering	384	4,227	26	218
Chemical Engineering	281	2,139	33	184
Management Systems Engineering	171	2,399	8	31

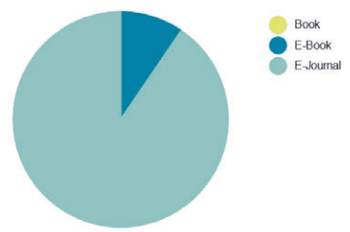
Program Area	Books	E-Books	Journals	E-Journal
FY 2021-2022	**7,247**	**71,671**	**587**	**3,848**
Engineering	5,656	49,795	393	2,746
Civil Engineering	229	1,747	29	181
Mechanical Engineering	274	4,088	42	225
Electrical Engineering	163	6,847	42	262
Materials Engineering	408	4,461	40	218
Chemical Engineering	383	2,251	33	185
Management Systems Engineering	134	2,482	8	31

SNAPSHOT OF SUBJECT AREA HOLDINGS (FY 2021-2022)

Collection Holdings by Format



Collection Usage by Format **



High Impact Journals

Journal Title	Impact Factor	USF Holding
Energy & Environmental Science	39.714	☑
Progress in Energy and Combustion Science	35.339	☑
Nature Electronics	33.255	☐
Nature Biomedical Engineering	29.234	☐
Applied Catalysis B-Environmental	24.319	☑
IEEE Transactions on Pattern Analysis and Machine Intelligence	24.314	☑
Advances in Applied Mechanics	17.400	☑
Bioactive Materials	16.874	☑
Chemical Engineering Journal	16.744	☑
Biomaterials Research	15.863	☑
BIOMATERIALS	15.304	☑
IEEE Signal Processing Magazine	15.204	☑
Proceedings of the IEEE	14.910	☑
npj Regenerative Medicine	14.404	☑
IEEE Transactions on Neural Networks and Learning Systems	14.255	☑
Journal of Hazardous Materials	14.224	☑
Engineering	12.834	☑
Journal of Industrial Information Integration	11.718	☑
Additive Manufacturing	11.632	☑
Automation in Construction	10.517	☑

Print Inventory by Subject Area

Subject	Print: Inventory	Print: Uses	Print: Percent Use	Print: Titles Used	Journal: Inventory
General & Civil Engineering	3,786	65	22%	55	2,255
Environmental Engineering	2,461	42	14%	25	1,058
Mechanical Engineering	1,381	40	14%	26	384
Electrical/Computer Engineering	5,523	97	33%	64	1,123
Chemical & Materials Engineering	1,592	35	12%	34	1,047
Industrial/Manufacturing Systems Engineering	1,959	14	5%	10	432
TOTAL:	**16,702**	**293**	**100%**	**214**	**6,299**

Electronic Inventory by Subject Area

Subject	E-Book: Inventory	E-Book: Uses	E-Book: Percent Use	E-Book: Titles Used	E-Journal: Inventory	E-Journal: Uses	E-Journal: Percent Use	E-Journal: Titles Used
General & Civil Engineering	13,854	4,115	26%	1,346	404	29,759	19%	334
Environmental Engineering	3,475	1,932	12%	430	71	18,268	12%	63
Mechanical Engineering	4,602	1,980	12%	364	117	9,782	6%	96
Electrical/Computer Engineering	15,572	4,945	31%	1,340	186	11,515	7%	157
Chemical & Materials Engineering	4,298	2,201	14%	576	249	33,403	22%	208
Industrial/Manufacturing Systems Engineerin	2,277	403	3%	197	61	4,526	3%	49
Engineering - General	1,174	427	3%	88	948	47,592	31%	431
TOTAL:	**45,252**	**16,003**	**100%**	**4,341**	**2,036**	**154,845**	**100%**	**1,338**

Subject Area Databases

Database Title
Scopus
Web of Science
Science Direct
Springerlink
IEEE Xplore
ACM Digital Library
Applied Science & Technology Source
Inspec

For Collection Usage by Format: Usage counts are not available for print journals. For Inventories: Subject areas differ by format. High Impact Journal data was collected from Journal Citation Reports. ** For the "Collection Usage by Format" pie chart: the small numerical value for the "Book" format wedge makes it not visible. Please consult the inventories for value comparisons.

Figure 1. Example of a report card for engineering.

Print Inventory by Subject Area

Subject	Print: Inventory	Print: Uses	Print: Percent Use	Print: Titles Used	Journal: Inventory
General & Civil Engineering	3,786	65	22%	55	2,255
Environmental Engineering	2,461	42	14%	25	1,058
Mechanical Engineering	1,381	40	14%	26	384
Electrical/Computer Engineering	5,523	97	33%	64	1,123
Chemical & Materials Engineering	1,592	35	12%	34	1,047
Industrial/Manufacturing Systems Engineering	1,959	14	5%	10	432
TOTAL:	**16,702**	**293**	**100%**	**214**	**6,299**

Electronic Inventory by Subject Area

Subject	E-Book: Inventory	E-Book: Uses	E-Book: Percent Use	E-Book: Titles Used	E-Journal: Inventory	E-Journal: Uses	E-Journal: Percent Use	E-Journal: Titles Used
General & Civil Engineering	13,854	4,115	26%	1,346	404	29,759	19%	334
Environmental Engineering	3,475	1,932	12%	430	71	18,268	12%	63
Mechanical Engineering	4,602	1,980	12%	364	117	9,782	6%	96
Electrical/Computer Engineering	15,572	4,945	31%	1,340	186	11,515	7%	157
Chemical & Materials Engineering	4,298	2,201	14%	576	249	33,403	22%	208
Industrial/Manufacturing Systems Engineering	2,277	403	3%	197	61	4,526	3%	49
Engineering - General	1,174	427	3%	88	948	47,592	31%	431
TOTAL:	**45,252**	**16,003**	**100%**	**4,341**	**2,036**	**154,845**	**100%**	**1,338**

Figure 2. Report card data sheet.

Action Plan

The team developed an action plan for the Ebook Call Number Project that included collecting call number data for electronic resources in Alma, experimenting with methods to add call numbers to records in Alma in batches, exploring ways to retrieve and insert call numbers into records, developing an automation workflow for completing the call number retrieval and insertion, collecting statistics on the project, and creating custom analytics reports to make report cards.

Phase one of the action plan was to collect data from Alma, the Library Services Platform. A team member conducted an extensive survey of all the electronic ebook collections activated in Alma for the University. The collections were exported as MARC records and analyzed in MarcEdit. Field counts were tracked for LC call number, author, title, and subject across vendors and ebook collections in an Excel spreadsheet. Figure 3 shows a sample tracking spreadsheet.

The analysis revealed that in the spring of 2023, USF Libraries had seventy-five ebook collections activated in Alma's Community Zone (CZ) from twenty-eight different vendors. Filters were used to exclude primary source materials, government documents, and other material types, leaving only books. Approximately eighteen of the electronic resource vendors already provide LCC call numbers on 90 percent or more of their records. These collections consisted mainly of resources from major ebook vendors. Figure 4 shows vendor counts highlighting the percentage of call numbers present in the MARC records.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Vendor	Collection	Alma Electronic Collection Name	Portfolio Count	MARC Total Count(from MarcEdit)	Call Number Percentage 050/total	035	050	090	100	110	111	130	245	264	505	520	650	830	Notes
26	ACLS	ACLS Humanities ebook	ACLS Humanities eBook	5962	5924	87.68%	5924	5194	2128	5264	33	9	37	5924	2565	2961	2389	5357	3372	
27	ABC-Clio	ABC-Clio ebooks	ABC-CLIO ebooks	339	339	87.91%	339	298	0	269	0	0	0	339	78	243	186	303	110	
28	Cambridge	Cambridge EBA	Cambridge EBA (USF)	47283	47283	89.83%	47283	42474	338	35535	96	333	93	47283	47257	23146	47126	39125	22141	5/5/2(
29	EBSCO	owned	EBSCOhost Ebooks	68708	68701	89.96%	68701	61803	1016	51308	2232	426	140	68701	29245	39114	35773	59750	16312	no marc record for collectio
30	Project Muse	owned	Project Muse All Books	6160	6149	92.03%	6149	5659	54	5090	20	25	24	6149	4609	3970	4401	5773	2771	no marc record for collectio
31	LOEB		Loeb Classical Library	641	292	92.12%	292	269	0	256	0	0	5	292	80	152	154	202	272	no marc record for collectio
32	JSTOR	ebooks owned	JSTOR Books	6514	6365	92.38%	6365	5880	7	5427	34	49	24	6365	5274	3107	4883	6020	2493	no marc record for collectio
33	Brill	Brill EBA	Brill Ebooks EBA (Brill - All E-Book Titles)	27172	27155	93.32%	27155	25342	115	16731	442	323	145	27155	20917	21537	22290	23693	21433	5/5/2(
34	Wiley	Online Library ebooks	Wiley Online Library - AutoHoldings Books	6349	6349	93.38%	6349	5929	209	3292	88	74	55	6349	3836	4538	3732	5879	2157	
35	Knovel	Knovel Civil Engineering & Construction Materials Academic	Knovel Civil Engineering & Construction Materials Academic	1108	1108	93.68%	1108	1038	6	561	302	96	4	1108	926	411	638	1065	290	
36	Talor & Francis	Owned/EBA Autoload	Taylor & Francis eBooks Complete	36288	35610	94.24%	35610	33559	208	23921	108	140	107	35610	32661	19337	22710	33342	15532	5/5/2(
37	Brill	Brill EBA	EBA Brill All	22890	22878	94.68%	22878	21661	49	13968	425	264	145	22878	18077	18352	18295	19785	19386	5/5/2(
38	Ebook Central	Owned	Ebook Central Perpetual, DDA and Subscriptiion Titles	246012	243581	94.81%	243581	230930	5945	169656	11979	5082	630	243581	160681	174940	161060	222245	100974	5/5/2(
39	Knovel	Knovel Electrical & Power Engineering Academic	Knovel Electrical & Power Engineering Academic	487	486	94.86%	486	461	15	314	62	7	3	487	336	249	260	469	164	
40	Knovel	Knovel Mechanics & Mechanical Engineering Academic	Knovel Mechanics & Mechanical Engineering Academic	1062	1062	95.10%	1062	1010	16	693	190	20	4	1062	803	413	584	1040	246	
41	Sage		Sage (USF)	110	110	95.45%	110	105	5	9	0	0	8	110	62	2	99	101	1	no marc record for collectio
42	Elsevier (Science Direct platform)	Elsevier ScienceDirect Books (Elsevier ScienceDirect Books Complete)	Elsevier ScienceDirect Book Series Pa ckage - All Subjects	5689	5689	96.59%	5689	5495	85	1308	27	30	52	5689	3124	3776	3370	5587	3078	
43	IEEE Xplore platform	IEEE Xplore Wiley	IEEE Xplore Wiley eBooks Library (IEEE Xplore)	1175	1159	96.81%	1159	1122	8	878	0	2	1	1159	1145	1069	985	1125	559	

Figure 3. Data collection spreadsheet: analysis of ebook call numbers.

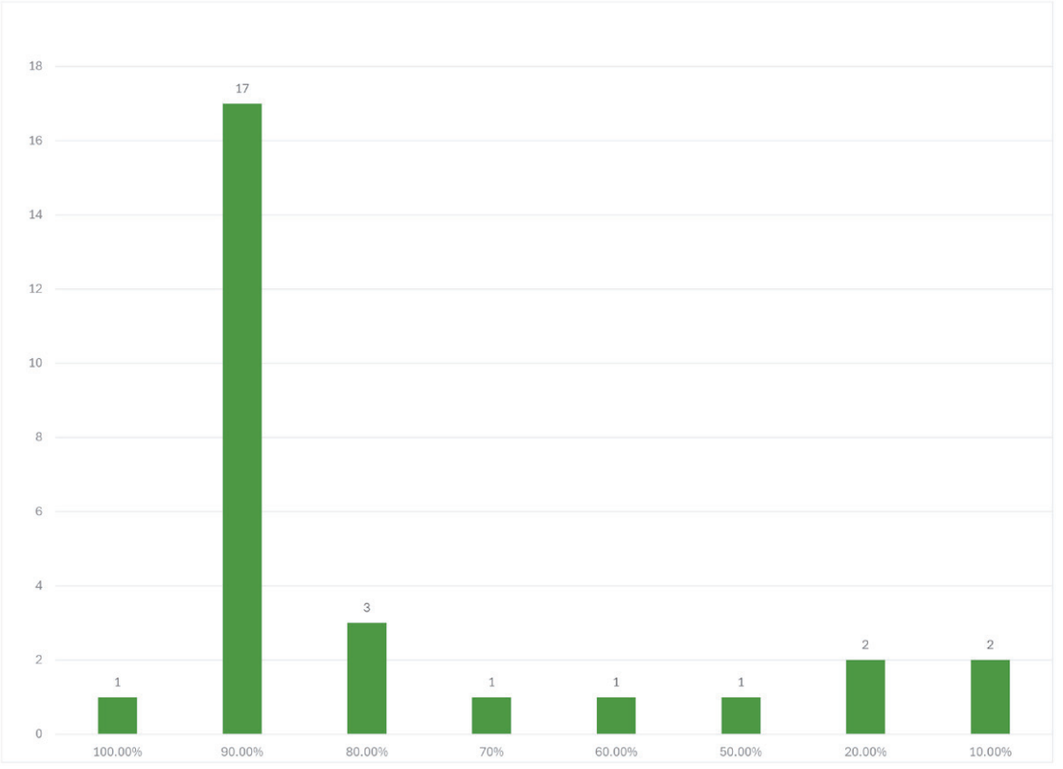


Figure 4. Vendor counts by call number percentage.

The remaining ten vendors were mostly smaller or specialized publishers. These vendor collections required more call number enhancements. A total of more than 60,000 ebook records without LC call numbers required remediation in the catalog. These records represented critical academic disciplines,

including foreign language materials, electrical engineering, psychology, archeology, medieval history, reference materials, and music and drama collections. It was important for USF Libraries to represent these disciplines accurately in the report cards. Assigning call numbers to these collections added clarity and completeness while improving the accuracy of analytics call number reporting, including those used on the report cards. Thus, staff time and effort invested in this project yielded valuable results.

The Ebook Call Number Team proceeded to the second phase of the action plan after determining that it would be useful for USF Libraries to add LCC to all ebooks. In this phase, the team experimented with various methods to batch add call numbers to MARC records in Alma. Most of the USF Libraries' electronic resources were activated in Alma's CZ, creating an Institution Zone (IZ) record linked to the CZ record for local use. Updating the CZ records themselves is problematic for individual institutions because Ex Libris maintains the CZ records with metadata supplied by the various vendors. Institutions do have the ability to add certain local fields to the CZ record, which creates the local fields in the IZ linked record. The local field must include special coding (local extension) to prevent the call number from being overwritten when Ex Libris updates metadata in the CZ. The team elected to use the 090 field with a local extension to insert call numbers into the IZ copies of the CZ records. Inserting 60,000 LC call numbers manually was not feasible. The Ebook Call Number Team experimented with two approaches to batch-processing LC call numbers into Alma. One option was to create a set and apply normalization and indication rules to insert the same call number via a job. A second option involved retrieving MARC records with the proper call number and overlaying them into Alma using an import profile. The team ultimately used both processes throughout the project. The second process (overlay) worked well during the automation phase, and the first process (normalization and indication rules) was useful during the cataloging quality-control phase.

In the third phase of the action plan, the team needed to devise a method to automate call number retrieval. Processes explored included using a macro to scrape call numbers from WorldCat, using OCLC Classify, the MarcEdit/Alma API, OpenRefine, the OCLC Connexion API, or ChatGPT. The macro process worked well, but it retrieved only the call number, not the whole MARC record. OCLC Classify created a call number from the OCLC subject heading, but OCLC discontinued this plug-in. Although OpenRefine is excellent for cleaning metadata and provides APIs for data retrieval, the team was unable to retrieve OCLC call numbers in batch. MarcEdit, an excellent cataloging tool, has an API for retrieving records from Alma, which could then be updated in MarcEdit and returned to Alma. Because MarcEdit's call number plug-in was tied to the discontinued OCLC Classify, the team could not use this tool.

The Connexion API had the greatest potential for the call number project because it enables OCLC searches that return either specific fields or complete MARC records. With assistance from the C&D Department's programmer, the team used this feature to retrieve call numbers, insert the local 090 extensions, and then reload the records into Alma. Because the USF Libraries are an OCLC member, it is permissible to retrieve OCLC records and call numbers via API and use them in the IZ of Alma. Institutions without a programmer on staff may find this solution difficult because it requires extensive programming expertise to write API data retrieval code. The team lastly explored using ChatGPT to

retrieve call numbers. Although ChatGPT had great potential for future projects, the team ultimately decided not to use it for this project due to the extensive training required to teach ChatGPT the LCC system.

The Metadata/Cataloging Unit collected statistics on the Ebook Call Number Project throughout the project to assess its success, evaluate its effectiveness, and inform improvements to the workflow for future endeavors. Each team member recorded consistent, identifiable metrics for each task and for every step within that task. Statistics collected at the beginning of the project indicated the number of ebooks in Alma with call numbers and which collections required remediation. Statistics collected during the project indicated the number of call numbers successfully corrected by the automation process, the number requiring manual updates, and the number requiring manual correction by catalogers. Statistics collected at the end of the project determined the total number of ebooks that now had an LC call number and the number that still required one. These statistics proved invaluable to the team for monitoring progress and assessing the success of the automation process. Figure 5 shows the project's statistics tracking spreadsheet.

Implementation Workflow

Once the action plan phases were completed, the gathered information was used to develop an implementation workflow for this project. The workflow consists of the following processes: record set creation, automation, quality control, and report creation.

The first process in the implementation workflow involved creating MARC record sets for ebooks that required call numbers. A team member created sets in Alma for each electronic ebook collection activated in the CZ. Each set was then exported as full MARC bibliographic records. These MARC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	# of records with call number automatically inserted	# of records with call number manually inserted	# of records with Automation Correction	# of records with 050 inserted post run	# of records NOT imported	Ready for Liang	load back to Alma	Marc records no 050 field (after removing collection marc)	Vendors	Alma Electronic Collection Name	Current Number of Portfolios Active in IZ	MARC Total Count(from MarcEdit)	Call Number Percentage 050/total	Notes
27	43	0	6			X	X	43	Bloomsbury	Bloomsbury Medieval Studies	199			
28	5	0	0			X	X	5	Bloomsbury	Bloomsbury Medieval Studies 2020-2021 Update	24			
29	4	0	0			X	X	4	Bloomsbury	Bloomsbury Medieval Studies Annual Update 2022	37			
30	3	0	0			X	X	3	Bloomsbury	Bloomsbury Medieval Studies: Core Collection - Annual Update 2023	16			
31	67	0	0			X	X	67	Bloomsbury	Bloomsbury History: Theory & Method - Launch Collection	219			
32	14	1	0			X	X	15	Bloomsbury	Manchester Medieval Sources Online (MMSO)	41			
33	802	4	802			X	X	806	Bloomsbury	Bloomsbury Drama Online Core Collection US/Canada	1167			
34	13	0	13			X	X	13	Bloomsbury	Bloomsbury Drama Online Core Collection 17-18 Update US/Canada	116			
35	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online: Core Collection 2018-2019 Update US	98			
36	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online Core Collection - 2020 Annual Update US	54			
37	10	0	10			X	X	10	Bloomsbury	Bloomsbury Drama Online: Core Collection Annual Update 2021	63			

Figure 5. Statistics tracking for call number insertion.

records were run through MarcEdit, and field counts for all call number fields were recorded; these were later used to assess the effectiveness of the automation process. The team stored this data in a project statistics spreadsheet organized by vendor and ebook collection for easy retrieval and analysis.

The second process in the workflow involved the automation process. The programmer created a program that retrieved LC call numbers from OCLC by using the Connexion API. These call numbers were inserted into a 090 local extension field. The metadata specialist created an import profile to reload MARC records into Alma with the local 090 extension field. The MARC records were then reloaded into Alma by vendor and collection. The total number of automated call numbers was recorded on the project statistics spreadsheet.

The third process of the workflow involved quality control. While the Ebook Call Number Team felt automation was the best approach, for quality-control purposes, catalogers still needed to monitor it. The complex copy catalogers checked the reloaded records for accuracy and corrected any errors in the automation. The catalogers also manually inserted call numbers into records that could not be automated. The team performed final quality control on project completion and recorded the statistics in the project statistics spreadsheet.

The final process in the workflow involved creating analytical queries for building report cards. The metadata quality-control manager used Alma Analytics to build custom reports for the C&D Department, which were in turn used to create USF Libraries report cards.

Record Set Creation Process

A team member conducted an advanced search in Alma to ensure the data was relevant and accurately reflected the intended collection. The search sought book titles in electronic collections that lacked call numbers. Figure 6 displays search parameters used in Alma.

The results were saved as a set, and then the bibliographic data from the set was exported using Alma's "Export Bibliographic Records." The downloads were too large to handle efficiently, so the files were

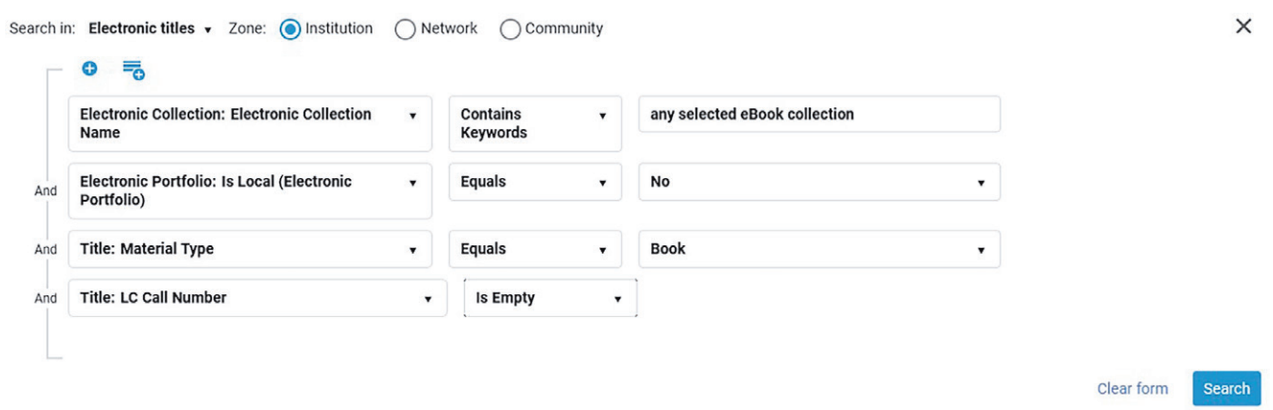


Figure 6. Advanced search in record set creation process.

parsed into smaller sets of 1,000 records. This approach improved processing times and allowed the programmer to detect errors more easily. Renaming the parsed files to a standard naming convention helped avoid issues during automated processing.

Automation Process

This section provides a comprehensive automation workflow for the Ebook Call Number Project. The automation process used Python to enhance bibliographic records. The heart of the automation process was a Python program designed to streamline the retrieval and integration of call numbers into MARC records.

The core of the automation strategy was a custom Python program.[18] Inspired by earlier efforts to use a macro for automation, the programmer developed a solution tailored to the project's specific needs. In the article "Opening the Door: A First Look at the OCLC WorldCat Metadata API," Terry Reese mentioned that OCLC WorldCat APIs support read and write access to the WorldCat database, as well as offer new functions and workflows that directly interact with the WorldCat database.[19] Using the capability described by Reese, the programmer wrote a Python program that used an API to search for bibliographic records one by one in the OCLC database, using different identifiers and prioritizing them based on their relevance and availability. Upon retrieving the call numbers from bibliographic records, the program automatically inserted them into the MARC records, specifically into the 090 field. This automation ensured that bibliographic records were accurately updated with the correct call numbers, requiring minimal manual intervention.

Before writing the Python program, the programmer needed to complete two tasks. The first one was to request a WorldShare Key at the WSKey Management platform. Obtaining this key was crucial for integrating OCLC's resources into the automation process because it provided access to two critical services: the WorldCat Search API, which enabled searching for bibliographic records in WorldCat; and the WorldCat Metadata API, which facilitated the retrieval and management of metadata from WorldCat. The second task was to prepare the programming tools. For this project, the programmer utilized PyCharm, specifically the professional version. PyCharm is a robust Integrated Development Environment (IDE) that supports Python programming with features such as code completion, debugging, and project management.[20] The program relied on several Python libraries to function effectively: Pandas 3.0.3, Pymarc 5.1.0, and BookOps-WorldCat 0.5.0. Python 3.9.8 was used to ensure compatibility with the libraries and tools.

The programming logic addressed the limitations of the OCLC API and ensured accurate retrieval of bibliographic information. The OCLC API could retrieve full bibliographic records with call numbers only when searching by OCLC number. If searches used other identifiers such as ISBN or titles, the API would return brief records containing OCLC numbers but not call numbers. The program's approach was to find the OCLC number and use it to retrieve the full MARC record and extract the first call number field found on the record.

Figure 7 displays a flowchart of the Python program. The following explains how the program processed a single record. First, the program read a MARC record and initiated a search using available identifiers. If an OCLC number was present, the program issued a follow-up request using that number to retrieve a full bibliographic record that included the call number. Once the call number was retrieved, it was automatically inserted into a local 090 field of the MARC record. If there was no OCLC number, then the program searched using other available identifiers and then retrieved brief records with OCLC numbers. The program then made a follow-up request using the OCLC numbers to fetch full bib records, which would include call numbers. Once the first call number was retrieved, it was automatically inserted into the local 090 field of the MARC record.

The Python program generated three distinct file types for each set of bibliographic records, each serving a specific purpose in the bibliographic data enhancement workflow. The program's primary output was a set of MARC records with call numbers automatically inserted into the local 090 field. These records overlayed existing records in Alma. The second output was a Comma-Separated Values (CSV) file that catalogers used for quality control, to review and correct, to assign call numbers as needed, or to make necessary adjustments. The third output consisted of program log files, which the programmer used to monitor the program's performance, check for correct execution, and identify any errors or issues.

According to Toaza and Esztergár-Kiss, the development and implementation of Python programs have significantly streamlined the bibliographic data management process.[21] By automating the retrieval and integration of call numbers using Python programming, the team enhanced accuracy, reduced manual effort, and improved overall efficiency.

The final step in the automation process was to overlay the MARC records in Alma with the primary output from the Python program. This was done using an import profile. The import profile was configured to ensure that records were updated correctly and that the new call numbers were properly incorporated into the existing bibliographic data. To optimize the import process and improve efficiency and the accuracy of statistical tracking, the 1,000-record files were combined by vendor when possible.

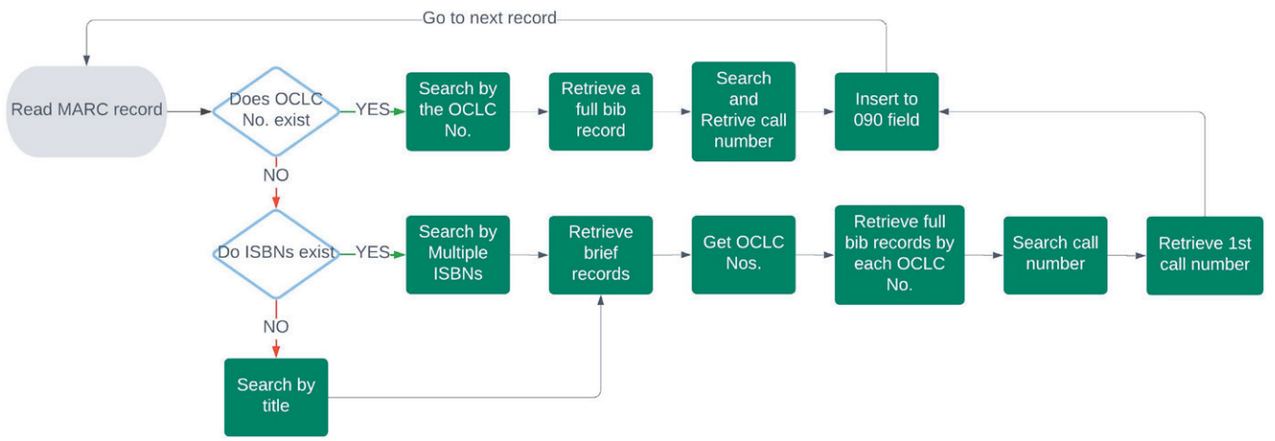
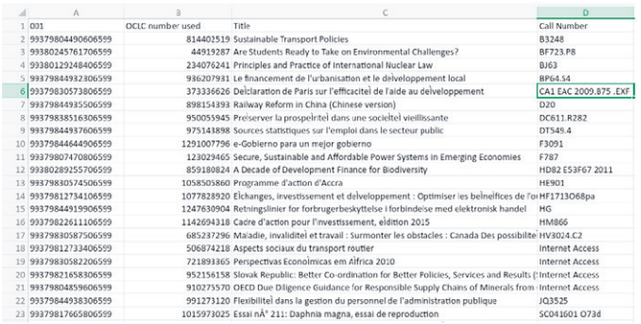


Figure 7. Customized Python program automation workflow chart.

Quality-Control Process

The third process in the Ebook Call Number Project workflow was quality control. Park and Tosaka indicate that it is essential to perform thorough quality control to ensure the accuracy of the bibliographic data.[22] The Python program, although efficient, occasionally retrieved incorrect call numbers or included irrelevant content. Catalogers manually reviewed the automated call numbers to verify that the assigned LC class was correct and relevant by matching the assigned call number to the subject headings; they corrected any discrepancies or errors identified during the review process directly in Alma. The catalogers also created call numbers for records that could not be automated.

Catalogers received reports in the form of CSV files as part of the Python program's second output. See figure 8 for an example of the fields in the quality-control report. After converting the CSV file to Excel, the catalogers verified that the retrieved call numbers were accurate.



	A	B	C	D
1	001	OCLC number used	Title	Call Number
2	9937980449060659	814402519	Sustainable Transport Policies	B3248
3	9938024576170659	44919287	Are Students Ready to Take on Environmental Challenges?	BF723.P8
4	9938012924840659	234076241	Principles and Practice of International Nuclear Law	BJ63
5	9937984493280659	936207931	Le financement de l'urbanisation et le dévelopgement local	BP64.34
6	9937983057380659	373336626	Déclaration de Paris sur l'efficacitél de l'aide au dévelopgement	CA1 EAC 2009.875 .EXF
7	9937984493550659	898154393	Railway Reform in China (Chinese version)	D20
8	9937983851630659	950055945	Préiserver la prospelrité dans une societé vieillissante	DC611.R282
9	9937984493760659	975143898	Sources statistiques sur l'emploi dans le secteur public	DT549.4
10	9937984464906659	1291007796	e-Gobierno para un mejor gobierno	F3091
11	9937980747080659	123029465	Secure, Sustainable and Affordable Power Systems in Emerging Economies	F787
12	9938028955706659	859180824	A Decade of Development Finance for Biodiversity	HD82 E53F67 2011
13	9937983057450659	1058505860	Programme d'action d'Accra	HE901
14	9937981273410659	1077828920	Elchanges, investissement et dévleoppement : Optimiser les belnelfices de l'oi	HF1713O68pa
15	9937984491990659	1247630904	Retningslinier for forbrugerbeskyttelse i forbindelse med elektronisk handel	HG
16	9937982611060659	1142694318	Cadre d'action pour l'investissement, éldition 2015	HM866
17	9937983058750659	685237296	Maladie, invaliditél et travail : Surmonter les obstacles : Canada Des possibilte	HV3024.C2
18	9937981273340659	506874218	Aspects sociaux du transport routier	Internet Access
19	9937983058220659	721893365	Perspectivas Econolmicas em Alfrica 2010	Internet Access
20	9937982165830659	952156158	Slovak Republic: Better Co-ordination for Better Policies, Services and Results (	Internet Access
21	9937980485960659	910275570	OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from	Internet Access
22	9937984493830659	991273120	Flexibilitél dans la gestion du personnel de l'administration publique	JQ3525
23	9937981766580659	1015973025	Essai nÂ° 211: Daphnia magna, essai de reproduction	SCD41601 O73d

Figure 8. Cataloger's quality-control report.

The cataloger used the report to identify and correct incorrect call numbers and to assign call numbers to items that had not received one. Although the percentage of incorrect call numbers was low, some were present. For example, if the wrong OCLC number was in Alma's CZ record, the program pulled call numbers from records that did not match the item. When no OCLC number was available, the program performed a title search, which sometimes yielded inaccurate results. Sometimes the OCLC record did not have an actual LC call number but contained phrases such as "Internet Access" or "Electronic Resource" in a call number field. Additionally, the call number field sometimes contained incomplete call numbers: it included only the alphabetical portion of the LC call number or call number ranges (e.g., AB12-56).

Depending on the number of call numbers that needed correction, the cataloger either intervened manually or ran a batch process. During manual correction, catalogers would search OCLC by title or ISBN to obtain a call number. If no record with an appropriate call number was found, the cataloger would search for subject headings in OCLC and/or the Alma CZ and use ClassWeb to determine a corresponding call number. Catalogers also used Dewey Decimal numbers (if present) and searched in ClassWeb to find an LC call number. As a last resort, the ebook itself was consulted to assign a call number for the item using subject analysis. Once a call number was identified, the cataloger would insert a 090 local extension field into the IZ record in Alma and delete any incorrect call number fields added through automation.

Typically, when more than 100 call numbers needed to be corrected or added, a batch process was used to save time. During a batch process, records with similar subject headings were assigned the same LC call number and updated in bulk using a normalization rule. Catalogers used the same procedure for determining batch call numbers as the earlier process for creating manual call numbers. Normalization

rules are queries used to modify bibliographic records in Alma. After creating a set of records in Alma, the cataloger would export the bibliographic records from Alma and use a MarcEdit tool to convert to a spreadsheet with the following fields: identifier, OCLC numbers, title, call number, and subject headings. Abrahamse wrote, "This tool can provide a useful overview of large or even not-so-large files of MARC records."[23]

The catalogers sorted the MarcEdit report by the MARC 050 field and removed records that already contained call numbers. These call numbers were most likely added to the IZ record after the team exported the original MARC record set. Next to be sorted are the local 090 fields that were incorrectly assigned by the automation process; these items were collected and converted into a set in Alma. Once the set was created, the local 090 fields were bulk deleted from the records using a normalization rule. The catalogers would then sort by subject headings, and main class call numbers were assigned based on these headings. It was not necessary to create cutters for the call numbers because the USF Libraries report cards represent twelve broad discipline areas and use only the top level of the LCC structure. This made it possible to add one general call number to the entire subject area. For example, items with a subject heading about soil moisture were assigned the LC call number S594. All items assigned the call number S594 were grouped into a single set in Alma. Before running the job that inserts the correct local 090 fields, the cataloger would edit the normalization rule associated with the job. Using the same example, the normalization rule was adjusted to insert the value S594 into the local 090 field. The normalization rule must be updated to reflect the call number to be inserted into the records before each run. Examples of the normalization rules created for this project are in figure 9.

Analytics Reports Process

With the data now enhanced in Alma through the completion of the Ebook Call Number Project and the uploading of COUNTER reports via SUSHI harvesting, the Report Card process could now be automated. The fourth process in the workflow was to build Alma Analytics reports to facilitate the creation of the USF Libraries report cards. LCC was built into Alma's reporting capabilities. This allowed the construction of analytics reports to capture inventory and usage counts for print and electronic resources by LC Class. The consortium built Alma Analytics reports by LCC for USF Libraries; the Ebook Call Number Team further customized these reports by adding LCC to the consortia-created IPEDS reports. This allowed USF Libraries to report on the print book and journal inventory counts, as well as ebook and e-journal counts, by discipline for the annual report cards. The usage statistics

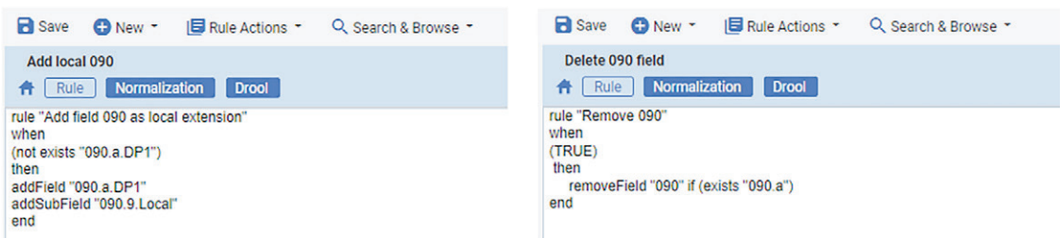


Figure 9. Examples of Alma normalization rules.

were compiled into two separate analytics reports: one for print loan usage by the LC class and one for electronic COUNTER report usage by the LC Class.

To test the reliability of the Alma Analytics–created report cards, the previous year's data from the manual report card creation was compared with the Alma Analytics–created reports of the current year; the number counts for inventory and usage were found to be very similar. The Alma Analytics report card data was also compared with vendor data and SUSHI reports to verify accuracy. This new automated process using Alma provided much more comprehensive data and generated incredible time savings in report creation.

Outcomes

The Call Number Project Team kept statistics on the various components of the project throughout the entire process. Initially, statistics were collected on the current percentage of ebooks with call numbers. During the project, the team collected statistics on the following: the number of call numbers added by automation; the number of automated call numbers requiring corrections; the number of call numbers manually added by catalogers; and the number of records with 050 fields added since the start of the project. Finally, toward the end of the project, the team ran analytics reports to determine how many ebooks had call numbers after the project and performed advanced searches in Alma to identify ebooks without call numbers for future cleanup.

Considering the ratio of call number records before and after the project, the data confirms that before the project, there were 631,129 records with call numbers and 59,040 without call numbers. After the project, there were 685,724 records with call numbers and only 4,445 without call numbers. All records included in the project results exclude primary resources, music scores, and audiovisual electronic materials because the focus of the project was ebooks. The 4,445 records without call numbers may either no longer exist or were unintentionally excluded; these records will be reconciled in future cleanup efforts. Looking at the same statistics as percentages, before the project, 90 percent of the ebook records contained call numbers, and 10 percent lacked them. After the project was complete, 99 percent of the ebook records contained call numbers, and only 1 percent lacked them.

Considering the time and labor required to insert call numbers into the local 090 field, the data indicates that of the 59,040 records with no call numbers, 44,653 call numbers were inserted through the automation process, and only 2,641 of the automated call numbers required correction. Finally, the catalogers had to enter 9,942 call numbers manually. The automation data, shown as percentages, indicates that 94 percent of automated call numbers were correct, and only 6 percent required modification. Data further shows that 78 percent of the call number insertions were automated, 5 percent needed corrections as a result of the automation process, and 17 percent required catalogers to manually insert call numbers. See figure 10 for automation statistics.

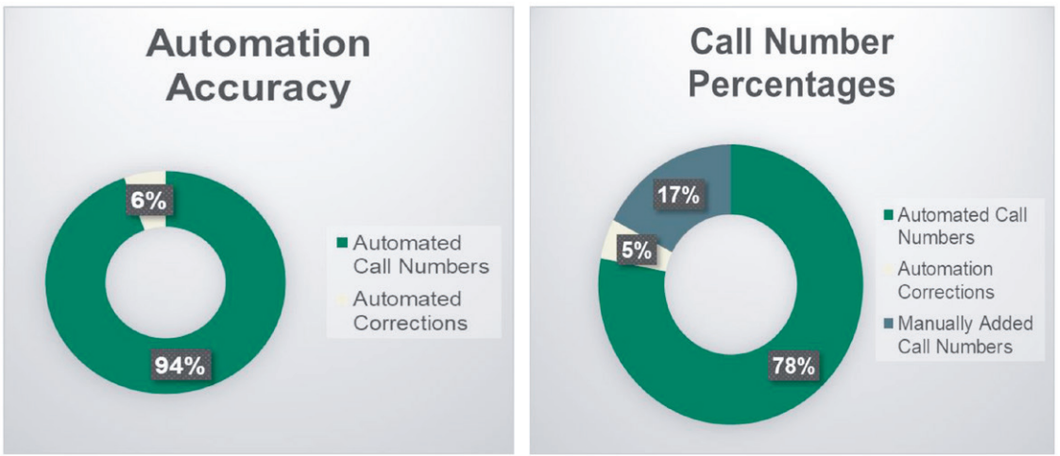


Figure 10. Statistics on the call number automation.

Conclusions

The automation process retrieved call numbers with an accuracy rate of greater than 94 percent. Not only was the Ebook Call Number Project successful, but its benefits were numerous and immediately apparent. Statistics for the annual report cards no longer had to be compiled manually from vendor websites, reducing the time required to produce the report cards. The accuracy of Alma Analytics reports improved by almost 50 percent, making the report cards more accurate. Adding LCC numbers enabled the USF Libraries to sort library resources by discipline, a key factor in accreditation reports for new programs. USF Libraries were now able to provide more comprehensive report cards to the various departments and colleges at the USF.

Lessons Learned

The Call Number Project Team identified several valuable lessons; however, five key lessons may help other institutions undertaking this project. First, due to the nature of Alma, most of the USF Libraries' electronic records are linked to the CZ. Still, some ebook collections are loaded locally and are not available in the CZ. The team did not consider this possibility when running sets at the beginning of the project and inadvertently captured some local records that were not linked to the CZ during the reload process. The import profile was set to update the CZ records with a local extension that placed the call number in our IZ linked records; accordingly, these local records would not overlay because they were not linked to the CZ and needed to be updated manually. The typical error was "Match Found but Not Imported." As noted by Chen, Kim, and Montgomery, overlaying local bibliographic records often requires specific import profiles and workflows.[24] In the future, these local records would be excluded from the record set and updated separately using a different import profile. Second, the team learned that between the time the records were exported from Alma and reloaded, some records had been removed because of CZ updates and automatic holdings updates from certain vendors. Because these records were no longer in Alma, they did not reload. The typical error message was "No Match Found."

This was also an issue noted by Rathemacher, Ragucci, and Doellinger in their article "Don't Wait, Automate! Industry Perspectives on KBART Holdings Automation."[25]

Third, the team discovered that exporting an entire collection from Alma sometimes included records for materials other than ebooks. Some collections in the CZ include ebooks, audio, streaming media, and e-journals in the same collection, all for the same vendor and on the same platform. This oversight highlighted the importance of setting correct facets from the beginning. In the future, the team will need to filter collections by material type and export only ebook records. Fourth, the team discovered that ebooks purchased by the consortium were activated in the Network Zone for distribution to consortia members; USF Libraries were unable to automate updates in batch to Network Zone records. Finally, during the MARC record export process in Alma, a collection level bib record was exported in each file. This affected both the automation process and the project's statistical analysis. To resolve this, a team member used MarcEdit to remove these collection bib records before finalizing the files.

Although the Python program has proved effective, the API request limitations have been a challenge. One potential issue was the risk of being temporarily blocked by OCLC because of excessive API requests. Several times, the programmer encountered temporary access restrictions when the program sent numerous requests in a short period. The typical resolution was to wait a couple of hours before resuming the requests. To mitigate this risk, the team opted to keep files small, limiting each file to 1,000 records. This approach, recommended by Terry Reese, helped manage API load and reduced the likelihood of hitting request limits.[26]

The team also established best practices to ensure quality control. For example, when searching for records in OCLC, if there was no OCLC number or ISBN, the Python program would perform a title search for the record in OCLC. Although this had a 50 percent success rate, it also returned incorrect call numbers that required manual correction. Future projects will not include title searching; records with neither an OCLC number nor an ISBN will be moved directly to manual processing.

Many of the corrections to automated changes were due to faulty metadata in the MARC records. Many OCLC records contained "electronic resource" or "online" in the call number field instead of an actual call number, and these records had to be updated manually. The programmer resolved this issue by programmatically filtering these exceptions from the results and then searching for another record with an actual call number. Another change for future projects will be to include subject headings in the program reports for each record. Without the subject headings, catalogers matched call numbers based on the title alone, which could introduce a margin of error depending on the depth of context the title conveyed. Including subject headings from the project's outset will enable more accurate verification of call numbers and eliminate the need to run sets through MarcEdit to find them.

Overall, the workflow for the Ebook Call Number Project performed well, with only a few minor, unexpected issues that the team corrected as the project progressed. Some of the workflow improvements include filtering Alma searches by material type and filtering out local collections in Alma. A separate workflow will be designed to add call numbers to local collections, including

creating an import profile to update the call numbers of local records and filtering local collections only when searching. Removing title searching from the programming improved the API's accuracy, and writing advanced scripting filters eliminated false hits in the call number field, further improving the automation process. During manual call number insertion, adding subject heading retrieval to the automation process made manual cataloging of call numbers easier for the cataloger. These improvements have made a big difference in the workflow performance.

Alternative Methods for Institutions Without a Programmer

The workflow discussed in this article was created through an iterative process. During the development phase, the Call Number Project Team explored various alternative techniques to accomplish different parts of the project. These were ultimately discarded in favor of a more efficient process developed by a programmer. It's important to note that not all institutions have programming resources available. Therefore, the authors document alternative methods to provide alternative paths to accomplish a similar project. Without programming expertise, UI.Vision's RPA, a macro-recording program (https://ui.vision/), and ChatGPT (https://chatgpt.com/) are two tools that librarians could consider using to automate the process of finding call numbers for ebooks, with several caveats discussed in the following paragraphs.

UI.Vision RPA was successfully used to automate the process of looking up ISBNs on the FirstSearch website and retrieving LC call numbers. This free browser extension allows users to record macros, which means that mouse clicks and keystrokes can be recorded as the user navigates through a website. The strengths of UI.Vision RPA include user-friendly basic functionality, the ability to use spreadsheets for input and output, thorough product documentation (https://ui.vision/rpa/docs), and free cost.

The completed workflow involved providing a spreadsheet of ISBNs to a macro, which then searched the OCLC FirstSearch website and recorded the LC call number back onto the spreadsheet. Although successful in trials, the drawback was the time required to obtain results. The macro searched the website in real time, meaning it had to fill out forms, click buttons, and load webpages for each search. Although effective, this process was inefficient compared with a programming solution. Developing the macro was also not as straightforward as recording. The data returned on search result pages was not static and not always in the same location. Extracting the LC number required identifying the correct Xpath, and the macro also needed additional adjustments to account for variables such as page loading times.

Macros created using UI.Vision RPA can automate the process of retrieving call numbers from FirstSearch. This tool could be beneficial for institutions that lack programming support; however, setting up macros can be complex and finicky. Fortunately, UI.Vision provides excellent documentation, and the community offers strong support (https://forum.ui.vision/).

ChatGPT was used experimentally to identify call numbers for a fringe case. This project was completed in ChatGPT-4, the current version at the time, using the Data Analyst GPT, a paid service. Since then, the results have been replicated in the free version of ChatGPT-5. One collection of ebooks that needed

LC numbers lacked ISBNs but had subject headings. ChatGPT was used to look up the subject headings from the LC Subject Headings Word document and return the LC number.

The advantage of using ChatGPT in this case was that it could search across different file types to return results. ChatGPT was given an Excel file containing the ebook titles and a Microsoft Word file with the LC Subject Headings. It was able to look up the subject headings from the Excel file in the Word file and then print the results back into the Excel file. ChatGPT's strengths include its ability to handle multiple file types and to perform complex tasks via conversational prompts without requiring programming or specialized knowledge. However, obtaining the exact output often requires iterative experimentation with the application. A careful examination of the results is also required because ChatGPT can make errors or be incomplete in its responses.

The use of ChatGPT in cataloging has been controversial. For instance, the article "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging" asserted that ChatGPT could make successful MARC records.[27] This set off a flurry of letters to the editor and a scathing review of the article titled "ChatGPT Not Useful as a Tool to Streamline Library Cataloguing Processes."[28] The criticisms centered around how error-ridden the artificial intelligence (AI)-generated MARC records were and how the author, not a cataloger, was unqualified and unable to determine the quality of the records.[29] Lowagie, in his careful, nuanced approach to creating an ethical universal catalog, says that AI, as a tool, "plays a pivotal role in recognizing patterns and relationships across datasets."[30] This project takes a similar view of ChatGPT as a singular, useful tool within a broader methodology. In this use case, ChatGPT accomplishes a specific task with easily verifiable results, making it a useful tool in a cataloger's arsenal.

Automating the retrieval of call numbers for ebooks is still possible even without programming resources. Applications such as UI.Vision RPA and ChatGPT can reduce substantial burdens by automating more labor-intensive tasks in a project, improving reporting and visibility of e-resources.

Future Plans

On completion of the project, USF Libraries' immediate plans included enhancing metadata for e-journals; however, lessons learned from the ebook enhancement enabled the e-journal project to proceed more quickly, and it was completed in the summer of 2024. The Metadata/Cataloging Unit plans to run an annual update of ebooks and e-journals using a modified workflow. Because annual updates will be in smaller batches, using AI applications to validate call numbers against subject headings and enhance metadata could improve the automation process. Additionally, the unit is considering expanding the metadata enhancement to include subject headings and call numbers. To make reports and statistics available to stakeholders across USF Libraries, the Metadata/Cataloging unit will create a dashboard that allows any department to run reports as needed. High-quality metadata is critical to improving the library's ability to generate accurate reports that display library resources for the benefit of the university.

Notes

1. Association of Research Libraries. Systems and Procedures Exchange Center, *Qualitative Collection Analysis: The Conspectus Methodology* [SPEC Kit,0160-3582 ;151] (Association of Research Libraries, Office of Management Services, 1989). https://catalog.hathitrust.org/Record/001078642.
2. Jennifer Benedetto Beals and Ron Gilmour, "Assessing Collections Using Brief Tests and WorldCat Collection Analysis," *Collection Building* 26, no. 4 (2007): 104–7, https://doi.org/10.1108/01604950710831898; Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *College & Research Libraries* 67, no. 1 (2006): 35–49, https://doi.org/10.5860/crl.67.1.35.
3. Michael Luther and Ana Guimaraes, "Applying the Principles of Total Library Assessment to Inform Sustainable Collection Development," *Proceedings of the 2016 Library Assessment Conference* (2017): 181–87; Mary H. Munroe and Jennie E. Ver Steeg, "The Decision-Making Process in Conspectus Evaluation of Collectons: The Quest for Certainty," *The Library Quarterly (Chicago)* 74, no. 2 (2004): 181–205, https://doi.org/10.1086/382846.
4. Jennifer Benedetto Beals, "Assessing Library Collections Using Brief Test Methodology," *Electronic Journal of Academic & Special Librarianship* 7, no. 3 (2006): 1, https://research.ebsco.com/linkprocessor/plink?id=2aae0fc0-a059-30d2-aa1f-e827f0d84517.
5. Jennifer E. Knievel, Heather Wicht, and Lynn Silipigni Connaway, "Use of Circulation Statistics and Interlibrary Loan Data in Collection Management," *College & Research Libraries* 67, no. 1 (2006): 35–49, https://doi.org/10.5860/crl.67.1.35; Mary H. Munroe and Jennie E. Ver Steeg, "The Decision-Making Process in Conspectus Evaluation of Collections: The Quest for Certainty," *The Library Quarterly (Chicago)* 74, no. 2 (2004): 181–205, https://doi.org/10.1086/382846.
6. Karen C. Kohn, "Usage-Based Collection Evaluation with a Curricular Focus," *College & Research Libraries* 74, no. 1 (2013): 85–97, https://doi.org/10.5860/crl-295.
7. Sheila S. Intner and Elizabeth Futas, "The role and impact of Library of Congress Classification on the assessment of women's studies collections," *Library Acquisitions: Practice & Theory* 20, no. 3 (1996): 267–79, https://doi.org/10.1016/0364-6408(96)00067-1.
8. Jonathan Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus: Providing Subject Categorization for Users of Electronic Resources," *Journal of Internet Cataloging* 5, no. 4 (2002): 67–80, https://doi.org/10.1300/J141v05n04_05; Kimberly Looby, "Subject Librarians Vs Collection Data: How to Use Data to Visualize a Small Monograph Collection," *Technical Services Quarterly* 38, no. 1 (2021): 33–53, https://doi.org/10.1080/07317131.2020.1854576.
9. Stephen Paul Davis, "HILCC: A Hierarchical Interface to Library of Congress Classification," *Journal of Internet Cataloging* 5, no. 4 (2002): 19–49, https://doi.org/10.1300/J141v05n04_03.
10. Karen Harker, Janette Klein, and J. Laurel Crawford. "Multiplying by Division: Mapping the Collection at University of North Texas Libraries," *Where Do We Go From Here?: Charleston Conference Proceedings 2015* (2016): 333–40, https://doi.org/10.5703/1288284316279.
11. Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus," 67–80; Looby, "Subject Librarians vs Collection Data," 33–53.
12. Rothman, "Bridging the Gap Between Materials-Focus and Audience-Focus," 67–80.

13. Intner and Futas, "The role and impact of Library of Congress Classification on the assessment of women's studies collections," 267–79.

14. Michael Hughes, "How I Stopped Worrying and Learned to Love the Usage Data," *College & Research Libraries* 81, no. 7 (2020): 1168–83, https://doi.org/10.5860/crl.81.7.1168; Michelle Leonard and Steven Carrico, "Developing a Sustainable Collection Assessment Strategy," *Proceedings of the 2016 Library Assessment Conference* (2017): 172–80.

15. Suzanne Saskia Kemperman et al., "Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper" (OCLC Online Computer Library Center Inc., 2014): 2–24, https://www.oclc.org/content/dam/oclc/reports/data-quality/215233-SuccessStrategies.pdf.

16. Isidoro Gil-Leiva, Maria Carolina Andrade e Cruz, Franciele Marques Redigolo, and Mariângela Sptti Lopes Fujita, "Assignment of Subject Headings and Classification Codes to e-books in South American University Libraries," *Revista Española de Documentación Científica* 47, no. 2 (2024): e385, https://doi.org/10.3989/redc.2024.1513.

17. Donna E. Frederick, "EBook Discovery Metadata," in *Managing E-Book Metadata in Academic Libraries: Taming the Tiger* (Chandos Publishing, 2016): 105–201.

18. USF Libraries Call Number Automation Python Programming is available on request. Please contact Bonita Pollock at pollockb1@usf.edu.

19. Terry Reese, "Opening the Door: A First Look at the OCLC WorldCat Metadata API," *Code4Lib Journal*, no. 25 (2014), https://journal.code4lib.org/articles/9863.

20. JetBrains, "PyCharm," accessed December 10, 2024, https://www.jetbrains.com/pycharm/.

21. Bladimir Toaza and Domokos Esztergár-Kiss, "Automated Bibliometric Data Generation in Python from a Bibliographic Database," *Software Impacts* 19 (2024): 100602, https://doi.org/10.1016/j.simpa.2023.100602.

22. Jung-Ran Park and Yuji Tosaka, "Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms," *Cataloging & Classification Quarterly* 48, no. 8 (2010): 696–715, https://doi.org/10.1080/01639374.2010.508711.

23. Ben Abrahamse, "A Cataloger's Toolkit, Part Two: Using MarcEdit with Spreadsheets," *Technicalities* 42, no. 6 (2022): 16–18.

24. Mingyu Chen, Misu Kim, and Debbie Montgomery, "Ebook Record Management at The University of Texas at Dallas," *Technical Services Quarterly* 33, no. 3 (2016): 251–67, https://doi.org/10.1080/07317131.2016.1169781.

25. Andrée Rathemacher, Matthew Ragucci, and Stephanie Doellinger, "Don't Wait, Automate! Industry Perspectives on KBART Holdings Automation," *The Serials Librarian* 82, no. 1–4 (2022): 91–7, https://doi.org/10.1080/0361526X.2022.2019545.

26. Reese, "Opening the Door: A First Look at the OCLC WorldCat Metadata API."

27. Richard Brzustowicz, "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging," *Information Technology and Libraries* 42, no. 3 (2023), https://doi.org/https://doi.org/10.5860/ital.v42i3.16295.

28. Andrea Miller-Nesbitt, “ChatGPT Not Useful as a Tool to Streamline Library Cataloguing Processes,” *Evidence Based Library and Information Practice* 19, no. 2 (2024): 145–7, https://doi.org/https://doi.org/10.18438/eblip30524.

29. Christine DeZelar-Tiedman, “Response to ‘From ChatGPT to CatGPT.’” *Information Technology and Libraries* 42, no. 4 (2023), https://doi.org/https://doi.org/10.5860/ital.v42i4.16991.

30. Hannes Lowagie, “From Bias to Transparency: Ethical Imperatives in AI-Based Library Cataloging” (paper presented at the IFLA WLIC 2023 Satellite Meeting on Universal Bibliographic Control, Brussels, Belgium, August 18, 2023), https://repository.ifla.org/items/351270b4-31b2-4481-a67f-8412e48198bb.

Book Review

Collection Management Basics, Eighth Edition. By Margaret Zarnosky Saponaro, John Novak, and G. Edward Evans. New York: Bloomsbury, 2025. 432p. $67.95 softcover (ISBN 9798216170259); $130.00 hardcover (ISBN 9798216170297).

Collections Management Basics is a textbook focused on all aspects of collections—from selection to preservation—for all types of libraries. This updated edition is the eighth in the series and continues to discuss different aspects of the collection management cycle in easy-to-understand chapters. Each chapter draws attention to information with callout boxes such as "Check This Out," which highlights specific articles or examples. Another callout box titled "Something to Watch" captures a new trend in collections. In addition, "From the Advisory Board" provides a narrative from one of the three advisory board members from public and school libraries who helped contribute to this edition. The "From the Advisory Board" callout boxes resembled the quick tips that you would hear in between sessions at a library conference akin to "at my library we did this" type advice, which could either be helpful or impractical depending on the reader.

This textbook appears to be comparable to *Fundamentals of Collection Development and Management* by Peggy Johnson and Mary Beth Weber and *Collection Development and Management for 21st Century Library Collections* by Vicki L. Gregory.[1] Although Johnson and Weber's textbook is more in depth and Gregory's is a concise text less concerned with the history of collections and more focused on present day, the authors of this textbook strike a balance between the two that is more approachable for library and information science students.

While *Collections Management Basics* is meant to serve as a library and information science textbook, it is also a great resource for new or mid-career librarians who have taken on a role within the collection management cycle. Each element of collections is introduced, explained, and then dissected, providing the history and updated challenges that many library staff encounter when working in collections. For example, the chapter "Assessing User Needs" provides a background on the importance of conducting a needs assessment based on library users (78), explains how to get started with collecting data (82), and then goes in depth on how to find demographic data (87) and how to communicate value using data visualizations (89).

This eighth edition of *Collections Management Basics* covers the unique challenges brought on by the COVID-19 pandemic. The authors discuss the importance that was placed on e-resources during this period (8), the way collaboration changed due to remote work (243), and the need for additional disaster preparedness plans to ensure access to information (349). Many of the chapters touch on how the pandemic affected all types of libraries and what decisions were made as a result of this new shift in patron behavior. For example, many universities reduced their print serials based on the lack of need by patrons during this time of limited print access. Library staff had time to complete assessment reviews of their print serials collections to see what could be removed in order to save space and future staff time (290).

In addition to the pandemic coverage, this edition explores trends to watch closely that were emerging at the time of publication. Some of these trends are called out specifically in the "Something to Watch" sections, while others are interspersed within the chapters. Artificial intelligence started to impact libraries in a significant way during the writing of this edition, and its implications for copyright were only beginning to be explored. Other new trends that affected library collections were #BookTok and review bombing. The #BookTok trend has contributed to an increase of patron requests and will continue to impact book selection in public and academic libraries (50). In determining which books to purchase, library staff often read online reviews, including reviews on the Goodreads website. Goodreads addressed recent "review bombing," which is when users post negative reviews that have nothing to do with the book itself and often are used to harass authors (106). This anecdote served as a reminder to read multiple reviews for a book before making a final purchasing decision.

Overall, this textbook allows library and information science students, as well as library staff, to gain an understanding of each aspect of the collection management cycle and does so in a manageable fourteen chapters. Each chapter ends with a "Points to Keep in Mind" section that summarizes the chapter in a short, bullet-point format. There is also a list of references in every section that allows readers to go deeper into an example or resource from that chapter. While the three authors are experienced academic librarians and administrators, the addition of the public and school advisory board members enables this resource to be beneficial to public and academic library staff alike. However, there are few examples from special libraries. Future editions could include anecdotes or advisory board members from special libraires to assist library staff in hospital, law, or other unique libraries. The examples from academic and public libraries can help all collections library staff from all libraries understand the basics of the collection management cycle and how to handle challenges with intellectual freedom, open access, digital resources, and financial constraints. The authors have updated this edition to account for new emerging trends and new ways to perform collection management basic duties. —*Rayla E. Tokarz (rtokarz@unr.edu*), University of Nevada, Reno

Note

1. Peggy Johnson and Mary Beth Weber, *Fundamentals of Collection Development and Management*, 5th ed. (Facet Publishing, 2025); Vicki L. Gregory, *Collection Development and Management for 21st Century Library Collections: An Introduction*, 2nd ed. (ALA Neal-Schuman, 2019).