Library Resources & Technical Services

ISSN 2159-9610
April 2021
Volume 65, No. 2

Core LEADERSHIP INFRASTRUCTURE FUTURES

65 | 2

Library Resources & Technical Services, https://journals.ala.org/lrts (ISSN 2159-9610) is published quarterly by the American Library Association, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601. It is the official publication of the Core: Leadership, Infrastructure, Futures, a division of the American Library Association, and provided as a benefit to members. Subscription price to nonmembers is $100. Individual articles can be purchased for $15. Business Manager: Kerry Ward, Executive Director, Core: Leadership, Infrastructure, Futures, a division of the American Library Association. Submit manuscripts using the online system at https://journals.ala.org/index.php/lrts/login. Mary Beth Weber, Editor, *Library Resources & Technical Services*; mbfecko@rulmail.rutgers.edu. Advertising: Core: Leadership, Infrastructure, Futures, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601; 312-280-5038; fax: 312-280-5033; alcts@ala.org. ALA Production Services: Tim Clifford and Lauren Ehle. Members may update contact information online by logging in to the ALA website (http://www.ala.org) or by contacting ALA Membership Relations and Services—*Library Resources & Technical Services*, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601; 1-800-545-2433. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to *Library Resources & Technical Services*, Subscription Department, American Library Association, 225 N. Michigan Ave., Suite 1300, Chicago, IL 60601; 1-800-545-2433; fax: 312-944-2641; subscriptions@ala.org.

Library Resources & Technical Services is indexed in Library Literature, Library & Information Science Abstracts, Current Index to Journals in Education, Science Citation Index, and Information Science Abstracts. Contents are listed in CALL (Current American—Library Literature). Its reviews are included in Book Review Digest, Book Review Index, and Review of Reviews.

Instructions for authors appear at https://journals.ala.org/index.php/lrts/about/submissions#authorGuidelines. Copies of books for review should be addressed to Elyssa M. Gould, University of Tennessee Libraries, 1015 Volunteer Boulevard, Knoxville, TN 37996-1000; lrts-bookreviews@lists.ala.org.

LRTS was available in print (ISSN 0024-2527) from 1957 through 2014. Single print issues from volume 38 through volume 58 can be purchased for $30 each. Contact core@ala.org with purchase requests.

Visit *LRTS* online at https://journals.ala.org/lrts.

Library Resources & Technical Services

ISSN 2159-9610 **April 2021** **Volume 65, No. 2**

Cover image: Ideal Beach, NJ, Mary Beth Weber, September 30, 2020.

Editorial

Mary Beth Weber

This past week marked the one-year anniversary of my staff and me working from home. When we packed up our cubicles and offices in March 2020, no one expected to be working remotely from home for long. We honestly expected to return in a few weeks or at least by the end of April. In the meantime, the university kept extending our work-from-home agreements. During the past year, we have acquired new skills and ways of working. For example, we have mastered how to use WebEx and Zoom for meetings and have realized that this technology can make our meetings more effective. Although some people complain of Zoom fatigue, we have found that our meetings are shorter, and no one lingers afterward. We may start meetings with small talk, but when we are done, people are ready to sign off. Participants who may have difficulty speaking up can choose to use the chat box, and entering terms like "stack" in the chat box helps to ensure that everyone gets a chance to speak and in a predetermined order. It avoids having everyone try to speak at once and ensures all have a chance to speak.

The lack of a commute for many of us has meant an earlier starting time or starting a day without rushing out the door and navigating traffic jams or construction. It has enabled some of my employees (as well as myself) time to exercise before work. It can also mean time to eat a more substantial breakfast or lunch that is not rushed. Time spent with families, loved ones, and pets is also more easily managed.

There are challenges in working from home that include sharing a workspace with children, spouses, and pets. Parents may have the additional challenge of having children at home who are learning remotely while they are working. There might also be disruptions that one may not get when working at the office, such as garbage pick up or landscapers outside the window shouting and using leaf blowers. People frequently have had to make do with what spaces and equipment they have at home. Most libraries lack the funding to provide equipment to employees working remotely, and others permitted employees to bring home chairs and desktops. Despite these challenges, we as a profession have persevered.

Several of my colleagues from other libraries have returned to work on-site. Some work hybrid schedules, and others have returned to work 100 percent on-site. In my case, my staff returned to work in August and September 2020, while the librarians continue to work remotely. It has not been determined who will return to work and when, or how our library system will reopen. It is the hope that we will return to work in the fall of 2021, and it will be a gradual return.

Going forward, there is no doubt that this experience will change how we work and provide services. We have proven that technical services functions can be provided remotely or in a hybrid environment. Vendors who provide outsourced services have successfully done so for years. In addition to ensuring uninterrupted and efficient services to users, the ability to work remotely or on a hybrid schedule can enhance job performance (and job satisfaction) for some employees. I suspect there will be an increase in libraries that provide flexible schedules or

reconfigured workspaces to ensure a safe and hygienic environment. Some of my colleagues from other academic libraries have reported that their university administrators have seen this as an opportunity to reconsider library real estate. There have been many references to "returning to normal," but that raises the question of how "normal" is defined. We cannot return to how we worked previously for several reasons. The pandemic has led us to examine how we work, staffing levels, and priorities. We have seen a gradual shift in the materials we handle and associated processes due to a greater emphasis on acquiring and making available electronic resources. Some libraries no longer have serials librarians or departments and that work has been subsumed under electronic resources management. Archives and special collections, which deal with realia and rare materials, also have taken a major focus on digitized collections and finding aids, for example. Cataloging departments have focused more on original materials as vendor supplied cataloging in the form of record sets has become common as staffing has decreased. We are already shifting as a profession, and the shift in how and where we work, plus how we deliver those services, is a logical next step.

Not surprisingly, there has been a proliferation of presentations and papers on how COVID-19 has impacted our lives and work. There are numerous calls for chapter and paper proposals on the topic. This issue includes the first paper I have received related to remote operations due to COVID. On a personal note, I am also working on a publication related to the topic. This leads to my usual overview of the contents of this issue:

- In their paper "Exploring the Impact of Digitization on Print Usage," Teper and Kuipers explore the belief held by librarians and administrators that digitization and access of items through the HathiTrust Digital Library may reduce or eliminate demand for the corresponding print content. They provide a data-driven examination of the use of their institution's print items that correspond to the digital materials deposited into HathiTrust, and detail the results and process by which data was gathered, managed, and digested to yield the results.
- "On the State of Genre/Form Vocabulary: A Quantitative Analysis of LCGFT Data in WorldCat," Bitter and Tosaka report on a quantitative analysis of the LCGFT vocabulary within a large set of MARC bibliographic data retrieved from the OCLC WorldCat database. Their intent was to provide a detailed analysis of the outcomes of the LCGFT project that launched by the Library of Congress in 2007. The findings point to a moderate increase in LCGFT use over time, yet the vocabulary has not been applied to the fullest extent possible in WorldCat.
- Gentry's paper "Digital Collections at a Distance: Telework during the COVID-19 Pandemic" details how a team at her library that was tasked with the creation of digital collections succeeded at telework and executed essential functions despite not being able to digitize new content from March to July 2020 during the COVID-19 pandemic. It is her hope that managers of similar types of units will gain strategies to create similar telework projects at their institution and she shares the lessons learned while working and supervising employees remotely.
- Book reviews, courtesy of *LRTS* Book Review Editor Elyssa Gould.

Exploring the Impact of Digitization on Print Usage

Thomas H. Teper and Vera S. Kuipers

Librarians and administrators speculate that the digitization and access of items through the HathiTrust Digital Library may reduce or eliminate demand for the corresponding print content. This belief feeds into a perception that monographs housed in academic libraries and delivered via such services are ripe for deduplication or outright withdrawal, yet other institutions may remain dependent upon those holding titles to provide print-based access for their patrons. Embracing HathiTrust's emerging Shared Print Monograph Program, more than seventy-nine member institutions committed to retain print monographs that correspond to those digitized from their collections. Putting aside concerns expressed by some about the meaningfulness of those commitments, not all members made such commitments. Moreover, retention commitments are not always publicly displayed, leading to scenarios in which such commitments may be used by other institutions to withdraw from their collections, based on these holdings. This paper provides a data-driven examination of the use of one research library's print items that correspond to the digital materials deposited into the HathiTrust, detailing both the results and the process by which data was gathered, managed, and digested to yield the results.

In the early stages of library digitization, assumptions arose about the potential that the digitization and delivery of items online would reduce demand for the corresponding print titles. By the middle of this century's first decade, this belief furthered speculation that the reduced demand served to advance the goals of preservation by diminishing wear and tear on items, facilitated the goals of collection managers by easing decision-making about relocating items to storage facilities, and the served those interested in developing new and innovative services once such materials were relocated.[1] Digitized back file content acquired from commercial vendors reduced the demand for print copies of much of the commercially published literature as the ease of on-demand, desktop access supplanted the need to consult print journal runs.[2] In recent years, reported circulation numbers for print resources declined, providing ample evidence to make a conjecture that preservation needs are declining due to reduced wear and tear.[3] Libraries also reported factoring the availability of digital surrogates into many of their collection management decisions.[4]

Some remained skeptical about factoring these changes into collection management decision making, expressing trepidation ranging from concern about the book as object, quality of the scanning, the accuracy of the metadata underlying discovery, and the uncertainty about the availability of print copies through lending networks—the fragility of which the COVID-19 pandemic laid bare. Although some of these concerns pre-dated contemporary mass digitization efforts, they assumed a new urgency in the last decade as digitization and

Thomas H. Teper (tteper@illinois.edu) is Associate University Librarian for Collections & Technical Services at the University of Illinois at Urbana-Champaign. **Vera S. Kuipers** (verav2@illinois.edu) is a recent graduate of the iSchool at the University of Illinois at Urbana-Champaign.

Manuscript submitted September 20, 2020; returned to authors for minor revision January 21, 2021; revised manuscript submitted March 10, 2021; accepted for publication March 18, 2021.

The authors presented preliminary results of this research at ACRL (Association of College and Research Libraries) 2019 Conference in Cleveland, Ohio.

deposit into services such as HathiTrust increased, speculation about reductions in demand for corresponding print content again arose, and suspicion about administrative intention resurfaced.

Despite this, there appears to be remarkably little published data on the actual impact of digitization on the use of their physical counterparts, either locally or through borrowing networks. With a history of more than one hundred years of developing and maintaining resource sharing networks, many research libraries in the pre-COVID-19 era embraced the notion that, in some cases, their institutions would remain dependent upon those holding physical titles to provide print-based access for their patrons. Indeed, the notion of resource sharing remains a foundational assumption of discussions around collective collections.[5] Knowing that this cross-institutional dependency exists, many HathiTrust member institutions committed to retain print monographs that correspond to those digitized from their collections. However, such commitments are not universal among the membership or collectively displayed to other libraries or members, meaning that retention commitments remain challenging to identify. Knowing more about how the availability of digital surrogates may impact the usage of print monographs is a critical component of the developing collective collection.

Problem Statement

To draw meaningful conclusions about the relative use of volumes after digitization, the project lead developed a series of questions and charged a research team to gather and evaluate datasets from three different sources. The primary challenge was that the datasets harvested to gather this information did not directly correspond to one another. To surmount this challenge, the research team pursued the following steps: (a) compiled several locally developed datasets, (b) imported the datasets into an MS SQL Server database, (c) performed data cleaning and manipulation, (d) determined unique item identifiers to connect the datasets, (e) wrote and ran SQL queries, and (f) created data visualizations in Tableau to illustrate answers to the questions. The three types of datasets initially imported included

- a set of 10.7 million records of every physical item within the University of Illinois at Urbana-Champaign Library's (U of I Library) Voyager catalog as of January 2018;
- a record set of 8,622,399 items from the U of I Library's "*archive transactions table*" that detailed circulations (checked out and returned items, not including renewals)—of all physical collection items from the library during the period spring 2002–December 31, 2019 (the entirety of the recorded transactions on the library's integrated library system), and "*current transactions table*" that logged the 81,207 items currently checked out. The project team merged these two datasets into one circulation dataset with total of 8,703,606 records, providing a complete record of circulation history from 2002 through 2019; and
- a record set of the 847,247 items digitized from the U of I Library's collections that are available via HathiTrust Digital Library. The project team downloaded an initial dataset from the HathiTrust's Hathifiles repository as a tab-delimited text file that included bibliographic records for every item in the HathiTrust collection, which contained 17,153,606 items as of January 1, 2020.[6] Using the source bibliographic record, the authors narrowed the dataset to include only the items digitized from the U of I Library's collection.

One challenge in drawing conclusions using the available datasets is that the records associated with a particular digitization date are not precise enough to pinpoint exactly how circulation dates and the digitization date fell chronologically within a particular year. Consequently, with a circulation record covering complete years running from 2002 through 2019, data about an item digitized in 2010 required reporting information into periods before and after digitization that consisted of entire calendar years. In this case, data about the item in question required reporting from the years 2002 to 2010 (to count circulation before digitization) and from 2011 through 2019 (to count circulation after digitization).

Method

Using the three types of datasets gathered into the MS SQL Server database, the research team explored the local circulation for volumes from the University of Illinois at Urbana-Champaign's collection that are digitized and available via HathiTrust. This analysis broke down usage by disciplinary fields with the intent of developing a more nuanced understanding of usage for print resources both prior to and after their digitization. This project sought to explore the following research questions:

1. Were there subject-based differences in the ongoing demand for the original print resources?
2. Was there a measurable difference in demand for these print resources from the periods before and after an item was digitized?
3. Was there a difference in demand after digitization for

those items that are freely available as full-text (most are pre-1923 through the period of the study) as compared to those in which copyright or other restrictions limited the digital access?

Literature Review

There is an extensive body of literature on the development, underlying premise, and perceived flaws inherent in print retention agreements for both serial and monographic literature. This literature further details the emerging overlap of collections and holdings within regionally defined areas and the challenges posed by image quality in both commercially digitized content and content digitized and delivered via HathiTrust. Yet there appear to be no published assessments that specifically examine the potential impact of digitization on the usage of corresponding print resources save for a limited study conducted by IFLA/UNESCO that generalizes about the use of original special collection items post-digitization.

A voluminous literature currently surrounds the development of print retention agreements and the possible flaws that may undermine the successful implementation of cross-institutional deduplication efforts. Most of the publications about these are relatively recent, although the earliest calls for a "national lending center" specifically intended to avoid unnecessary duplication date to the late 1800s, and calls to develop a National Periodicals Center date to 1973–80, when Steven proposed a national serials repository. Although that effort failed, partly due to the political climate and lack of federal funding, it set the groundwork for further discussions.[7] More recently, the Center for Research Libraries (CRL) assumed a leadership role in trying to coordinate print retention efforts for serials. When CRL convened the 2004 conference "Preserving America's Printed Resources," the organization effectively embarked on a series of discussions and iterative developments that resulted in them assuming a central role in the development of a serials print registry.[8] CRL's continued engagement in discussions with the serials retention programs increasingly form a part of the collection management strategies for North American academic libraries. CRL sponsored the 2015 "Preserving America's Print Resources II: A North American Summit," and published outcomes from that meeting in *Print Archiving and Shared Print in North America: A Preliminary Analysis and Status Report*.[9]

Monographic print retention presents different challenges to libraries than corresponding programs focused on serial literature. This is partly due to higher instances of bibliographic uniqueness among monographs, which publishers often produce in multiple editions and over many years. Monographs frequently include purposeful and accidental changes made by the author, editor, or typesetter within their pages. Moreover, due to the lower instance of duplication among titles, lower return on time invested in deduplication, and the relatively recent availability of significant bodies of digitized monographic literature, monographic collections did not garner the initial attention of those advocating deduplication for purposes of space savings. Yet discussions about the transformative value, underlying framework, benefits, and weaknesses of monographic print retention schemes are not new. While the widespread application of copyright deposit eased the adoption of nationwide monographic print retention schemes in European academic libraries, the idea took longer to catch on in the United States as the complicated patchwork of political, consortia, and educational bodies magnified the challenges faced by a geographically larger nation. Among the earliest meaningful recent works focused on the United States are Kieft and Payne's "A Nation-Wide Planning Framework for Large-Scale Collaboration on Legacy Print Monograph Collections," Nadal and Peterson's "Scarce and Endangered Works: Using Network-Level Holdings Data in Preservation Decision-Making and Stewardship of the Print Record," and Malpas' *Cloud-Sourcing Research Collections: Managing Print in the Mass-Digitized Library Environment*.[10] These works influenced the potential for such programs, their value as mechanisms to preserve our cultural heritage, how they might be constructed, and the potential for overlapping holdings to be viewed as expendable. They influenced discussions about the subsequent development of monographic print retention programs. The most prominent of the monographic retention programs—the HathiTrust Print Monograph Archive—resulted from a ballot initiative developed for the 2011 HathiTrust Constitutional Convention. From this, HathiTrust emerged as the leader in developing the closest thing to a national print retention program. Whereas this proposal did not explicitly call for any institution to withdraw content, it operated on the assumption that HathiTrust would exert a transformative influence on the management of print collections and that some institutions would withdraw content based upon the presence of digital surrogates in the HathiTrust Digital Library. It sought to create a baseline framework for ensuring continued retention and access to print titles that corresponded to the digitized monographs in HathiTrust.[11] What this and other programs lack is the presence of a concerted national framework, a point highlighted by the 2016 report "Concerted Thought, Collaborative Action, and the Future of the Print Record."[12]

The developing monographic print retention models have their own strengths and weaknesses. Their weaknesses as tools to manage local collections include the two most prominent issues: (a) concerns about the quality of the digitized content and its metadata, and (b) concerns about

how the retention commitments made by other institutions may be used by librarians to guide the deduplication of local holdings. Although both papers focused on commercially digitized content, the challenges inherent in making collection management decisions and withdrawing print titles based on the availability of digital surrogates featured prominently in Joseph's "Image and Figure Quality: A Study of Elsevier's Earth and Planetary Sciences Electronic Journal Back File Package" and her 2012 follow-up study.[13] With respect to HathiTrust, image quality featured prominently in Conway's more recent work. In "Preserving Imperfection: Assessing the Incidence of Digital Imaging Error in HathiTrust," Conway reported on a study of image quality for titles digitized and delivered via HathiTrust, seeking to quantify the prevalence of errors in pre-1923 items.[14]

The other concern regarding the utilization of digital availability via HathiTrust as a tool for driving local print retention decisions centers on the challenge of accurately determining the duplicate status or condition of materials held locally or across multiple institutions. Stauffer tackled this challenge in "My *Old Sweethearts*: On Digitization and the Future of the Print Record," and Teper sought to further explore this topic with her paper "Considering 'Sameness' of Monographic Holdings in Shared Print Retention Decisions."[15] Stauffer's work expressed concerns over the high level of variance among the items in his sample set, and Teper appears to have verified many of the conclusions drawn by Stauffer.

A quantitative study that draws a direct correlation between print usage and online availability is a 1999 publication jointly issued by the International Federation of Library Associations (IFLA) and UNESCO.[16] Among questions associated with digitization practices, the survey examined post-digitization access to original items. Focusing on the use of special collections materials from several national libraries after digitization, the study included a note indicating findings that post-digitization demand for items can increase. Again, this study focused on special collections and indicated that discovery could spur a higher interest in the original items and an increased instance of use.[17] That said, there seem to be few studies that directly compare pre- and post-digitization use of general collections materials. As early as 1999, the Council on Library and Information Resources published *Scholarship, Instruction, and Libraries at the Turn of the Century.* In this publication, the authors highlighted reports from multiple academic task forces, one of which noted that the enhanced discoverability of digitized materials increased the demand for corresponding print materials.[18] Smith referenced this finding in her 1999 CLIR publication *The Future of the Past: Preservation in American Research Libraries.*[19] Yet the lack of quantifiable studies about the impact of digitization on demand for print monographs remains a challenge for those tasked with collection management decision making. Aggregating and analyzing that data is critical to future collection management decisions in the context of the collective collection.

Analyzing and Managing the Data

Based on the aggregation of datasets of bibliographic and item level data representing 10.7 million items (10,601,294 when deduplicated) at the U of I Library, circulation data for the same items dating from 2002 through the end of 2019, and the digitization and availability of these items via HathiTrust Digital Library, the research team conducted a circulation analysis of the aggregated data.

With respect to the specifics of this study, the research team sought to quantify changes in the usage of printed resources after digitization and delivery via HathiTrust compared to the period prior to digitization. To accomplish this, it was necessary to link three datasets that shared no single common point of intersection and to identify the usage of individual items.

To overcome that challenge, the team devised the following solution. In the 10.6-million item deduplicated dataset of the library's print collection, a unique item identifier for each physical item record is *Item_ID*. In the combined circulation transactions dataset, each record represents a single circulation (not a single item); thus, to count how often a particular item circulated, the team counted the number of records in which that *Item_ID* appears. If an item never circulated, no records appear in the combined circulation transactions dataset. Unfortunately, the dataset of digitized items does not include the library's item identifier (*Item_ID*), or any other common identifier. This makes it difficult to match the dataset of digitized items with either the library's print collection dataset or the circulation transactions dataset since they do not share common unique identifiers. However, the HathiTrust dataset provides *htid*, a permanent HathiTrust item identifier.[20] For items digitized from the U of I Library's collection, *htid* contains an item's barcode information. Using an item's barcode, the authors found an item's *Item_ID* for digitized items in the *Item_ID/Barcode* dictionary for the library's print collection. Using the digitized items' *Item_ID*, the authors matched the HathiTrust dataset with the combined circulation transactions dataset by *Item_ID* and retrieved information about circulations of digitized items (see figure 1).

The primary problem emerged when trying to detect the digitized item's *Item_ID* based on the permanent HathiTrust item identifier (*htid*). The identifier consists of two parts that are divided by a (.) dot. The authors identified the section before the dot *htid_prefix*, and the section following the dot—*htid_suffix*. For digitized items from

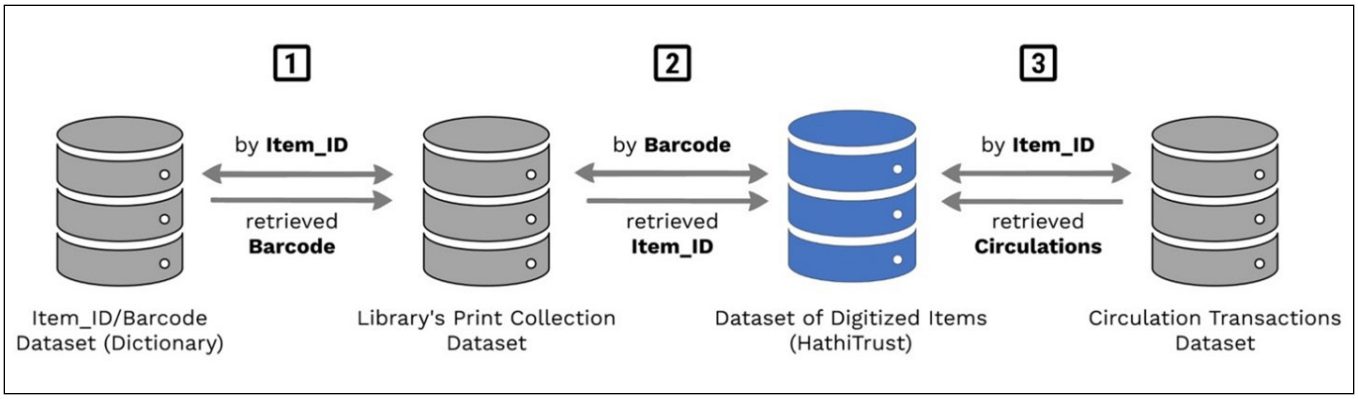


Figure 1. Retrieval of the Circulation Information for the Digitized Items

the U of I Library's collection, *htid_prefix* indicates the source of content and the organization that digitized the content. In the dataset, *htid_prefix* has the following distinct values: *uiuc*, *uiug*, *uiuo*, *uiul*. All four prefixes start with *uiu*, indicating that the source of content is the U of I Library, and end with one of the letters *c*, *g*, *o*, *l*, which specifies the digitization source. Thus, *uiuc* means that an item is locally digitized (i.e., digitized by the U of I Library), *uiul* indicates digitization by the U of I's Law Library, *uiug* is assigned to the items digitized by Google, and *uiuo*—by OCA (Internet Archive). Consisting of 847,247 items, the library's digitized collection was almost 84 percent digitized by Google (710,706 items), 10 percent (10.3 percent) digitized by the Internet Archive (87,562 items), and slightly less than 6 percent digitized by the library itself. This 6 percent comprises 39,241 items digitized by the Law Library and 9,738 items digitized by the Main Library (see table 1).

Table 1. Digitization of U of I Library Collection by Institution

htid_prefix	Institution Name	No. of Digitized Items	% of Total Digitized Collection
uiug	Google	710,706	83.9
uiuo	Internet Archive	87,562	10.3
uiul	U of I Law Library	39,241	4.6
uiuc	U of I Library	9,738	1.2

The *htid_suffix*, which is the item's identifier, varies and depends on how the digitizing institutions manage the digitized items and requires the source of content institution to submit metadata. For example, Google and the authors' Law Library use the print item's barcode as a digitized item's identifier; locally digitized single-volume monographs contain the item's bibliographic identifier *Bib_ID* in *htid_suffix*, whereas multi-volume monographs and serials—*Bib_ID* combined with volume, issue, or publication year information; and the Internet Archive assigns its own number to an item as an item's identifier, which starts with *ark:/*. Understanding the origin of *htid*, the authors started their search for *Item_ID*s for the items when metadata contained the print item's barcode information (749,947 digitized items) by matching the item's barcode with barcodes in the *Item_ID/Barcode* dictionary (see figure 1). The dictionary is a dataset with all the *Item_ID*s from the U of I Library's Voyager catalog and their active barcodes, and all previous (inactive) barcodes. This also includes the date the barcode was assigned to an item. Thus the authors pulled *Item_ID*s for 747,706 (99.7 percent) digitized items. The remaining 0.3 percent of the items either contained typos in their barcode metadata or lacked items in the library catalog.

For the remaining items with *Bib_ID* or *ark:/* as the item identifier, (97,300 digital items) and items with misspelled barcodes, the project team used a different approach to obtain *Item_ID*. Since both the library's Voyager catalog data set and HathiTrust's dataset of digitized items included the bibliographic identifier (*Bib_ID* and *source_bib_num*, respectively) and volume information fields, the authors used that metadata to match the datasets (see figure 2). In the library's Voyager's print collection dataset, the volume's enumeration and chronology data are in the separate fields, *Enum* and *Chron*; whereas in the HathiTrust dataset, only one *description* field describes both types of metadata. Thus, after preliminary cleaning and manipulation of the enumeration and chronology metadata to match the datasets, the authors identified *Item_ID*s for another 62,120 digitized items. That resulted in a total of 809,826 items, which is 95.6 percent of the entire dataset of digitized items.

Less than 4.5 percent of the digitized collection remained unidentified due to several reasons, including the metadata in the HathiTrust's *description* field did not coincide with how library personnel recorded the data in the library's Voyager's *Enum* and *Chron* fields.

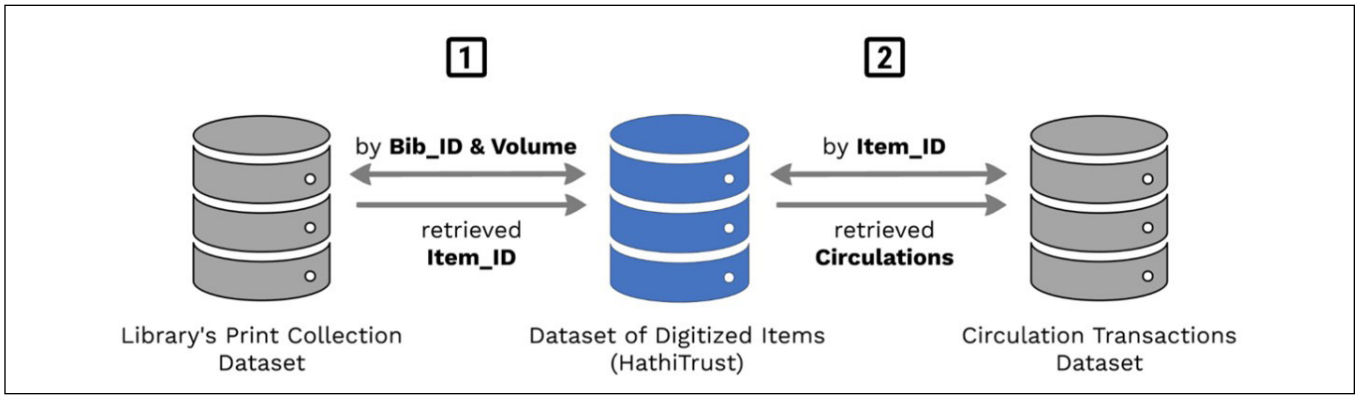


Figure 2. Retrieval of the Circulation Information for the Digitized Items by Bibliographic Identifier and Volume Information

For example, in addition to the added or removed spaces/commas/dots/colons/brackets between enumeration and chronology information, the chronology month and year data were swapped/pruned/modified, and the abbreviation of the word "volume" appeared in various forms. In the case of bound volumes, the library's print collection dataset consists of one record that provides the range of volumes, while the HathiTrust's data set provides a separate record for each volume in the bound volume. The different combinations of these inconsistencies resulted in a variety of ways for how volume details are represented in the HathiTrust's *description* field, and, consequently, require extensive and time-consuming data cleaning.

Other reasons why the U of I Library's print collection data set and the data set of the digitized items did not match by volume information in the print collection data set center on the following factors: (a) enumeration and chronology fields were not provided for the multi-volume monographs and serials, and (b) there is more than one copy of the item in the print collection. Because the authors focused on circulations at the item level, not title level, it became necessary to match datasets by volume information and not just by bibliographic identifier (*Bib_ID*). However, when items from one dataset did not match items from another by both *Bib_ID* and volume information, the authors narrowed the library's print collection dataset to single-volume titles and then matched the datasets only using the *Bib_ID* field. Finally, there were cases when items from the HathiTrust's dataset (which is an extract as of January 2020) did not appear in the library's print collection dataset (extract as of January 2018) because an item's record was added to the library's Voyager catalog after the dataset was extracted for the analysis. Thus the record from the dataset of the digitized items does not have a counterpart in the print collection dataset. The authors verified the *Item_ID* for more than 95 percent of the collection of the digitized items, which let them precisely determine the usage of print counterparts for 809,826 digitized items.

Additionally, the research showed that the library's print collection included an item's circulation analysis at the subject level. Since the library implemented the use of non-standardized Dewey Decimal Classification (known as "Exceptional Dewey") and subject headings in the 1960s to provide more nuanced discovery for literature in a research collection, determining subject heading information required considerable work to assign subject headings to the 10.6 million deduplicated items based on their call numbers. Like many institutions, the U of I uses several classification schemes, including Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), the US Superintendent of Documents Classification (SuDocs), United Nations Documents Classification, and locally developed schemas for specialized collections. Furthermore, some call numbers include a prefix or several prefixes that catalogers assigned based on format, book size, or collection. Thus more than 140 different prefixes were identified, for example, Quarto (Q.), Folio (F.), Biography (B.), quarto Biography (Q.B.), Bibliographies (A.), Textbook (TEXT.), school collection S-Collection (S.), folio S-Collection (F.S.), quarto S-Collection biography (Q.SB.), picture books S-Collection (SE.), Government Documents (DOC.), CD-ROM Government Documents (CDROMDOC.), Microfiche (MFICHE), Digital video disc (DVD), Cavagna Sangiuliani Collection (Cavagna), Carl Sandburg Collection (SNDBRG), quarto Sandburg Collection (SNDBRGQ), and microfilm Sandburg Collection (SNDBRGFILM). A large variety of prefixes and their combinations, along with spelling inconsistencies and typos, further complicated the task of determining an item's classification and subject heading. In all, approximately, 1,205,432 of the records were classified with LCC, and 375,138 were government publications, including many with SuDoc classification. The bulk, totaling 6,595,595 records, were classified using the previously discussed "Exceptional Dewey." The remaining items were classified with other, locally developed schemes applied to a multitude of specialized collections, records for withdrawn

items, and records that contained errors in call numbers or other critical identifying metadata.

Limitations

In analyzing this data, the project team considered numerous constraints. First, the data itself contained limitations. Compiled through decades of work by individual library personnel, the catalog data itself contained variances and errors that required manipulation and massaging. Moreover, the data brought together multiple datasets that required remediation to ensure common links between them.

In addition to limitations of the data sets themselves, local configurations impact the circulation data gathered into the library's ILS. Areas of scholarly interest on individual campuses and among lending networks shift with trends, popular events, and even the presence of key faculty with specialized research areas. Furthermore, this study did not consider the influence that purchased commercially digitized backfiles might exert on usage of print counterparts.

Finally, the Association of Research Libraries has documented a strong decline in print usage.[21] This appears to be a general trend across research libraries. Some of the decline clearly results from the replacement of print journals with digitized journal backfiles. Versaket et al. documented this in an arXiv preprint in 2014.[22] There is, however, no direct link established between the general decline of circulation and digitized monograph literature.

Results

The analysis answered three distinct questions that focused on the usage of the print resources, differences in the usage of print items after their digitization, and whether those differences varied based upon the full-text or partial, or "snippit," view presented due to copyright restrictions. The analysis provided subject-based data to present a more nuanced understanding of usage as it directly impacts collection management activities. In this analysis, the authors noted that the highest level of print circulation over nearly two decades fell within the social sciences, that there is a measurable decrease in demand for print items after their digitization, and that those items available as full-text experienced a slightly greater decline in print usage. The authors discuss each of these findings in more detail in the following results sub-sections, which address a specific research question.

Research Question 1: Are there subject-based differences in the ongoing demand for print resources?

Results indicate that there are measurable differences in the overall usage of print resources in the library's collection, based on their classifications. The total number of items from the library's print collection data set used in the subject-based analysis is 7,797,819, where 1,204,687 records were classed using LCC, and 6,593,132 items were classed with a local variation of DDC. Overall, 23.9 percent of the items, which is more than 1.86 million, in all formats represented in the dataset, circulated between 2002 and 2019. The total number of circulations for those items is 6,209,034 times. At the beginning of 2002, the library's migration from DRA to Voyager meant that the catalog failed to fully capture all circulation data during the first quarter of that year. That explains the significant difference in total number of circulations between 2002 and 2003 years and peak in the latter (see figures 3,4, and 6).

During the entire 2002–2019 period, in DCC, the highest demand was for print items in the [300]—Social Sciences subject. However, after 2003, the subject showed a steady decrease in circulations except for the 2005 and 2006 years when the circulations were nearly the same. Within nine years, from 2003 to 2012, an annual number of checkouts dropped by a third, from 89,767 to 61,557, and within the next seven years dropped to 28,640 by the end of 2019. Over the years, all subjects in the library's DDC range experienced a gradual decline in the demand for print resources. By 2019, they all showed one third of the circulations totals that they had in 2003 (see figure 3).

In LCC, [P]—Language and Literature and [M]—Music and Books on Music subjects stand out by their annual number of circulations, which is higher in comparison with other subjects in the classification. During 2002–2019, annual circulations in [P]—Language and Literature subject ranged between 37,467 and 19,518, and in [M]—Music and Books on Music—from 24,453 to 11,728. Other subject areas experienced much more modest usage (see figure 4).

When the authors compiled the circulation of LCC and DDC classified titles, they found that the most highly circulated subjects fell within the [300]—Social Sciences with total 1,121,234 checkouts. Subjects that circulated least frequently (under ten thousand total circulations) are the following LCC subjects: [Z]—Bibliography, Library Science, etc., [C]—Auxiliary Sciences of History, [A]—General Works, [S]—Agriculture, [U]—Military Science, and [V]—Naval Science (see table 2).

Furthermore, the analyzed data showed a linear correlation between the number of items in the subject and the number of corresponding circulations. In figure 5, the scatter plot displays the relationship between two variables—a

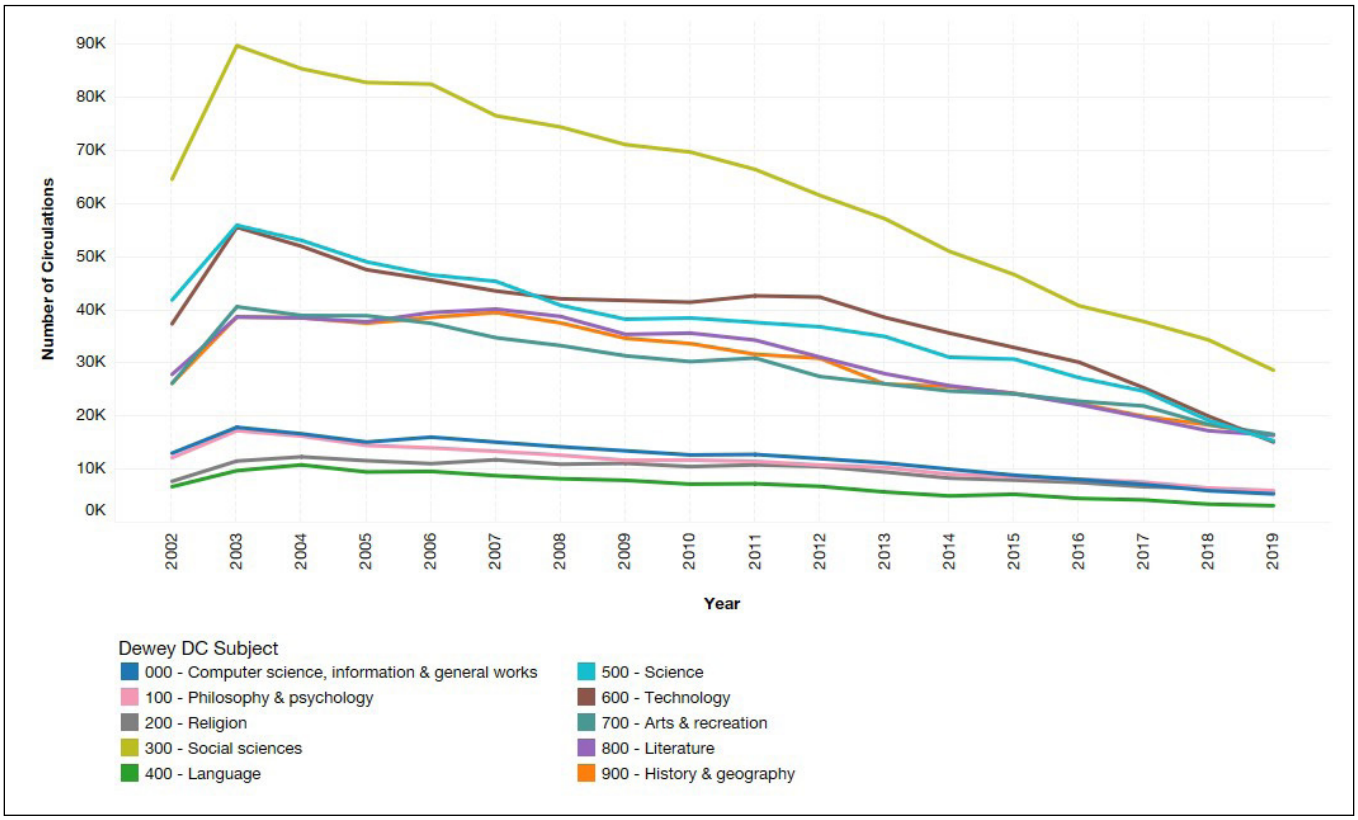


Figure 3. Circulations of the Dewey Decimal Classification Subjects by Year

total number of items in the subject and a total number of their circulations (see values for the variables in table 2). The straight line on the graph is a trend line, which demonstrates a positive linear correlation between the variables. The correlation coefficient is the measure of the strength of the relationship between variables and takes values between -1 and 1. For the authors' variables, the correlation coefficient is equal to 0.9, which indicates a strong relationship. Thus, it leads to the conclusion that the more items the subject collection offers for checkout, the more circulations the collection will show.

Additionally, some subjects experienced a greater total percentage of items circulated from within their subject areas. Not surprisingly, [E]—History of the Americas ranked highly with 66.1 percent of the volumes circulating. The next two highest subjects, however, were surprising as [R]—Medicine (56.4 percent) and [T]—Technology (51 percent) ranked quite high in terms of the percentage of the collection that circulated.

The frequency of item circulations, which is a ratio of the total number of circulations to the total number of circulated items, varied for each subject, and on average it ranged as low as 1.9 times per item for [A]—General Works and as high as 4.8 times for [R]—Medicine. The analysis revealed five subject collections with the highest percentage of checked out items at least once during the 2002–2019 period, and the highest circulation rate for those items in demand. The subjects are as follows: [E]—History of the Americas, [R]—Medicine, [T]—Technology, [Q]—Science, and [M]—Music and Books on Music, and all have a frequency of items circulations ranged on average between 4.3 and 4.8 checkouts per circulated item (see table 3).

Research Question 2: Is there a measurable difference in demand for these print resources from the periods before and after an item was digitized?

The results indicate that, when average annual usage is calculated, a measurable difference in demand appears for these print resources in the periods before and after their digitization. The total number of items digitized between 2010 and 2018 from the U of I Library's collection is 697,059. Almost half of the entire digitized collection falls in the years 2014 and 2018, with an annual total of 171,372 and 160,151 per each year, respectively. Nine percent (63,352 items) of the print counterparts of the digitized items showed evidence of circulations during 2002–2019,

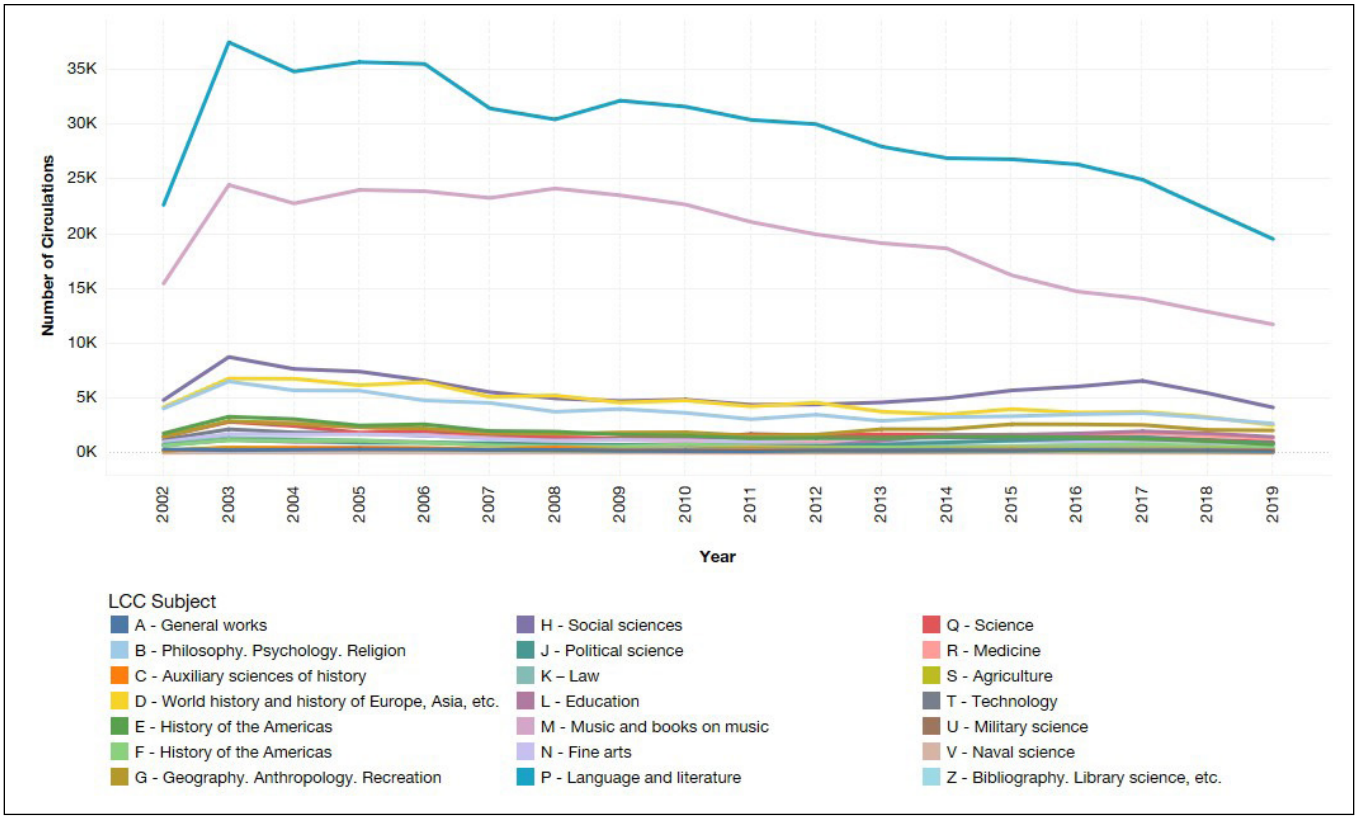


Figure 4. Circulation of the Library of Congress Classification Subjects by Year

with 102,540 total circulation transactions. Since 2012, less than 15 percent of items digitized in each year circulated (see table 4).

The decrease is evident when reviewing the average circulation per year for the pre- and post-digitization periods. For the pre-digitization period, the total average annual circulation for the print counterparts of the items digitized from 2010–2018 had 6,710 circulations. When reviewing the post-digitization period, the corresponding annual usage data is 2.72 times less, which is 2,464 circulations. By examining the 2010 digitization year, one can view data for digitized items with a balanced number of years before and after digitization, which is nine. The project team included circulations recorded in the year of digitization as part of the number of circulations pre-digitization. For example, the average number of circulations for items digitized in 2010 was calculated as a sum of circulations between 2002 and 2010 divided by nine years. The average number of circulations per year after digitization equaled a sum of circulations between 2011 and 2019 divided by nine years. Finally, the total average annual circulation for all digitized items was calculated as a sum of the average number of circulations per year for all nine years of digitization (see table 4).

One reason for the decreased demand for items following digitization can be an overall natural decline in the collection's circulation (see figure 6). The total number of annual item checkouts for the library's print collection fell from 482,206 in 2003 to 180,146 in 2019. At a rate of 2.68 times less, it is nearly the same as the decline for the average annual circulation of the print counterparts of all digitized items pre- and post-digitization. The reduced demand for the library's print collection might have resulted from the library's expanding collection of electronic subscriptions to various journals, databases, and other electronic resources. Their usage could have served to displace the use of some of library's print collections.

In total, annual circulation demand was higher for the pre-digitization period than for post-digitization. To gain a deeper perspective on this result, the authors will conduct a comparative analysis of the usage of digitized copies versus circulations of their print counterparts. Additionally, an analysis can be done by an item's publication year and subject to learn more about the library user's demand and interest for the items chosen for digitization.

Table 2. Total Number of Items and their Circulations by Subject, 2002–2019

Subject (Library of Congress & Dewey Decimal Classifications)	No. of Circulations	No. of Items in Subject	Ratio of Circulations to No. of Items
300—Social sciences	1,121,234	1,694,519	0.66
600—Technology	689,616	1,029,920	0.67
500—Science	666,961	747,657	0.89
800—Literature	550,965	996,036	0.55
900—History & geography	540,132	728,953	0.74
P—Language and Literature	526,647	349,638	1.51
700—Arts & recreation	524,631	463,503	1.13
M—Music and books on music	352,509	171,403	2.06
000—Computer science, information & general works	215,355	437,922	0.49
100—Philosophy & psychology	201,739	157,261	1.28
200—Religion	171,590	183,482	0.94
400—Language	123,469	153,879	0.8
H—Social sciences	101,747	92,528	1.1
D—World history and history of Europe, Asia, etc.	83,472	130,221	0.64
B—Philosophy. Psychology. Religion	71,861	72,019	1
G—Geography. Anthropology. Recreation	38,482	146,392	0.26
E—History of the Americas	32,052	10,610	3.02
Q—Science	29,814	12,735	2.34
R—Medicine	29,799	11,100	2.68
T—Technology	25,391	11,512	2.21
N—Fine arts	21,776	18,992	1.15
L—Education	20,221	15,125	1.34
J—Political science	16,852	28,784	0.59
K—Law	13,089	59,087	0.22
F—History of the Americas	12,882	8,601	1.5
Z—Bibliography. Library science, etc.	7,431	15,954	0.47
C—Auxiliary sciences of history	6,284	10,314	0.61
A—General works	4,455	31,681	0.14
S—Agriculture	4,387	4,353	1.01
U—Military science	3,873	3,329	1.16
V—Naval science	318	309	1.03

Research Question 3: Is there a difference in demand after digitization for those items that are freely available as full-text (most being pre-1923) when compared to those in which copyright or other restrictions limit the digital access?

Due to copyright limitations, most items published in 1923 and later are not available as full-text via HathiTrust. The results indicate that there is a difference in the local demand for the print counterparts of those items that are freely available as full-text (as defined by pre-1923 date of publication and an "allow" status) when compared to the ongoing demand for those published in 1923 and later. The results appear to confirm that an item's availability as a full-text resource corresponded to a more significant decrease in the use of its print counterparts. As noted above, the overall circulation rate appears to decline post-digitization. The average circulation per digitized item for those published after 1923 was lower than that for the pre-1923 publications.

In the authors' data set, the library held 697,059 items digitized between 2010 and 2018. To assign the digitized

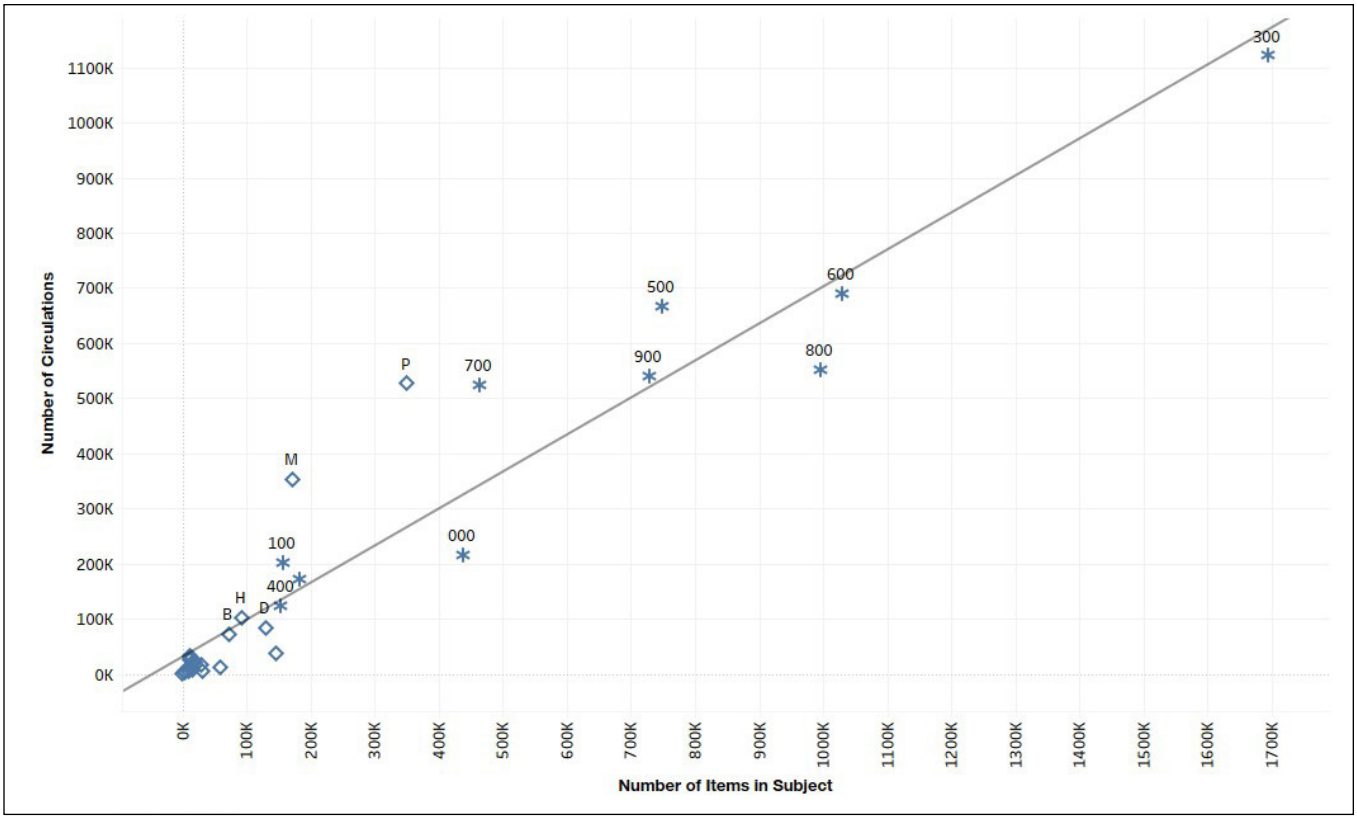


Figure 5. Correlation between Total Number of Items in the Subject and Total Number of their Circulations

item to one of the categories, such as "pre-1923 publications," "post-1923 publications," or "items with bad publication date," the *rights_date_used* field from the HathiTrust data set was used. Correcting for those items with a bad or incorrectly entered publication date reduced the sample pool by 34,139 items. Of that final body of 662,920 items, 36.7 percent, or 243,610, included pre-1923 publication dates, and 63.3 percent, or 419,310, were published in 1923 or later. Despite the difference of more than one and a half times in number of digitized titles from each publication period, both pre- and post-1923 print counterparts showed similar numbers of circulated items over an eighteen-year period from 2002 to 2019, which is 30,550 and 29,572, plus the number of circulations, 48,534 and 48,753, respectively. It follows that the percentage of circulated items is 12.5 percent for the titles published before 1923, and 7.1 percent for the titles with publication dates of 1923 or later. As for the frequency of items' circulations, the average number of circulations per one digitized item is 1.7 times higher for the pre-1923 titles than for post-1923 publications, 0.199 versus 0.116 (see table 5).

Comparing rates of circulation for pre- and post-1923 publications before and after digitization led the authors to speculate about the impact of full-text access. The HathiTrust dataset has an *access* field that indicates whether users can view the item. The field contains one of the following two values: "*allow*" when end users can view the item, and "*deny*"—when cannot. Of the final body of 243,610 digitized items published before 1923, 243,576 items, or 99.99 percent, have "*allow*" as the access value, and thirty-four items, or 0.01 percent, have "*deny*" as the access status. This means that nearly all pre-1923 publications in the analysis are available as full-text after digitization. Since a low percentage of the pre-1923 publications are not available for full-text access, they were not considered in the analysis. In the case of post-1923 publications, 170,978 of 419,310 digitized items, which is 41 percent, are available as full-text via HathiTrust Digital Library, and for 248,332 items, or 59 percent, copyright or other restrictions limit the digital access.

To identify if there is a difference in demand post-digitization, the authors compared the total average annual number of circulations before and after digitization for both pre- and post-1923 publication periods. For the items digitized in each of the 2010–2018 digitization years, the average annual number of circulations of their print counterparts equaled as a sum of circulations recorded prior to and including the year of digitization for the pre-digitization

Table 3. Percentage of Circulated Items (2002–2019) of Total Number of Items in the Subject

Subject (Library of Congress & Dewey Decimal Classifications)	% of Circulated Items	Frequency of Items Circulations
E—History of the Americas	66.1	4.6
R—Medicine	56.4	4.8
T—Technology	51.0	4.3
Q—Science	49.5	4.7
M—Music and books on music	45.2	4.6
L—Education	39.9	3.4
F—History of the Americas	39.3	3.8
100—Philosophy & psychology	37.6	3.4
U—Military science	36.1	3.2
V—Naval science	35.9	2.9
700—Arts & recreation	35.1	3.2
P—Language and literature	34.8	4.3
N—Fine arts	34.5	3.3
200—Religion	32.1	2.9
S—Agriculture	28.3	3.6
400—Language	27.7	2.9
H—Social sciences	27.4	4
900—History & geography	27.2	2.7
B—Philosophy. Psychology. Religion	26.7	3.7
D—World history and history of Europe, Asia, etc.	22.5	2.9
500—Science	22.1	4
800—Literature	21.6	2.6
300—Social sciences	21.0	3.2
J—Political science	20.0	2.9
600—Technology	18.0	3.7
Z—Bibliography. Library science, etc.	16.8	2.8
000—Computer science, information & general works	16.7	2.9
C—Auxiliary sciences of history	16.7	3.6
K—Law	9.5	2.3
G—Geography. Anthropology. Recreation	7.8	3.4
A—General works	7.5	1.9

period and as a total of circulations after the year of digitization for the post-digitization period divided by the corresponding number of years participated in the calculation. The total average annual number of circulations is a sum of all average annual number of circulations for all years of digitization. Thus, for the pre-1923 publications with full-text available via HathiTrust, the total average annual number of circulations equaled 3,313 before digitization and decreased by more than three times after digitization to 1,014 circulations (see table 6). The post-1923 publications that are also available as full-text after digitization showed a drop in total average annual number of circulations as well, from 960 circulations in the pre-digitization period to 178 circulations in the post-digitization period, which is a 5.4 decrease. For post-1923 publications with copyright or other restrictions limiting their digital access, the decline in circulations was not as steep, with 2,102 circulations before digitization versus 1,093 afterwards. This is less than by 2 times (see table 7). Thus, considering that the items from the same publication period, which is post-1923, having only limited viewing rights, show a different circulation decrease rate after digitization, which is three times as much for items whose full-text is available after digitization compared with those with restricted full-text. This led the authors to conclude that users chose electronic over print. To further establish the nature of this relationship, the authors plan to conduct further research that would include circulation data not just for print counterparts of the digitized items, but also the usage data for the electronic items. The usage information for the digitized copies will show if users had checked them out. Moreover, a general drop in the number of circulations after digitization for both pre- and post-1923 publications might be associated with an overall reduced demand for the library's print collection.

Conclusion

To complete a study of the impact of digitization on the circulation of printed items in a research library's collection, one needs to compile information on the items in the collection, their digitization status, and their recorded circulation information. Many of the systems that libraries use to maintain or gather these data sets do not directly interface with one another. In this study, the research team needed to compile four different data sets that included not only the identifying information for more than 10 million items, but digitization histories for 847,247 items and circulation transaction logs that tracked 8,703,606 individual transactions over an eighteen-year period. With respect to the questions about the impact of

Table 4. Print Items Circulations (2002–2019) by Year of Digitization

Year of Digitization	Total			Before Digitization			After Digitization		
	No. of Digitized Items	% of Circulated Items	No. of Circulations	No. of Years	Total No. of Circulations	Avg. No. of Circulations per Year	No. of Years	Total No. of Circulations	Avg. No. of Circulations per Year
2010	6,845	49.5	4,453	9	4,230	470	9	223	25
2011	136	39.7	76	10	69	7	8	7	1
2012	72,862	5.2	4,968	11	4,813	438	7	155	22
2013	8,824	14.1	1,508	12	1,459	122	6	49	8
2014	171,372	8.2	22,327	13	18,986	1,461	5	3,341	669
2015	97,909	12.3	19,738	14	18,114	1,294	4	1,624	407
2016	102,368	10.5	18,123	15	16,627	1,109	3	1,496	499
2017	76,592	6.1	7,516	16	7,063	442	2	453	228
2018	160,151	8.3	23,831	17	23,228	1,367	1	603	605
Total	697,059		102,540		94,589	6,710		7,951	2,464

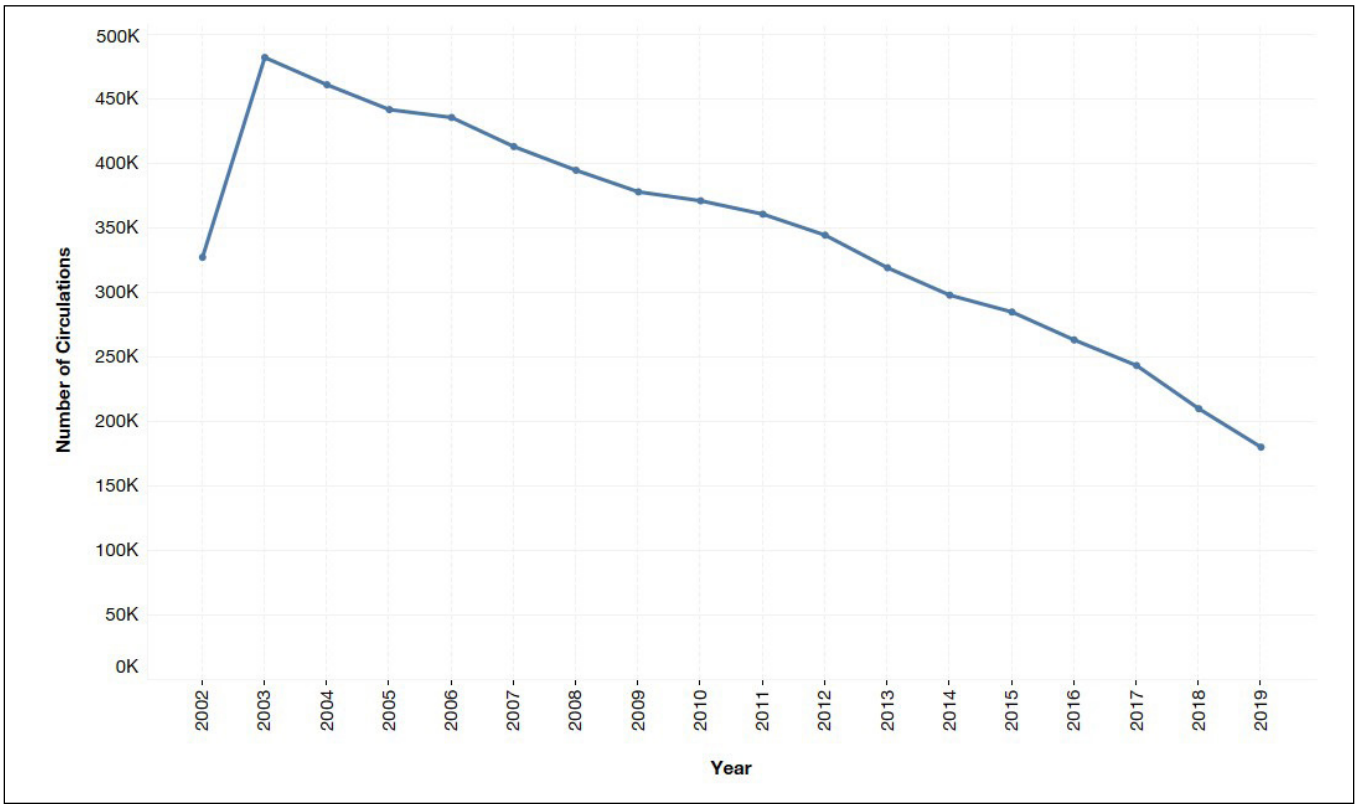


Figure 6. Annual Number of Circulations for Library's Print Collection

digitization on the circulation of printed items in a research library's collection, the conclusion from the data provided seems to indicate that there is a diminished amount of annual average usage for items in the periods after their digitization. With respect to differences in the demand for pre-1923 and post-1923 publications after digitization,

Table 5. Number of Digitized Items and Circulations by Publication Period

Publication Period	No. of Digitized Items	No. of Circulated Items	% of Circulated Items	No. of Circulations	Circulations per one Digitized Item
Pre-1923 publications	243,610	30,550	12.5	48,534	0.199
Post-1923 publications	419,310	29,572	7.1	48,753	0.116

Table 6. Number of Digitized Items and Circulations by Year of Digitization for Pre-1923 Publications

Year of Digitization	Pre-1923 Publications											
	Total		Before Digitization					After Digitization				
	No. of Digitized Items		# years	No. of Circulations with Access "Allow"		No. of Circulations with Access "Deny"		# years	No. of Circulations with Access "Allow"		No. of Circulations with Access "Deny"	
	With Access "Allow"	With Access "Deny"		Total	Avg. Annual	Total	Avg. Annual		Total	Avg. Annual	Total	Avg. Annual
2010	6,242	-	9	3,766	418	-	-	9	176	20	-	-
2011	132	-	10	65	7	-	-	8	7	1	-	-
2012	11,434	-	11	1,004	91	-	-	7	114	16	-	-
2013	3,470	-	12	1,188	99	-	-	6	41	7	-	-
2014	84,332	1	13	11,542	888	-	-	5	1,810	362	-	-
2015	74,452	6	14	14,350	1,025	6	0	4	1,150	288	0	0
2016	20,095	27	15	3,967	264	8	1	3	362	121	0	0
2017	17,203	-	16	2,773	173	-	-	2	187	94	-	-
2018	26,216	-	17	5,913	348	-	-	1	105	105	-	-
Total	243,576	34		44,568	3,313	14	1		3,952	1,014	0	0

the evidence points to a greater level of demand on print counterparts for items with restricted access. Overall, while there are significant differences in the demand on print resources by subject area, ascertaining whether the differences result from their digitization remains impossible at this point.

In comparison, it is possible to examine overall trends. The evidence thus far points to a marked decline in usage for print counterparts of the digitized items—a presumed confirmation of much speculation from years past—and a confirmation that existing print stocks can likely serve broader populations of users when borrowing networks resume regular operations. Among items that cannot be factored into this data are changes in the scholarly demand for resources or subjects, the impact of the digital availability from other, commercial sources such as e-book backfile packages on the use of individual titles or within disciplinary areas, or how those who used digital surrogates interacted with the resources (using online, printing, etc.). However, the evidence does point to a decline in usage post-digitization as a general trend.

Additionally, this data does not further the understanding of how the availability of access to items digitized and shared via HathiTrust might impact both local circulation and the rate of interlibrary loan and document delivery for such items. Determining the impact on the borrowing/lending behaviors of local communities is a critical step in determining how our institutions might approach the management of these collections in the future.

What this means for libraries and scholars is unclear. Some will look at this selective set of data and assume that collections can be managed more aggressively, lending credence to those concerned in the scholarly community that libraries are not stewarding our cultural heritage. Others will view the data as incomplete or flawed, using it to support stonewalling local and collective efforts to rationally manage low-use collections that occupy significant portions of campus buildings where broader bodies of students and scholars may benefit from direct access to other services. In the end, the findings can point us in directions, to encourage the scholarly community to sharpen its arguments about the value of preserving elements of our shared cultural

Table 7. Number of Digitized Items and Circulations by Year of Digitization for Post-1923 Publications

Year of Digitization	Pre-1923 Publications											
	Total		Before Digitization					After Digitization				
	No. of Digitized Items		# years	No. of Circulations with Access "Allow"		No. of Circulations with Access "Deny"		# years	No. of Circulations with Access "Allow"		No. of Circulations with Access "Deny"	
	With Access "Allow"	With Access "Deny"		Total	Avg. Annual	Total	Avg. Annual		Total	Avg. Annual	Total	Avg. Annual
2010	196	260	9	188	21	194	22	9	11	1	34	4
2011	1	3	10	1	0	3	0	8	0	0	0	0
2012	55,209	2,815	11	3,183	289	454	41	7	32	5	6	1
2013	4,566	142	12	167	14	55	5	6	3	1	0	0
2014	19,962	55,860	13	1,214	93	4,936	380	5	107	21	1,205	241
2015	2,377	18,140	14	818	58	2,277	163	4	43	11	324	81
2016	16,769	62,206	15	2,934	196	9,055	604	3	144	48	925	308
2017	20,995	36,110	16	818	51	3,143	196	2	30	15	214	107
2018	50,903	72,796	17	4,054	238	11,754	691	1	76	76	351	351
Total	170,978	248,332		13,377	960	31,871	2,102		446	178	3,059	1,093

heritage without advocating that the community of research libraries tackle the impossible by preserving everything, to support collection stewards as they seek to manage their collections, and to further the discourse around how we curate these resources.

References and Notes

1. Oya Y. Rieger, *Preservation in the Age of Large-Scale Digitization: A White Paper* (Washington, DC: Council on Library and Information Resources, 2008), https://www.clir.org/wp-content/uploads/sites/6/pub141.pdf. This assertion also appears in Tony Horva, "Challenges and Possibilities for Collection Management in a Digital Age," *Library Resources & Technical Services* 54, no. 3 (2010): 147.
2. Chandra Prabha, "Shifting from Print to Electronic Journals in ARL University Libraries," *Serials Review* 33, no. 1 (2007): 4–13, https://doi.org/10.1080/00987913.2007.10765086.
3. Rick Anderson, "Less Than Meets the Eye: Print Book Use Is Falling Faster in Research Libraries," Scholarly Kitchen, August 21, 2017, https://scholarlykitchen.sspnet.org/2017/08/21/less-meets-eye-print-book-use-falling-faster-research-libraries/. This article examines ARL data from 1995 to 2008. The ARL Statistics for more recent years reflect this ongoing trend with mean circulation declining almost every year. More recently, popular periodicals discussed this trend. See Dan Cohen, "The Books of College Libraries are Turning into Wallpaper," *The Atlantic*, May 26, 2019, https://www.theatlantic.com/ideas/archive/2019/05/college-students-arent-checking-out-books/590305/.
4. Roger Schonfeld and Ross Housewright, *What to Withdraw? Print Collections Management in the Wake of Digitization* (New York: Ithaka S+R: 2009), https://sr.ithaka.org/wp-content/uploads/2015/08/What_to_Withdraw_Print_Collections_Management_in_the_Wake_of_Digitization.pdf.
5. Lorcan Dempsey et al., *Operationalizing the BIG Collective Collection: A Case Study of Consolidation vs. Autonomy* (Dublin, OH: OCLC Research, 2019), https://doi.org/10.25333/jbz3-jy57.
6. Information about the HathiFiles datasets may be found here: https://www.hathitrust.org/hathifiles.
7. Mary Biggs, "The Proposed National Periodicals Center, 1973–1980," *Resource Sharing & Information Networks* 1, nos. 3–4 (1984): 1–22, https://doi.org/10.1300/J121v01n03_01.
8. The Center for Research Libraries published the collected papers of the 2003 PAPR conference in *Library Collections, Acquisitions, & Technical Services* 28, no. 1 (2004).
9. *Print Archiving and Shared Print in North America: A Preliminary Analysis and Status Report* (Chicago: Center for Research Libraries, 2015), http://www.crl.edu/sites

/default/files/attachments/events/PAPR_summit_preliminary_analysis2_revised.pdf.

10. Robert H. Kieft and Lizanne Payne, "A Nation-Wide Planning Framework for Large-Scale Collaboration on Legacy Print Monograph Collections," *Collaborative Librarianship* 2, no. 4, (2010): 229–33, https://digitalcommons.du.edu/collaborativelibrarianship/vol2/iss4/8; Jacob Nadal, Annie Peterson, and Dawn Aveline, "Scarce and Endangered Works: Using Network-Level Holdings Data in Preservation Decision-Making and Stewardship of the Printed Record," accessed January 30, 2019, http://www.jacobnadal.com/wp-content/uploads/2011/05/ScarceAndEndangeredWorksv7.pdf; Constance Malpas, *Cloud-Sourcing Research Collections: Managing Print in the Mass Digitized Library Environment* (Dublin, OH: OCLC Research, 2011), https://www.oclc.org/content/dam/research/publications/library/2011/2011-01.pdf.
11. HathiTrust Collections Committee, *HathiTrust Distributed Print Monographs Archive Proposal* (Ann Arbor, MI: HathiTrust, 2011), https://www.hathitrust.org/constitutional_convention2011_ballot_proposals#proposal; HathiTrust Print Monograph Archive Planning Task Force, *HathiTrust Print Monographs Archive Planning Task Force: Final Report* (Ann Arbor: HathiTrust, 2015), https://www.hathitrust.org/files/sharedprintreport.pdf.
12. Future of the Print Record Working Group, *Concerted Thought, Collaborative Action, and the Future of the Print Record: A White Paper* (New York: Modern Language Association, 2016), https://printrecord.mla.hcommons.org/concerted-thought-collaborative-action-and-the-future-of-the-print-record/.
13. Lura E. Joseph, "Image and Figure Quality: A Study of Elsevier's Earth and Planetary Sciences Electronic Journal Back File Package," *Library Collections, Acquisitions, & Technical Services* 30, nos. 3–4 (2006): 162–68, https://doi.org/10.1016/j.lcats.2006.12.002; Lura E. Joseph, "Improving the Quality of Online Journals: Follow-up Study of Elsevier's Backfiles Image Rescanning Project," *Library Collections, Acquisitions, and Technical Services* 36, nos. 1–2 (2012): 18–23, https:/doi.org/10.1016/j.lcats.2011.08.001.
14. Paul Conway, "Preserving Imperfection: Assessing the Incidence of Digital Imaging Error in HathiTrust," *Preservation, Digital Technology & Culture* 42, no. 1 (2013): 17–30, https://doi.org/10.1515/pdtc-2013-0003.
15. Andrew Stauffer, "My *Old Sweethearts:* On Digitization and the Future of the Print Record," in *Debates on the Digital Humanities* (2016): 218–29, http://dhdebates.gc.cuny.edu/debates/text/70; Jennifer Hain Teper, "Considering 'Sameness' of Monographic Holdings in Shared Print Retention Decisions," *Library Resources & Technical Services* 63, no. 1 (2019): 29–45, https://doi.org/10.5860/lrts.63n1.29.
16. Sara Gould and Richard Ebdon, ed., *IFLA/UNESCO Survey on Digitization and Preservation,* International Preservation Issues, no. 2 (Wetherby, UK: International Federation of Library Associations and Institutions, 1999).
17. Gould and Ebdon, *IFLA/UNESCO Survey on Digitization and Preservation*, 30.
18. Council on Library and Information Resources, *Scholarship, Instruction, and Libraries at the Turn of the Century: Results from Five Task Forces Appointed by the American Council of Learned Societies and the Council on Library and Information Resources* (Washington, DC: Council on Library and Information Resources, 1999), https://clir.wordpress.clir.org/wp-content/uploads/sites/6/pub78.pdf.
19. Abby Smith, *The Future of the Past: Preservation in American Research Libraries* (Washington, DC: Council on Library and Information Resources, 1999), https://clir.wordpress.clir.org/wp-content/uploads/sites/6/pub82.pdf.
20. Information on the HathiFiles accessed July 31, 2020, https://www.hathitrust.org/hathifiles_description.
21. Anderson, "Less Than Meets the Eye."
22. Alex Verstaket al., "On the Shoulders of Giants: The Growing Impact of Older Articles," *arXiv.org*, 2014, https://arxiv.org/abs/1411.0275.

On the State of Genre/Form Vocabulary

A Quantitative Analysis of LCGFT Data in WorldCat

Colin Bitter and Yuji Tosaka

The purpose of this paper is to report on a quantitative analysis of the LCGFT vocabulary within a large set of MARC bibliographic data retrieved from the OCLC WorldCat database. The study aimed to provide a detailed analysis of the outcomes of the LCGFT project, which was launched by the Library of Congress (LC) in 2007. Findings point to a moderate increase in LCGFT use over time; however, the vocabulary has not been applied to the fullest extent possible in WorldCat. Further, adoption has been inconsistent between the various LCGFT disciplines. These and other findings discussed here suggest that retrospective application of the vocabulary using automated means should be investigated by catalogers and other technical services librarians. Indeed, as the data used for the analysis show somewhat uneven application of LCGFT, and with nearly half a billion records in WorldCat, it remains a certainty that much of LCGFT's full potentials for genre/form access and retrieval will remain untapped until innovative solutions are introduced to further increase overall vocabulary usage in bibliographic databases.

When the Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT) project began in 2007, the principal aim was to develop a vocabulary separate from Library of Congress Subject Headings (LCSH) to describe what a resource is rather than what it is about.[1] While LCSH has been also used to describe "is-ness" for decades in certain situations, there were several problems with using LCSH terms to describe genre and form.[2] Through the efforts of the Library of Congress (LC) partnering with various parties in the greater cataloging community, the LCGFT project has been successful in establishing a separate vocabulary that is both broad and deep. As of March 2020 (when the data were compiled for this study), 2,357 terms are organized under eight disciplines (art, cartography, law, literature, moving images, music, religion, and non-musical sound recordings) plus "general library materials," with twenty-one "top terms" that have other narrower terms organized hierarchically in each category.

The purpose of this paper is to conduct a quantitative analysis of a large set of MARC bibliographic data retrieved from the OCLC WorldCat database (henceforth WorldCat). Previous publications about LCGFT have been primarily limited to providing a broad overview of the history of genre and form and establishing a clear need for a robust genre/form vocabulary, while some have also outlined the process to create the new vocabulary. What is lacking in the literature is a detailed analysis of the outcomes of the LCGFT project within

Colin Bitter (bitterc1@tcnj.edu) is the Head of Cataloging and Metadata at The College of New Jersey. **Yuji Tosaka** (tosaka@tcnj.edu) is Cataloging/Metadata Librarian at The College of New Jersey.

Manuscript submitted January 26, 2021; returned to authors for minor revision March 10, 2021; revised manuscript submitted March 13, 2021; accepted for publication March 18, 2021.

bibliographic databases. Although the primary focus of the following study is on LCGFT terms recorded in MARC field 655 subfield $a, multiple data points within the records are used for the authors' analysis. Filling a clear gap in the literature, such quantitative analysis will provide a broad overview as to the state of LCGFT usage within a shared cataloging environment, and will make significant contributions related to multiple library stakeholders. It will give catalogers a much better, empirical understanding of the extent to which LCGFT has been applied within MARC bibliographic records. Additionally, detailed analysis of the vocabulary usage will offer insights into future cataloging practices and training needs in the technical services community. This paper's findings will also offer useful insights for public services librarians, as they will benefit from learning in depth about patterns of LCGFT application in bibliographic databases for their work with users to help them navigate front end systems utilizing such data for improved resource discovery.

Literature Review

The question of providing access to genre and form information in library catalogs has not received much attention in the library literature, although it has been long recognized as one of the key intellectual foundations of information organization. In his influential *Rules for a Printed Dictionary Catalogue*, Cutter noted that a key objective of the catalog was the collocating function, that is, enabling users to discover all resources in a particular genre or form of material, and by author and subject.[3] Genre and form are also an essential part of the bibliographic universe as defined in the current IFLA Library Reference Model [e.g., LRM-E2-A1: *Category* attribute].[4] For many years, some limited access to the genres and forms found in library collections had been provided by LCSH, either as main headings or subdivisions, although their primary function was always to describe the content of the work (aboutness). By the end of the twentieth century, more recent developments brought increasing attention to the genre/form access question, with the creation of GSAFD (Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc.) genre terms and the implementation of a new MARC subfield $v for "form" subdivisions in 6XX fields.[5] At the same time, LC announced its plan to develop new genre/form headings. And yet, it was not until 2007 that LC finally developed the new LCGFT thesaurus, starting with moving image materials and radio programs.[6] More than a decade after its inception, LCGFT has developed into a more fully fledged controlled vocabulary for genre and form access covering nine disciplines, including "general" materials.[7]

The steady development of the LCGFT thesaurus, however, has not yet yielded a new stream of scholarship on genre/form access in the cataloging literature, although there are several studies that have begun to look into the subject over the last decade. Perhaps the most important overview of the historical literature was provided by Lee and Zhang's 2013 paper in *Cataloging & Classification Quarterly*. The authors traced how genre and form terms had been conceptualized and treated in Anglo-American cataloging standards up to the implementation of RDA. Their comprehensive examination showed that genre had not been given the attention it deserved in the cataloging literature, despite the "expanding role genre plays in the current as well as future environments." Notably, the authors also concluded that the cataloging community had failed to establish clear definitions differentiating "genre" and "form."[8] This conceptual ambiguity is reflected in the current LCGFT manual, which defines both genres and forms as follows:

> Genres and forms may be broadly defined as categories of resources that share known conventions. More specifically, genre/form terms may describe the purpose, structure, content, and/or themes of resources.[9]

While other authors have also investigated issues relating to genre and form access in specific subject areas and specialist communities, such as audiovisual cataloging, there have been few published studies focusing on the LCGFT thesaurus itself.[10] Those few publications include Young and Mandelstam's 2013 paper in *Cataloging & Classification Quarterly*, in which they discussed, in addition to introducing the reader to its potential benefits and applications, how the LCGFT thesaurus was developed, often involving formal collaboration between LC cataloging policy specialists and outside library organizations.[11] Iseminger and others have also considered LCGFT development and applications in specialist communities, such as music cataloging.[12] As adding LCGFT headings to legacy metadata is clearly a very important step in fully realizing the benefits of the new vocabulary, Mullin examined the process for automatically assigning them for music resources retrospectively based on the presence of LCSH terms in their bibliographic records.[13]

Now that more than a decade has passed since the LCGFT thesaurus first became available for use in the library community, recent literature has finally started analyzing data on how the LCGFT thesaurus has been deployed in library catalogs and digital repositories. In 2018, Dragon contacted twenty-nine digital repositories in North American academic libraries and examined how they provided genre and form access for their digital collections, using such display labels as "Format," "Type," and "Genre."

For specific vocabularies being used, she found that DCMI (Dublin Core Metadata Initiative) Type Vocabulary and the Art and Architecture Thesaurus were both most widely used, while LCGFT was used at only two of the repositories that she contacted.[14] In contrast, Bitter and Tosaka decided to focus specifically on the usage of LCGFT headings in library catalogs and conducted a survey that revealed that the new thesaurus had gained wide, if somewhat uneven, adoption in the cataloging community. The survey data reported in their paper provided valuable insights into how the LCGFT thesaurus was currently used in copy and original cataloging practices and which types of resources were more likely to have their bibliographic records enriched with LCGFT terms.[15] Whereas these newer studies serve as good starting points for examining current LCGFT implementation, what is sorely needed in the literature is detailed research on how LCGFT terms have actually been deployed in bibliographic databases, such as local catalogs or WorldCat. This paper's intent is to make a significant contribution to studies on genre and form access by conducting a quantitative analysis of LCGFT usage patterns in selected MARC records retrieved from WorldCat.

Research Method and Data Retrieval

To fill the critical gap in the literature described above, this paper explores several areas of inquiry. Most broadly, dates within bibliographic records are used to investigate rates of LCGFT application over time. Second, format of material is examined to differentiate LCGFT use between different types of records—for example, does notated music contain more LCGFT than projected media? Third, records are grouped by LC classification (LCC) to examine patterns of LCGFT usage in terms of pre-existing classification—do certain areas of LCC see greater use of LCGFT? Finally, LCGFT terms applied within bibliographic records are delineated to analyze the extent to which broader or narrower headings have been used in terms of the hierarchies in LCGFT.

To explore the research questions outlined above, the authors examined MARC bibliographic record data retrieved from WorldCat. As WorldCat is a shared cataloging environment with close to half a billion records used by thousands of OCLC member libraries, analyzing portions of data from this database provided much needed empirical insights into the current state of LCGFT usage in the cataloging community. Although there were many possibilities for record selection, the approach selected for the current study was to examine WorldCat records based on the holdings of the authors' institution, The College of New Jersey, a mid-sized four-year public college in Ewing, New Jersey. The college is a comprehensive institution enrolling approximately 7,000 undergraduate students in a wide range of disciplines, and also offers master's and post-baccalaureate programs for over 600 students in a small number of graduate programs, such as Business, Counseling, Education, English & Humanities, Integrative STEM, and Nursing & Public Health. The authors' library is a typical academic library for a medium-sized institution. The only library serving the campus community, it holds over half a million titles in its physical collections, divided into seven main areas, including Archives, Children's/Young Adult, Curriculum & K-12, General, Music & Media, Periodicals, and Reference. The library also directly manages over 350,000 electronic titles, spread across various electronic collections. The vast majority of MARC records for both the physical and electronic collections are cataloged in WorldCat.

The authors believed that performing an analysis on this set of records selected from WorldCat would lead to a good snapshot of the current patterns of LCGFT usage within bibliographic records typically used by academic institutions. That is, overall patterns of LCGFT usage can be better inferred from this record set since the vast majority of these records are selected via copy cataloging from WorldCat and the authors have made efforts to include only high-quality best matches in their local catalog. That would contrast with analyzing the entire WorldCat database, which would contain a plethora of duplicates, to say nothing of lower quality bibliographic records that the authors feared would make their analysis much more complicated than necessary. Additionally, they decided to avoid analyzing bibliographic records in their local catalog for the obvious reason that those records do not include changes, including LCGFT headings added, since they were last copy-cataloged from WorldCat.

To obtain WorldCat master record data for their library's institutional holdings, the authors first turned to OCLC WorldShare Collection Manager, a cloud-based application designed to promote efficiencies in managing metadata for print and electronic collections held by OCLC member libraries. The feature used in Collection Manager was "query collection," which enabled the authors to retrieve master records for all of their library's local holdings. Using query collection was straightforward, as only a single criterion needed to be specified in the query, "li:NJT," which limited the resulting collection to holdings based on their library's OCLC symbol. Several files retrieved contained their library's entire institutional holdings, which totaled 846,862 records. It initially appeared as if this data could be used for the present study; however, authentic dates and times of latest transactions were not recorded in the MARC field 005. Each field 005 in the retrieved records in the query collection contained the same calendar date, "20200321 . . ." followed by hours, minutes,

seconds, and fractions of seconds, e.g., 20200321084945.4, that is, the download date of each record as WorldCat apparently considers this to be a record transaction date. This seemed to diminish the utility of the retrieved data for the authors' intended analysis because they had expected that the replace date of each record might be queried to expose varying rates of vocabulary application over time. Email communications with OCLC support representatives confirmed that the authentic replace date in the field 005 could not be retrieved via query collection. Although the retrieved data thus could not be used as originally planned, the authors set aside the 846,862 unique OCLC numbers contained in the collection as the basis for future data retrievals, as described below.

The authors decided to use the OCLC Bib API integration in MarcEdit, a freely available leading MARC data editing tool developed by Terry Reese, who is the Head of Digital Initiatives and Infrastructure Support at The Ohio State University Libraries. This tool enables users to retrieve WorldCat master records by OCLC number, ISBN, ISSN, or Title/Author. To use the OCLC Bib API, the authors first needed to contact OCLC to obtain API keys, which were then recorded in MarcEdit. Once the integration was established, it was then possible to use MarcEdit's OCLC Record Downloader and extract the needed MARC records using OCLC control numbers. For the present study, there were two major advantages to using the OCLC API integration in MarcEdit. First, the records delivered contained the authentic field 005, i.e., the last replace date in WorldCat. Second, much larger batches could be processed (the authors generally retrieved 50,000 records during a single session), thereby eliminating the ceiling of 9,999 records that would have been possible via batch searching in Connexion. Despite these advantages, the OCLC API also presented some drawbacks. First, it was highly error-prone—that is, the downloader would typically fail to retrieve every MARC record matching the OCLC number specified in the search. Therefore, it was necessary to cross-check the OCLC numbers in the resultant download file against the original query and then retrieve missing records in a quick follow-up session. The WorldCat master records matching all of the authors' institution's holdings were successfully retrieved in sets of 50,000 records each between April 24 and May 2, 2020. They were combined into a single file of 846,862 records (henceforth referred to as the base file), the contents of which are analyzed in the Analysis section that follows.

Beyond generating this base file, LCGFT terms from the vocabulary itself required organization for the present study. Two files of LCGFT terms were prepared, based on data compiled using *Classification Web* as of February 25, 2020, which were then brought up to date in early May with *Library of Congress Subject Headings Monthly List* 03 (March 16, 2020). The first file created, LCGFT-1, was a single list containing all unique LCGFT terms (2,357 terms). In compiling the LCGFT-1 file, the authors also divided all the LCGFT terms into four levels of hierarchy by applying numbers 1, 2, 3, and 4 to each term based on their hierarchical relationships. That is, 1 was the highest level assigned to the broadest terms (e.g., "Art"), while 4 was the lowest level assigned to more specific, narrower terms (e.g., "Pageants"). These scores were assigned in the LCGFT-1 file so that information about aggregate depth of indexing also could be garnered for LCGFT headings used in bibliographic records. The LCGFT Manual instructs catalogers to "assign terms that are as specific as the genres and forms exemplified in a resource" and some disciplines, such as music, have a well-developed hierarchy of LCGFT headings.[16] The authors recognized that a broader term may be used instead under certain situations (e.g., when a given term may subsume several narrower genre and form terms). They were thus interested as part of their data analysis in identifying the extent to which narrower, specific terms had been assigned in WorldCat records as they evaluated the overall patterns of LCGFT application. Four levels of hierarchy were chosen for the current study as deeper levels of specificity (e.g., fifth and sixth levels) did not seem too productive for analysis. Additionally, as LCGFT is polyhierarchical (i.e., some terms belong to more than one broader discipline, sometimes at different levels), it was necessary to find a consistent way of applying hierarchy levels to terms occurring in multiple LCGFT disciplines and/or at multiple levels of hierarchy. For the purposes of this study, LCGFT terms were coded in the highest possible hierarchy for each discipline. "Loose-leaf services" is an apt example as it occurs at different levels, under both top terms "Law Materials" and "Informational Works." Under "Law Materials," "Loose-leaf services" would be coded 2 since it is a second-level term. Under "Informational Works," it would be coded 3 since it is a third-level term. In the combined LCGFT-1 file, "Loose-leaf services" was coded 2.

The authors also created the second file, LCGFT-2, containing twenty-one separate lists for each of the LCGFT subject categories (art, cartographic materials, commemorative works, creative nonfiction, derivative works, discursive works, ephemera, illustrated works, informational works, instructional and educational works, law materials, literature, motion pictures, music, recreational works, religious materials, sound recordings, tactile works, television programs, video recordings, and visual works). LCGFT terms in these separate lists were also given annotations for depth of indexing respectively, with 1, 2, 3, and 4 assigned in the same fashion as in LCGFT-1. LCGFT-2 was used to examine prevalence and depth of indexing of the vocabulary used in each category, as will be discussed below in the Analysis section.

Findings and Analysis

Date

During the planning phase of the present study, the authors had expected that the MARC field 005 (Date and Time of Latest Transaction) would prove to be a useful data point in analyzing LCGFT usage in WorldCat. That is, as this field functions as a replace date in WorldCat, examining LCGFT usage against field 005 might provide interesting insights into changing rates of vocabulary application. Though they recognized that LCGFT headings for various disciplines were introduced over different years, the year 2007 was chosen as the point of demarcation for this simple exploratory analysis on the grounds that the LCGFT thesaurus was first established in that year. However, the base file used for analysis (864,862 records, as described in Research Methods and Data Retrieval) revealed that all records had been replaced within the last seven years. The oldest field 005 was dated June 6, 2013. Although many of these records would have been upgraded manually by catalogers within Connexion or by various OCLC member libraries via automated means (such as datasync, which automatically generates a new field 005), other records would also have been updated by WorldCat's internal automated processes, such as addition of RDA 33X fields or FAST subject headings.[17] Indeed, as the field 005 did not extend beyond the past seven years, the field was found to be effectively unusable for the intended analysis.

An alternative to replace date that was identified for the authors' analysis was the "Date 1" fixed field, available in the field 008 positions 07-10. Nearly all of the records in the base file had usable Date 1 data. However, some records had to be expunged due to incompleteness (for example, uuuu, ||||, 0002, and similar non-usable data values). After eliminating these records, 838,875 records (99.1 percent) remained and were used for this area of the analysis. Whereas the exact meaning of Date 1 data can vary based on the coding of the DtSt fixed field (Type of Date/Publication Status—008 position 06), the vast majority can be accurately linked to the manifestation being cataloged, be it in form of the year of production, publication, distribution, release, manufacture, or copyright as specified in the code in DtSt.

Using Date 1 values in the base file, the records were divided into two groups: before 2007 and 2007 to the present. Records prior to 2007 numbered 640,449 (76.3 percent of the base file); records from 2007 to the present numbered 198,426 (23.7 percent). First, examining these two sets of records for LCGFT application showed some increase in the latter group, which is not surprising given that LCGFT was not available for use before 2007. As seen in figure 1, 144,045 (22.5 percent) pre-2007 records and 58,489 (29.5 percent) records from 2007–present contained one or more LCGFT terms. Additionally, the average number of LCGFT terms for records containing LCGFT increased slightly, from 1.34 to 1.50. Date 1 values will continue to serve as a point of illumination in the sections that follow.

Format

Format of material was also examined to find disparities in LCGFT application, if any, between various types of resources. From the entire base file, 205,879 records had one or more 655 fields containing subfield $2 lcgft, representing 24.3 percent of all the records under examination. From this set of records, type of record (Leader position 06) was retrieved to examine the format of material described by each record. Table 1 illustrates the proportion of records containing LCGFT based on type. (As there were few resources coded as kit, manuscript cartographic material, manuscript notated music, mixed materials, and three-dimensional artifact or naturally occurring object, these formats are omitted in table 1 as they are not substantively significant for the purpose of this analysis.) Here it is worth noting the high rates of LCGFT application for a handful of format types. Indeed, over half of the records for five types contained one or more LCGFT terms: manuscript language materials (96.2 percent), projected media (88.3 percent), cartographic materials (65.8 percent), notated music (53.5 percent), and two-dimensional nonprojectable graphics (50.7 percent). In contrast, less than half of the records contained LCGFT for musical sound recordings (40.8 percent), nonmusical sound recordings (19.5 percent), language materials (18.8 percent), and computer files (13.1 percent).

Comparing the pre-2007 and 2007–present record sets revealed some other interesting data on changes in LCGFT application across format types. Of 202,534 records containing LCGFT and a valid Date 1 value (as described earlier in the *Date* section), 144,045 (71.1 percent) were pre-2007 and 58,489 (28.9 percent) were 2007–present. These two sets of records were compared against all the records in the base file containing valid dates (divided into two files, pre-2007 and 2007–present) to measure changes in LCGFT application over time. With the exception of musical sound recordings, all types showed an increase in LCGFT application in the 2007–present set, as evidenced in figure 2. The most significant increases were found in notated music (a 34.1 percent increase, from 52.4 percent to 86.5 percent), two-dimensional nonprojectable graphics (21.6 percent, from 44.3 percent to 65.9 percent), nonmusical sound recordings (20.7 percent, from 18.1 percent to 38.8 percent), projected media (15.8 percent, from 83.8 percent to 99.6 percent), and cartographic materials (12.9 percent, from 58.9 percent to 71.8 percent). A less noticeable change was apparent in computer files (8.2 percent,

from 12.3 percent to 20.5 percent) and language materials (6.0 percent, from 17.5 percent to 23.5 percent).

While these increases may be expected, the data also revealed an unexpected *decrease* in LCGFT usage for musical sound recordings (11 percent, from 46.2 percent to 35.2 percent). One possibility here is that pre-2007 materials had received LCGFT terms via retrospective application. Of course, not all materials are cataloged contemporaneously—manifestations predating 2007 could easily have been cataloged well past the initial implementation of LCGFT, although the year 2007 may be a rather arbitrary point of demarcation for this format in particular because LCGFT for musical works were implemented in 2015. Additionally, it might be possible that increased use of batch loading from external providers in the set of records from 2007 to the present may have increased the number of records lacking LCGFT—for example, newer records for streaming sound recordings. Regardless, this surprising result obviously seems to warrant a separate future inquiry. Lastly, it should be noted that kits, manuscript notated music, manuscript cartographic materials, and mixed materials did not present significant changes between the two periods (not presented in figure 2).

Library of Congress Classification

The authors also decided to take a close look at LC classification (LCC) in the data file to see if it might render different insights into patterns of LCGFT application in WorldCat records. Of the records containing one or more 655 fields with $2 lcgft (205,879 records), 158,125 records (76.8 percent) contained one or more LC call numbers. For the set of LCGFT records containing LCC, call numbers were extracted from fields 050 and 090 to perform classification analysis. The records were checked for internal duplication of classes and subclasses. For example, a record containing two instances of ML was only counted once toward ML. Also, for the purpose of the current study, records containing differing LC subclasses or classes were counted in each area; for example, if a record contained subclasses DS and PN, the record counted toward both subclasses and both overall D and P classes. Additionally, invalid call numbers were removed from the data set. For example, the authors found that many 050/090 fields contained Dewey or SuDocs numbers, or textual phrases such as "ISSN RECORD." These types of records were removed, and remaining LC classes could then be trimmed to their first letter alone for the analysis.

Extracting, cleaning, and deduplicating the call numbers from the set of 158,125 records containing LCGFT resulted in 163,067 valid instances of LCC classes. Figure 3 contains the entire distribution of LCC within records containing one or more LCGFT terms. P (language and literature, 28.2 percent) and M (music and books on music, 21.8 percent) represented half of the LCC classes in the authors'

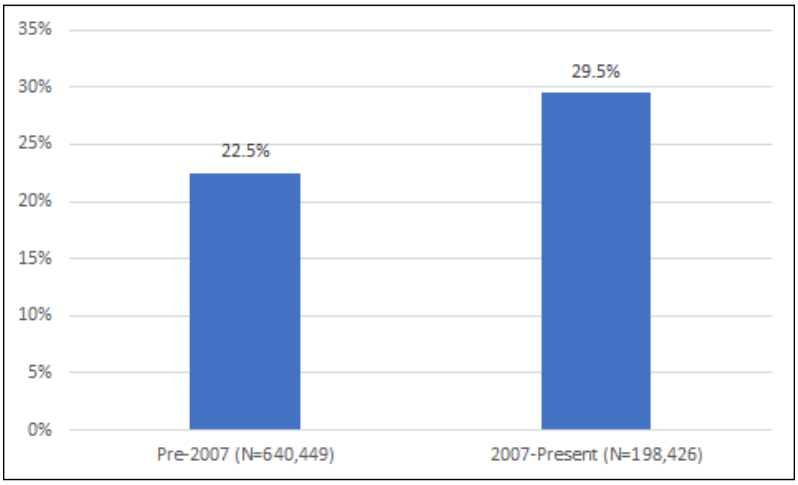


Figure 1. Percentage of Records with One or More LCGFT Term, by Date

Table 1. Percentage of Records with One or More LCGFT Terms by Type of Record

Type	Description	No. of Records with LCGFT	Total No, of Records in File	Percent
t	Manuscript language material	1,433	1,489	96.2
g	Projected medium	14,705	16,661	88.3
e	Cartographic material	4,595	6,979	65.8
c	Notated music	4,327	8,093	53.5
k	Two-dimensional nonprojectable graphic	205	404	50.7
j	Musical sound recording	51,402	125,898	40.8
i	Nonmusical sound recording	347	1,781	19.5
a	Language material	128,713	684,396	18.8
m	Computer file	133	1,013	13.1

base file; K (law, 9.9 percent) and H (social sciences, 7.4 percent) also revealed a moderate amount of representation in the file, followed by Q (science, 4.4 percent), D (world history and history of Europe, Asia, Africa, Australia, New Zealand, etc., 3.6 percent), and E (history: America, 3.2 percent). The remaining fourteen LC classes (each amounting to less than 3 percent) added up to just over 20 percent of the records containing LCGFT headings.

While this provides a broad picture of LCC distribution within the set of records containing LCGFT, a proportional analysis of this data against the entire base file provides a more accurate indication of the rate at which LCGFT has been applied within each class. For this analysis, records containing LCGFT were measured in each LC class against 622,777 records with 684,540 occurrences of valid LC classes from the base file. As seen in figure 4, M (music and books on music, 44.5 percent), P (language and literature, 41.1 percent) and K (law, 33.1 percent) still have high representation of records containing LCGFT; however, Z (bibliography/library science, 41.9 percent) has moved to second place, showing high levels of gene/form application for these resources. Although H (social sciences, 12.7 percent) ranked fourth in the earlier pure distribution, it dropped to the bottom half in the proportional analysis. C (auxiliary sciences of history, 26.4 percent), E (history: America, 26.3 percent), N (fine arts, 24.1 percent), and G (geography/anthropology/recreation, 22.4 percent) also showed moderate levels of LCGFT application.

Examining the number of terms applied by class per record also revealed interesting LCGFT application patterns, as shown in figure 5. For the 163,067 valid instances of LCC, there were a total of 220,668 fields 655 with $2 lcgft, yielding an average of 1.35 terms per record. There was some variability observed within this set; classes P (1.51) and M (1.48) show slightly higher levels of application (about 10 percent higher than the average), while K (1.02)—the LC class with the lowest level of LCGFT application—averaged only marginally higher than a single term assigned per record (about 25 percent lower than the average).

Another relevant area of analysis with regard to LCC was the distribution of LCGFT by Date 1. As described in the previous section on *Date*, the base file was divided into pre-2007 records and records from 2007 to the present. These two files of records were analyzed for LCC and Date 1; any record not containing a valid LCC class or Date 1 was omitted from this area of analysis. This resulted in 683,187 records total (80.7 percent of the base file). Out of this subset, 566,562 (82.9 percent) were in the pre-2007 group and 116,625 (17.1 percent) were in the 2007–present group. These two files were then examined for LCGFT; in the pre-2007 file, 125,630 records (22.2 percent) contained one or more LCGFT terms, while 36,754 records (31.5

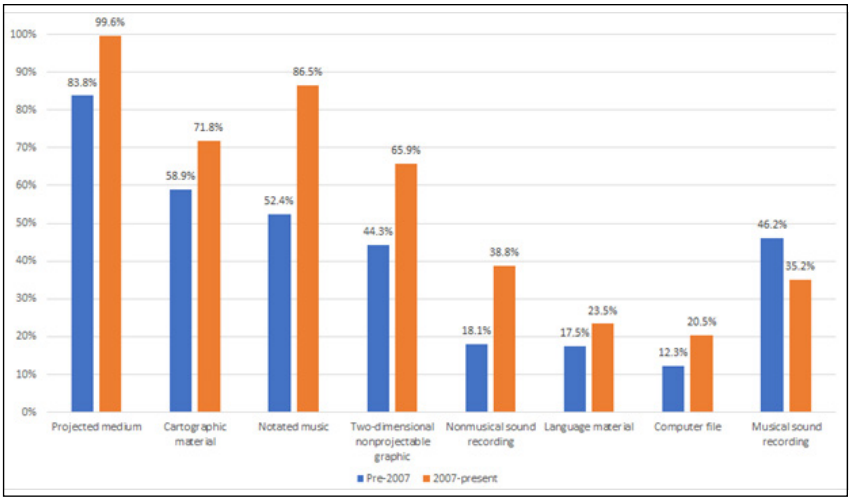


Figure 2. Percentage of Records with One or More LCGFT Terms by Type of Record, Grouped by Date

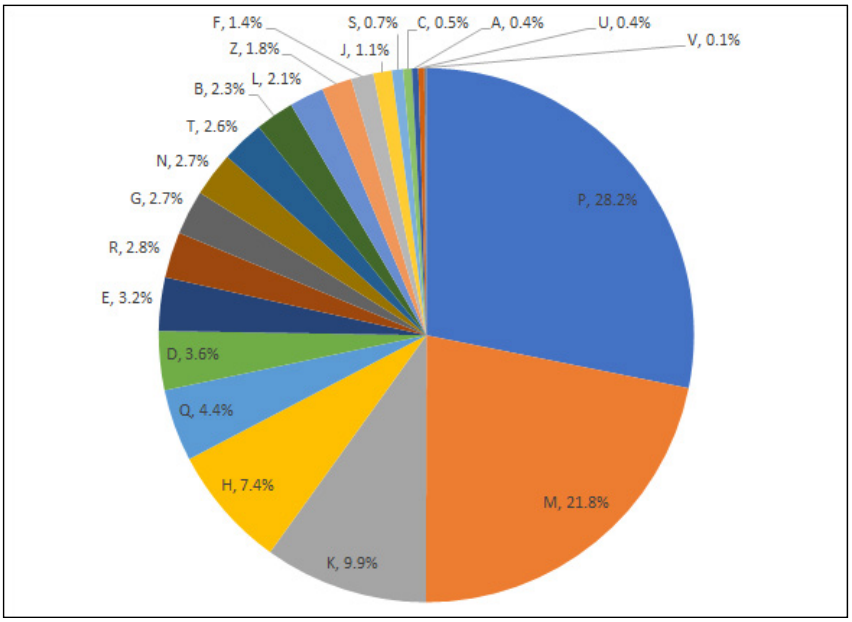


Figure 3. Distribution of LCC in Records Containing LCGFT, by Class (N = 163,067)

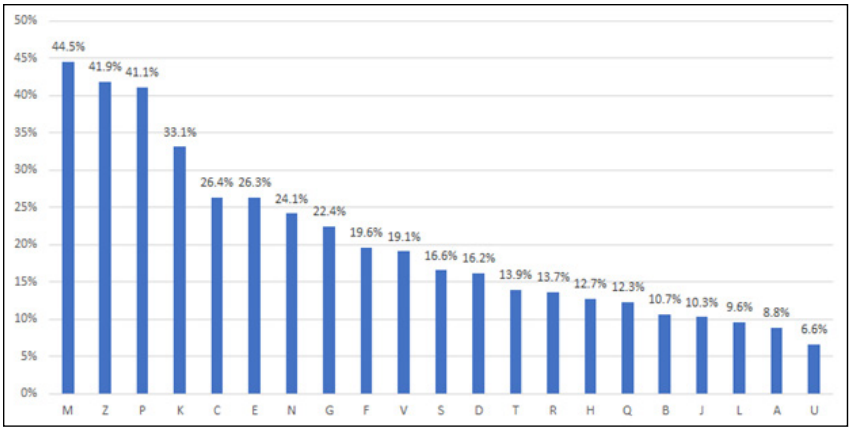


Figure 4. Proportion of Records Containing LCGFT, by Class (N = 684,540)

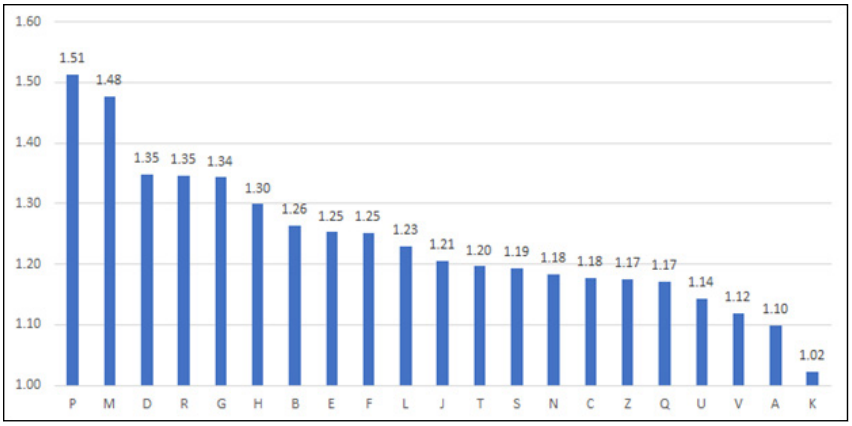


Figure 5. Number of LCGFT Terms Per Record by LCC Class, (N = 220,668)

percent) had LCGFT in the 2007–present file. Figure 6 shows the proportion of records containing LCGFT by class within each group of records, divided into pre-2007 and 2007–present records.

To further illuminate the data provided in figure 6, the authors also examined changes in the rate of LCGFT application in each LC class over time (figure 7). Two LC classes showed significantly increased rates of LCGFT application in more recent, 2007–present records, i.e., A (general works, 41.2 percent) and K (law, 38.2 percent). N (fine arts, 18.1 percent) also showed a moderate increase. As shown in the above analysis of LCGFT application by format, M (music, -13.2 percent) showed a moderate decrease in LCGFT application in the 2007–present group. LCGFT headings in Z (bibliography/library science, -10.6 percent) also decreased in the records representing more recent resources, despite its high representation of LCGFT against the entire base file (41.9 percent, see figure 4).

LCGFT Terms Assigned

Analyzing the individual LCGFT terms contained in the file was another relevant area of inquiry for the present study, as it revealed how the thesaurus had been used in WorldCat records. From the 205,879 records with one or more 655 fields containing subfield $2 lcgft (24.3 percent of the base file), terms in field 655 subfield $a were extracted to perform analysis. There were 284,964 655 fields with subfield $2 lcgft across this subset of records, with an average of 1.38 terms per record with LCGFT. After crossing each individual field 655 subfield $a against the master file of LCGFT (2,357 total terms), the authors found that 10,346 fields did not contain a valid LCGFT term. That is, 274,618 fields contained authentic genre and form terms, resulting in a 3.6 percent error rate in the file. For the 655 fields containing authentic LCGFT, 1,362 unique terms were present in the file, meaning that 57.8 percent of all LCGFT terms had been used in WorldCat records matched to the authors' local library holdings.

Analysis of the invalid LCGFT terms revealed a number of different types of errors. Many were simple typograpical errors (e.g., Stuides (Music), llustrated works), while others were missing qualifiers (e.g., Vespers, Rhapsodies, Thrillers). However, the majority of invalid LCGFT headings found were incorrectly assigned terms. Top offenders included "Electronic government information" (2,656 occurrences), "History" (806 occurrences), "Electronic Journals" (588 occurrences), "Juvenile works" (378 occurrences), and "Picture books for children" (217 occurrences). Table 2 contains every invalid LCGFT term that had more than 100 occurrences in the file.

Despite the fact that LCGFT terms can be easily controlled within the Connexion Client, the authors' data thus make it abundantly clear that invalid terms are still being deposited in field 655. One could easily infer any number of sources through which these invalid LCGFT terms had been introduced into WorldCat records. Some terms could have been simply misapplied by catalogers or there may be a deeper misunderstanding of the vocabulary. Conversely, terms may have been inadvertently added through improper authority control. For example, some authority systems might have flipped LCSH to LCGFT even though an equivalent term does not exist, i.e., 650 _0 $a Piano music.

changed to 655 _7 $a Piano music. $2 lcgft.[18] These headings could have easily ended up in WorldCat master records, particularly in light of ongoing data sync projects. As 3.6 percent is a relatively small portion of the file, one could argue that the problem is not so severe. However, given the ease of correcting many of these headings (for example, Sonatas (Piano) could easily be flipped to Sonatas), it seems regrettable that so many improper terms coded as LCGFT headings exist in WorldCat. Of course, many of these terms likely exist downstream in local library catalogs relying on WorldCat copy records, so the problem collectively has enormous cascading effects on the integrity of bibliographic databases across the wider library community.

Individual terms were further analyzed with a focus on the depth of indexing, that is, the extent to which broader and narrower terms have been assigned in terms of the hierarchies in LCGFT. As seen in figure 8, the overwhelming majority of LCGFT terms found in the authors' base file were coded in the second and third levels of hierarchy (45.7 percent and 42.5 percent, respectively). As expected, more specific, lower-level terms (level 4 in figure 8) were rarely used overall either because they are suited to describing few specialized resources, or because catalogers have applied broader terms for said resources instead. This result seems to reflect the basic guideline in the introduction to LCGFT Manual: "The preference is for broader, rather than narrower, terms. Most literary and artistic resources provide only a broad indication of their genres and forms. Broader terms can therefore expedite cataloging and also serve the users, who do not have to search several very narrow sub-genres or forms to find materials of interest to them."[19] (Note, however, that this guideline does have some conflict with the other guideline found in instruction sheet J 110: "Assign terms that are as specific as the genres and forms exemplified in a resource."[20]) Additionally, broadest, top-level terms (level 1) not surprisingly saw less use (7.9 percent) because these terms are intended more for collocation in each discipline; indeed, in many disciplines, top terms were rarely applied, if at all.[21] Based on the authors' analysis of LCGFT terms used in the base file records, some exceptions included "Sound recordings," "Literature," "Illustrated works," "Music," and "Video recordings," as illustrated by figure 9.

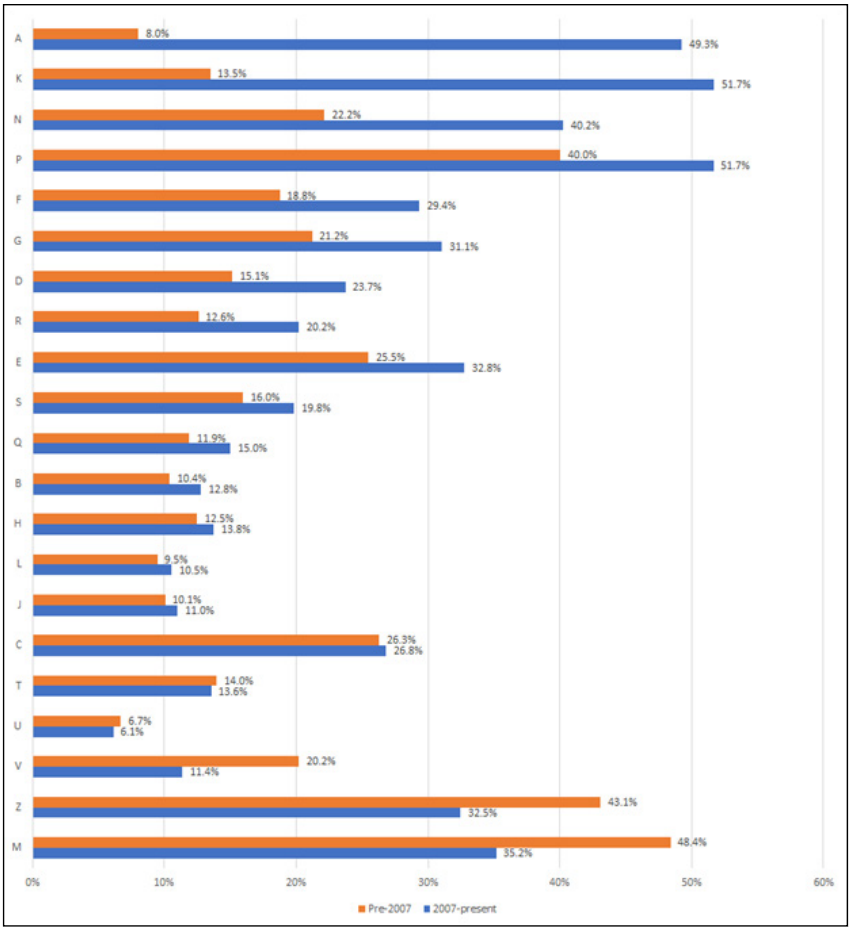


Figure 6. Proportion of Records Containing LCGFT, by Class and Year (N = 683,187)

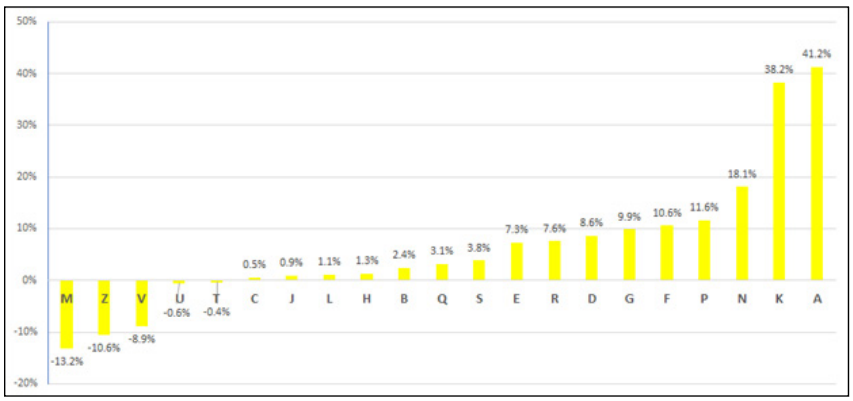


Figure 7. Percent Change in Number of Records Containing LCGFT, pre-2007 to 2007–present (N = 683,187)

In addition to the hierarchical distribution of LCGFT across the base file, records were also analyzed similarly

Table 2. Most Prevalent Invalid LCGFT Terms

Invalid 655 $a with $2 lcgft	No. of Occurrences
Electronic government information	2,656
History	806
Electronic journals	588
Juvenile works	378
Picture books for children	217
Detective and mystery stories	187
Sonatas (Piano)	174
High interest-low vocabulary books	170
Piano music	161
Electronic books	157
Criticism, interpretation, etc	154
Streaming audio	154
Photography, Artistic	147
Children's poetry	122
Compact discs	120
Young adult fiction	109
Children's stories	101

within each LCGFT category. With some exceptions, the most popular level of LCGFT application was level 2, or second-level terms. Thirteen of twenty-one LCGFT categories—commemorative materials, creative nonfiction, derivative works, ephemera, illustrated works, instructional and educational works, literature, motion pictures, religious works, tactile works, television programs, video recordings, and visual works—all followed this pattern. Some other vocabulary categories, by contrast, were applied more at level 3, or third-level LCGFT. These included cartographic materials, discursive works, informational works, law materials, music, and recreational works. The distribution of LCGFT terms in these categories favoring third-level LCGFT can be seen in figure 10. The two remaining categories, art and sound recordings, showed the greatest proportion of terms at level 1, first-level LCGFT. 72.3 percent of art terms were first level; this is not surprising given the relative sparse nature of art compared with other LCGFT categories—indeed, even at the second level only twelve terms are available as of this writing. As for sound recordings, 64.0 percent of terms were first level, which is understandable given the relatively broad applicability of the term. It should be noted, however, that the Music Library Association (MLA) states that "The term 'Sound recordings' is effectively a heading of last resort, i.e., it is a broad term that may be used to capture the sound recording aspect of a resource in cases where a narrower term is not available."[22] Despite these recommendations, the term "Sound recordings" had been available

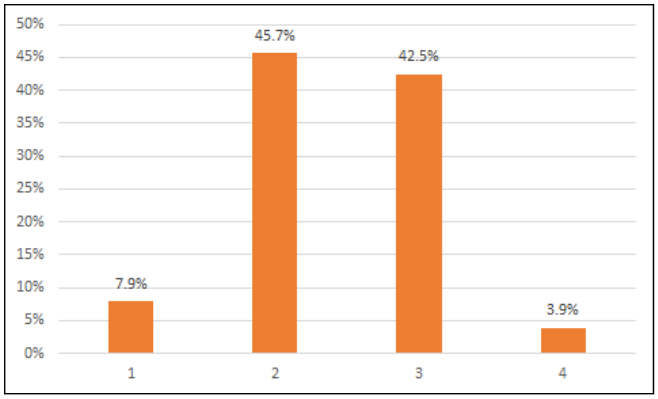


Figure 8. Hierarchical Level of LCGFT Terms as a Proportion of all LCGFT Terms (N = 274,618)

long before many of its narrower terms, which might explain why the top term was applied at greater levels. For example, "Sound recordings" was available as early as 2011, while "Studio recordings" first entered the vocabulary in 2019.

Lastly, two categories showed relatively high usage of fourth-level LCGFT, as evidenced in figure 10. Both cartographic materials (9.7 percent) and music (18.2 percent) exhibited somewhat heightened use of level 4; indeed, of the 21 LCGFT categories, only five showed application of fourth-level LCGFT at rates higher than five percent (cartographic materials, literature, motion pictures, music, and recreational works), with more than half of the twenty-one categories yielding less than one percent. Regarding cartographic materials, the position of both "Bathymetric maps" and "World atlases" within level 4 accounted for the majority of terms contributing to the rate of 9.7 percent in the authors' data. Examining the corresponding records reveals that the vast majority of these materials were for online government documents. As for music, relatively high fourth-level application is somewhat not surprising given the size and nature of the discipline; indeed, at 847 terms, music accounted for over a third of the entire LCGFT vocabulary. What is more, over half of music LCGFT terms (436 terms) in the vocabulary occurred within fourth level, indicating that all those specific terms were clearly regarded as necessary in describing musical resources when LC originally partnered with MLA to develop genre and form terms for music. Thus, it is far more likely for level 4 terms to be applied in this discipline than religious materials, for example, in which only 21.6 percent of terms occur at the lowest level of hierarchy.

Conclusion

The purpose of this study was to provide exploratory analysis of LCGFT within a large set of MARC bibliographic data. The authors retrieved their institutional holdings

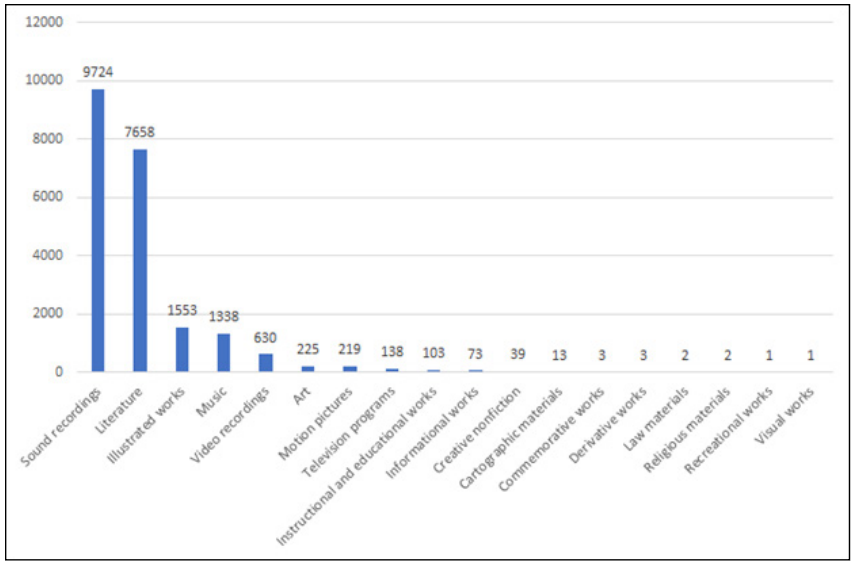


Figure 9. Number of Top Level Terms Used

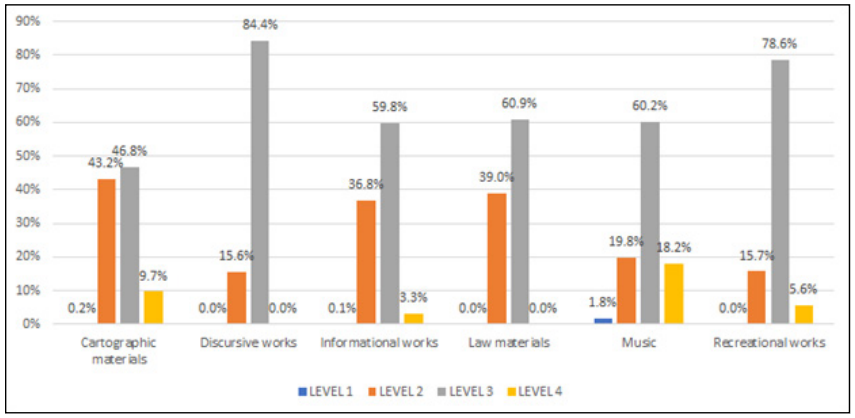


Figure 10. LCGFT Categories Favoring Level 3 Application

in WorldCat, using more than 800,000 master WorldCat records as the basis for analysis. From this base file, various data, such as date, format, call numbers, and LCGFT, were extracted and analyzed to explore a series of research questions related to the current status of LCGFT usage in WorldCat. With regard to changes in LCGFT application over time, there was an increase of seven percent between pre-2007 records and recent records from 2007–present (22.5 percent to 29.5 percent). Additionally, the average number of LCGFT terms increased in records containing them, from 1.34 to 1.50. When analyzing the data by format (e.g., type of record), most formats saw an increase in LCGFT application over time, with the exception of musical sound recordings. These findings are also supported by further analysis based on LCC; indeed, while many classes showed an increase in application between the pre-2007 and 2007–present sets, unexpected decreases were found in both M (music, -13.2 percent) and Z (bibliography/library science, -10.6 percent). The reason for such decreases could be that pre-2007 music materials may have received higher levels of retrospective application of LCGFT, or they could have been originally cataloged after music terms were added to LCGFT in 2015. Alternatively, the decrease might be explained by increased levels of batch loading of newer records for streaming sound recordings by external providers. A separate inquiry into these results and prospects for retrospective application would be warranted in view of the varied LCGFT application between formats and LC classes. Furthermore, this exploratory study used the year 2007—when LC first released the LCGFT thesaurus for moving image materials—as the point of demarcation to shed some light on changes in LCGFT application over time. Because LCGFT terms have been added in different disciplines over multiple years, it will be worthwhile to pursue further research on how LCGFT usage changed respectively when the LCGFT project was completed for a given discipline.

When examining the entire LCGFT vocabulary in terms of hierarchy, the authors found that second and third-level headings were assigned most frequently (45.7 and 42.5 percent, respectively). This was also evident for the overwhelming majority of individual LCGFT disciplines examined, such as motion pictures (favoring second-level) and music (favoring third-level). Perhaps the preponderance of second and third-level LCGFT headings used suggests that the hierarchical design of the vocabulary is working; it is reasonable to assume that they are specific enough, compared with the broadest, top-level terms, to describe the genres and forms exemplified in resources being cataloged, but not too narrow to impede efficient cataloging or confound the users as they try to find materials of interest to them. Further, the most specific, fourth-level LCGFT (which included fourth-level terms and below in the current paper) saw the least usage as these terms would naturally only be used for more specialized or unique resources; for example, cartographic materials and music, which had higher fourth-level usage than other disciplines. These results suggest that future efforts to add new terms to the vocabulary should aim to strike a balance

between specific and broad terms. Additionally, it should be noted that the application of first-level terms within certain LCGFT disciplines may warrant further analysis. For example, headings such as "art" and "sound recordings," particularly if the only genre form term recorded in the record, may not necessarily provide users with altogether helpful information, and further analysis might yield new insights that will be essential for any individual constituencies that wish to develop best LCGFT practices guidelines in these disciplines. Lastly, the number of erroneous terms in fields 655 subfield $a with $2 lcgft (10,346 total fields in the base file) points to some much needed data cleanup in WorldCat, as well as potential training and documentation for applying LCGFT terms correctly.

While the data reported in this study point to a moderate increase in LCGFT use over time, the amount of LCGFT within the base file suggests that the vocabulary has not been applied to the fullest extent possible in WorldCat. The results of the present study indicate that it is highly important that newly cataloged materials receive LCGFT application within records from the outset, so as to ensure that a more sizable portion of new bibliographic records include appropriate genre and form terms and lessen the need for retrospective application over time. As such, training needs to be increased in both libraries and library schools to facilitate broader LCGFT application. Increased communication with vendors may also be warranted, as LCGFT may be lacking (or incorrect) in vendor-supplied metadata. While these actions may further improve end-user retrieval based on genre and form, catalogers and other technical services librarians may need to begin to investigate more sophisticated methods in applying the vocabulary retrospectively to appropriate legacy records as well. Indeed, as the data contained in the base file show somewhat uneven application of LCGFT, and with nearly half a billion records in WorldCat as of the 2020 OCLC report, it remains a certainty that much of LCGFT's full potentials for genre/form access and retrieval will remain untapped until innovative solutions are introduced to increase vocabulary usage in bibliographic databases.[23]

References

1. Library of Congress, "Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials," 2019, https://www.loc.gov/aba/publications/FreeLCGFT/2019%20LCGFT%20intro.pdf.
2. Library of Congress, "Report on the Moving Image Genre/Form Project," 2008, https://www.loc.gov/catdir/cpso/movimgenre.pdf.
3. Charles A. Cutter, *Rules for a Printed Dictionary Catalog* (Washington, DC: US Government Printing Office, 1876); Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge: MIT Press, 2000).
4. Pat Riva, Patrick Le Bœuf, and Maja Žumer, *IFLA Library Reference Model* (International Federation of Library Associations and Institutions, 2017).
5. American Library Association. Subcommittee on the Revision of the Guidelines on Subject Access to Individual Works of Fiction, *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.* (Chicago: American Library Association, 2000); David P. Miller, "Out From Under: Form/Genre Access in LCSH." *Cataloging & Classification Quarterly* 29, no. 1-2 (2000): 169–88; Edward T. O'Neill et al, "Form Subdivisions: Their Identification and Use in LCSH." *Library Resources & Technical Services* 45, no. 4 (2001): 187–97.
6. Library of Congress, "Report on the Moving Image Genre/Form Project"; Janis L. Young and Yael Mandelstam, "It Takes a Village: Developing Library of Congress Genre/Form Terms," *Cataloging & Classification Quarterly* 51, no. 1–3 (2013): 6–24.
7. Library of Congress, "Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials."
8. Hur-Li Lee and Lei Zhang, "Tracing the Conceptions and Treatment of Genre in Anglo-American Cataloging," *Cataloging & Classification Quarterly* 51, no. 8 (2013): 891–912.
9. Library of Congress, "Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials."
10. Martha Yee, "Two Genre and Form Lists for Moving Image and Broadcast Materials: A Comparison," *Cataloging & Classification Quarterly* 31, no. 3–4 (2001): 237–95; Faye Leibowitz, "Form and Genre Headings in Serials Cataloging," *Cataloging & Classification Quarterly* 20, no. 3 (1995): 19–41; Spillane Wilson, "The Relationship Between Subject Headings for Works of Fiction and Circulation in an Academic Library," *Library Collections, Acquisitions, & Technical Services* 24, no. 4 (2000): 459–65; Carrie Newsom, Jimmie Lundgren, and Nancy Mitchell Poehlmann, "Genre Terms for Chemistry and Engineering: Not Just for Literature Anymore," *Cataloging & Classification Quarterly* 46, no. 4 (2008): 412–24.
11. Young and Mandelstam, "It Takes a Village."
12. Beth Iseminger et al., "Faceted Vocabularies for Music: A New Era in Resource Discovery," *Notes*, 73, no. 3 (2017): 409–31; Mark McKnight, "Are We There Yet? Toward a Workable Controlled Vocabulary for Music," *Fontes Artis Musicae* 59, no. 3 (2012): 286–92.
13. Casey A. Mullin, "An Amicable Divorce: Programmatic Derivation of Faceted Data from Library of Congress Subject Headings for Music," *Cataloging & Classification*

Quarterly 56, no. 7 (2018): 607–27.
14. Patricia M. Dragon, "Form and Genre Access to Academic Library Digital Collections," *Journal of Library Metadata*, 20, no. 1 (2020): 29–49.
15. Colin Bitter and Yuji Tosaka, "Genre/Form Access in Library Catalogs: A Survey on the Current State of LCGFT Usage," *Library Resources & Technical Services* 64, no. 2 (2020): 44–61.
16. Library of Congress, "Assigning Genre/Form Terms," 2016, https://www.loc.gov/aba/publications/FreeLCGFT/J110.pdf.
17. OCLC, "FAST Frequently Asked Questions," 2019, https://www.oclc.org/content/dam/oclc/fast/FAST-FAQ-Nov2019.pdf.
18. Bitter and Tosaka, "Genre/Form Access in Library Catalogs."
19. Library of Congress, "Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials."
20. Library of Congress, "Assigning Genre/Form Terms."
21. Library of Congress, "Frequently Asked Questions about Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT)," 2011, https://www.loc.gov/catdir/cpso/genre_form_faq.pdf.
22. Music Library Association. Cataloging and Metadata Committee Subcommittee, "Best Practices for Using LCGFT for Music Resources," 2019, http://cmc.blog.musiclibraryassoc.org/wp-content/uploads/sites/5/2019/07/BestPracticesforUsingLCGFT_Music_1.2_20190708_revURLs.pdf.
23. OCLC, "OCLC Annual Report 2019–2020," 2020, https://www.oclc.org/en/annual-report/2020/home.html.

Notes on Operations

Digital Collections at a Distance

Telework during the COVID-19 Pandemic

Laura M. Gentry

This case study explores how one team tasked with the creation of digital collections at The University of Alabama Libraries succeeded at telework to carry on its essential functions despite not being able to digitize new content from March through July 2020 during the COVID-19 pandemic. Managers of similar units will gain strategies to create similar telework projects at their institution and lessons learned while working and supervising employees remotely.

The COVID-19 pandemic forced The University of Alabama to limited business operations with its libraries shutting down its physical spaces and all library employees shifting to telework. Within The University of Alabama Libraries Special Collections, Digital Services functions as a specialized unit tasked with the digitization of unique materials, creation of digital collections, and digital preservation. The challenge was how to continue these functions without having physical access to materials and the ability to digitize new content. This case study explores how Digital Services responded to the challenge by leveraging technology and embracing the constraints of working remotely. By sharing these experiences, the goal is to help others who face similar challenges now and in the future, and to offer strategies that can be adapted to use at their institutions. Digital Services implemented four strategies while working remotely, including the repurposing of existing digital content, strengthening infrastructure, enhancing discovery, and trying a new approach. The historical significance of the COVID-19 pandemic pushed Special Collections to try something new with the creation of "Still Tide Together: Documenting Life during the COVID-19 Pandemic," whose submissions will represent the first intake of born-digital content without having an established born-digital program in place. Although the impetus for telework was the COVID-19 pandemic, the strategies presented apply to other unexpected situations including natural disasters, medical emergencies, building maintenance issues, or other short-term disruptions. This advice is also relevant for business continuity and emergency management plans and other planned disruptions, such as building renovations, familial job relocations, or extended medical leave, that take employees away from physical collections and equipment. Besides these strategies, this case study offers managers who oversee the creation of digital collections practical advice on how to develop projects in alignment with strategic goals, remote supervision, and supporting employees during extraordinary circumstances.

Laura M. Gentry (lmgentry@ua.edu) is the Special Collections and Digital Initiatives Librarian at the University of Alabama Libraries.

Manuscript submitted September 19, 2020; returned to author for revision December 18, 2020; revised manuscript submitted January 10, 2021; accepted for publication March 5, 2021.

The author included key ideas of this article in the presentation, "Digital Collections at a Distance" during the Creative Ideas in Technical Services Interest Group meeting on June 11, 2020, as part of ALCTS Virtual Interest Group Week.

Literature Review

In 2006, Robertson hypothesized how a pandemic would affect the daily operations of a library, and many of his predictions, including library closures and limited physical access to the physical buildings for extended periods, were correct. Yet, Robertson could not imagine how technological improvements over the next fourteen years would transform the library into an institution that could carry

on the majority of its functions remotely with employees effectively working from home.[1] The literature on teleworking work for library employees is limited but most likely will expand over the next few years because of the COVID-19 pandemic. The literature primarily focuses on four areas: setting up a telework agreement, employee-supervisor relationship, technology and communication tools, and personal experiences primarily from reference and instruction librarians and many technical services positions.

The overwhelming majority of library literature on this topic incorporated personal experiences and opinions towards telework and the process of setting up a teleworking agreement. Manley gave up telework after thirty minutes and drove into the library, believing that connection to people was fundamental to what defines a library.[2] Moe described her teleworking experience as a law librarian who found that online access to legal databases and electronic delivery of research made her position suited to telework.[3] Diehl Dietr explained the process of transforming her position as a theological librarian into a telework position.[4] Brooks-Kieffer negotiated with her university library administration a formal agreement to telework three days a week while working as an electronic resources (e-resources) librarian.[5] Throughout these papers, the authors repeated the importance of planning, outlining expectations, and deciding if a particular library position is suitable for telework and the best choice for an individual.

One reason for engaging in telework was to assess the feasibility of telework as part of a pilot program. Black and Hyslop participated in a two-day-a-week pilot program at Michigan State University Libraries in which librarians created original catalog records while working from home.[6] Luce and Hartman studied the six employees representing various departments of the Boulder Public Library who teleworked one day a week for one year in 1985. They found that employees experienced personal growth, increased productivity, a sense of control in planning work, and increased job satisfaction, yet also experienced distractions, feelings of isolation, workspace difficulties, and a need to structure work time. Their findings remain relevant to library employees who engage in telework today as there are advantages and disadvantages to working outside of a structured office environment.[7] While extremely dated in terms of technology used for telework, these authors believed that telework has promise for use in libraries, but the technology had to improve to make it an everyday reality instead of a pilot program.

The technology exists to make telework doable for library employees, but libraries have been reluctant to embrace telework, citing a need for employees to be with library collections. Hickey and Tang argue that an underlying fear drives this decision due to a lack of trust between a supervisor and employee or within the organizational culture. They also discussed the teleworker and the supervisor relationship using Tang's teleworking experience as a reference and instruction business librarian.[8] Circumstances, such as the need to move to another geographical location, force libraries to consider telework as an option for an employee. An unexpected temporary move led Duncan to telecommute for six months while working as an e-resources librarian.[9] Smith and Van Dyke shared the telecommuting experiences of their university's interlibrary loan department from both the perspective of the telecommuter and an on-site supervisor when Van Dyke relocated to another city three hours away.[10] These authors emphasized the importance of planning for telework, outlining expectations for both the employee and the supervisor, and establishing methods of communication. The physical distance added another dynamic between the employee and supervisor relationship. Manley argued that supervisors have assumed that visual observation of employees at work guarantees job performance instead of measuring an employee's performance by the deliverables produced.[11] Brooks-Kieffer suggested that results-based or outcome-based evaluation of employees lends itself to telework.[12] Hickey and Tang pointed out that the collaborative nature of the library work with many employees working in teams has stifled the acceptance of telework, yet telework can be effective in a team environment if a supervisor ensures communication and collaborations between employees.[13]

The COVID-19 pandemic necessitated the transition of library employees into telework with little time to plan, sometimes overnight. There have been a few contributions to the literature that specifically address the COVID-19 pandemic, with many more to follow. Technology has made it possible to simulate much of the team environment virtually with collaborative technologies. Rysavy and Michalak discussed the communication methods as library employees of Goldey-Beacom College in Wilmington, Delaware, transitioned to telework. Some of the technologies included Flipgrid, Slack, Zoom, Notion, and SharePoint. Rysavy and Michalak focused mainly on communication in the first weeks following the physical shutdown of the campus and represented an early contribution to the literature.[14] Walsh and Rana wrote about The University of Toronto's response during the first three months concluding with May 2020. They provided a bird's eye view of library operations with a focus on reference, circulation, interlibrary loan, and e-resources. Subsequent contributions to the literature should include examples of telework projects within a specific library area and the possible incorporation of telework in long-term plans and goals of that unit.[15]

No one was adequately prepared for the COVID-19 pandemic's effect on libraries. This case study positions itself as an in-depth look at how to create digital collections at a distance when facing no physical access to the

materials that need digitization and provides strategies for similar units for telework. By giving concrete examples of what a unit tasked with the creation of digital collections could accomplish remotely from March through July 2020, the author argues that telework is a feasible option to carry out the overwhelming majority of the functions of this unit and that the technological barriers that thwarted telework years ago no longer exist. This case study offers advice to managers in supervising employees remotely. Everyone has their own transition to telework story, and managers should reflect on what worked and what they could have done differently going forward in a world where telework might be the new normal for months and maybe even years to come.

Background

At The University of Alabama Libraries Special Collections, there has been a significant amount of change over the past two years that shaped the daily work of Digital Services. The Special Collections and Digital Initiatives Librarian who oversees Digital Services started in August 2018, and she supervises two full-time staff and two student employees. There was a decision to switch the digital asset management system (DAM) from a custom-built in-house system called Acumen to the cloud-hosted version of CONTENTdm. One of the reasons for this decision was that Acumen's infrastructure lived on an aging physical server that could not be upgraded to meet Special Collections' needs. This server also contained the digital archival storage consisting of TIFFs and metadata plus the Lots of Copies Keeps Stuff Safe (LOCKSS) infrastructure. The University of Alabama participates in the Alabama Digital Preservation Network (ADPN), a statewide LOCKSS network, for its digital preservation solutions. Digital Services froze additions to the existing server and deposited its last batch of new digital content into ADPN in June 2019 as part of its preparation for the migration to a new server. Digital Services was given access to a new virtual server in December 2019 with only the archival storage on it while the transfer of other components were scheduled to take place in summer 2020. The new public-facing digital collections website powered by CONTENTdm launched in April 2019, and Digital Services has been migrating content from one DAM to the other with the planned shutdown of Acumen in July 2020. The focus was on completing migration and redesigning workflows for CONTENTdm and technical processes for the new server. Digital Services digitized a few high priority collections, uploaded them to CONTENTdm, and stored the images and metadata on local network storage while waiting for a new server and LOCKSS to be fully functional. Digital Services planned to resume regular digitization production in summer 2020.

With the COVID-19 pandemic threat looming, students, faculty, and staff left for Spring break on March 14, 2020, fully expecting to return to class and work after this holiday. By March 17, 2020, The University of Alabama announced limited business operations, the switch to online classes for the rest of the semester, and instructed students not to return to campus from Spring break. All library employees transitioned to telework with a telecommuting agreement in place on March 23, 2020. It also meant the loss of the two student employees who digitize the majority of new digital content and were in the middle of digitizing two large collections. When considering the feasibility of telework for library departments, Special Collections departments would be one of the least likely candidates because its staff requires access to its collection of unique and rare materials to perform both archival processing and digitization. The unforeseen transition to telework required managers to determine what telework looked like every day for their units by assessing technological needs, defining communication methods, and planning the scope of work while working remotely.

Technology

Without the ability to bring home scanners and materials from Special Collections, Digital Services had to determine what telework meant and what technology was needed to be successful when working from home. After compiling a list of all the technological needs ranging from software to server access, the best option for Digital Services was to use a virtual private network (VPN) and remotely log into their work computers using the remote desktop function from personal computers or a library loaned laptop. The reasons for this choice included the processing power of work computers designed to handle large image files, specialized software (e.g., Adobe Photoshop or CONTENTdm Project Client), and scripted workflows that relied on Perl and Python scripting languages for execution. The benefits of this approach meant less downtime in installing and downloading software on personal computers and configuring access to network drives and servers. The added benefit of using a remote desktop was the feeling of being at work with all the settings and bookmarks customized for that environment. During the first two weeks of remote work, the biggest problem was the number of connections available through the VPN and the stability of the connection. Luckily, the university's Office of Information Technology created a second VPN network on April 1, 2020, and there have been very few VPN connection problems since then.

Communication

The shift to telework meant determining methods of communication. Managers should determine what forms

of communication work best by seeking input from their team and what works best for them. During March 2020, the Special Collections and Digital Initiatives Librarian tried a continuous Word document in Box, a university provided cloud-based storage subscription, as a centralized place for work assignments and updates. The amount of time needed to provide written updates was onerous, led to more questions, and felt impersonal. Digital Services switched to a scheduled daily check-in meeting each morning using the university-provided Zoom subscription to give progress updates and most importantly, support one another emotionally through the grief, anger, sadness, and the uncertainty of living through a pandemic. It also provided a venue to share successes, happy moments, and to laugh mirroring the informal check-ins that had occurred while working on-site. These virtual meetings typically lasted thirty minutes and were longer when troubleshooting problems or conducting training. The other primary method of communication was email both for quick questions and to communicate information coming from other library departments and outside vendors (i.e., EBSCO or OCLC). By establishing and using these communication methods, employees felt empowered to adapt work hours to fit their needs. For example, one employee monitored the online schooling of children during the day and worked late at night to minimize distractions and increase productivity. By embracing the freedom and flexibility of telework, employees could blend personal and work time throughout the normal work hours to achieve a better work-life balance when working from home.

Scope of Work

After determining the technological needs and methods of communication while working from home, Digital Services focused on the unit's goal of completing the migration of all existing digital content to CONTENTdm and ensuring that the impending shutdown of Acumen on the existing server did not cripple workflows and digital preservation. With the number of COVID-19 cases daily rising, the state issued a "Stay at Home" order beginning in April 2020. In May 2020, the state relaxed its guidelines and issued a "Safer at Home" order, which lessened restrictions on non-essential businesses, but continued to encourage telework whenever possible. The University of Alabama responded with the announcement that remote learning for students would continue during the interim and summer 1 terms, and only essential employees would be permitted on-campus. It became apparent that telework would be the norm for the indefinite future. With this knowledge, the Special Collection and Digital Initiatives Librarian expanded the scope of work to include more than migration, ensuring that any assigned work coincided with existing objectives and goals. She adopted these strategies, including repurposing existing digital content, strengthening infrastructure, enhancing discovery, and trying new approaches to the unit's work.

Implementation of the Strategies at Other Institutions

The strategies proposed in this case study are intentionally generic to account for the differences such as staffing levels, organizational structure, and unit responsibilities that exist between similar units at other institutions. Other institutions can use these strategies and customize the implementation to fit their situation. Each strategy will have a brief definition with generic implementation examples followed by specific examples detailing application of this strategy at The University of Alabama from March to July 2020. By providing real-life examples, this paper intends to provide guidance to managers of similar units at other institutions planning appropriate projects for telework. With no new digitization possible due to the closure of the campus, it is important to demonstrate that telework is meaningful and continuing the everyday work of Digital Services.

Strategy of Repurposing Existing Digital Content

The first strategy was to repurpose existing digital content created originally for another purpose such as for imaging requests, exhibits, digital humanities projects, or grant-funded digitization. In a telework environment, which lacks a means to digitize new content, institutions must assess the viability of any existing digital content as a candidate for inclusion in its digital collections and digital exhibits. The strategy of repurposing digital content creates quantitatively measurable additions to digital collections. Managers of digitization units often report to library administration the number of images digitized and added to its digital collections as one statistical measure to demonstrate productivity. Repurposing digital content to create new digital exhibits can add context and interpretation and can serve as an outreach tool to create awareness of an institution's holdings. By using web analytics, both digital collections and digital exhibits can provide usage statistics.

Implementation of the Strategy of Repurposing Existing Digital Content

At The University of Alabama, the three options to repurpose existing digital content were non-migrated digital content from its legacy DAM, a grant-funded digitization project, and digital content created originally for physical

exhibits. By March 2020, Digital Services had migrated 288 digital collections to CONTENTdm, which represented 94 percent of the total digitized content of 524,941 images. Digital Services concentrated their efforts on the most popular or the largest digital collections. The remaining 6 percent of digitized content was from 259 small digital collections and was the most time-consuming and difficult digital collections to migrate to CONTENTdm. The migration to the new DAM required converting MODS metadata to Dublin Core (DC) metadata, which Digital Services had partially automated using a Python and Perl rules-based scripted workflow. The metadata transformation required the same number of steps regardless of collection size, which meant that smaller digital collections consisting of one item took as long as 1,000 item digital collections. By July 2020, Digital Services completed the migration of the remaining 259 digital collections consisting of 31,720 images and audio files, successfully completing the migration of 547 digital collections totaling 524,941 images and audio files.

Grant-funded digitization offers another means to find existing digital content. The University of Alabama Libraries Special Collections received a National Endowment for the Humanities grant in 2018 to digitize 100,000 pages of state newspapers for inclusion in the Library of Congress's Chronicling America digital collection as part of the National Digital Newspaper Program (NDNP). The inclusion of its newspapers would generate significant amounts of web traffic, be beneficial to researchers using the digital collections, and represented the largest amount of existing digital content with 50,000 pages available when telework started in March 2020. That number of pages continued to grow over the next few months as the vendor returned newly digitized content to meet the grant-mandated completion deadline of August 2020. Digital Services transformed NDNP-compliant metadata to align with its DC metadata template and uploaded the PDF version of each newspaper issue since OCR was already embedded within the file. By July 2020, Digital Services added 62,036 newspaper pages to its digital collections.

Another option might be to repurpose existing digital content created specifically for a physical or digital exhibit whose content does not already exist in the digital collections. In 2019, Digital Services successfully used this strategy with *The Alabama Forum* LGBTQ newspaper digital exhibit to modify existing metadata to make it compliant to DC standards, add optical character recognition (OCR) to previously digitized images, and upload to its digital collections. This strategy can also work in the reverse with the creation of digital exhibits using existing items from an institution's digital collections or digital content created specifically for physical exhibits. Digital exhibits provide a way to offer more context to digital content, showcase small yet important digital collections, can coincide with anniversaries and other historical events, and highlight holdings of marginalized groups to foster diversity, equity, and inclusion in campus culture at an institution. In summer 2020, the Reference Services and Outreach Coordinator used existing digital content to create two timely digital exhibits, "Woman Suffrage in Dixie" and "Unrest: Two Weeks of Protest at The University of Alabama, 1970."

Strategy of Strengthening Infrastructure

This strategy involves the development or modification of technical processes and workflows to accomplish a specific task or improve efficiency. Sometimes, in the pursuit of creating quantifiable statistics through images digitized and uploaded to digital collections, managers do not allocate enough resources towards strengthening infrastructure. The abrupt shift to telework due to the COVID-19 pandemic allows managers to re-visit resource allocation towards the development of technical processes and improvement of workflows. Strengthening infrastructure is necessary for the care and maintenance of digital content and is worth the time investment to ensure long-term digital preservation. Digital collections represent the public-facing view of what digitization units do, but the behind-the-scenes work of digitization, creating metadata, uploading digital content, and digital preservation represents the important work that happens before and after making a digital collection available online. Because of the differences inherent in similar units at other institutions, there is not a list of examples for ready implementation, but most managers know where deficiencies and inefficiencies lie.

Implementation of the Strategy of Strengthening Infrastructure

Digital Services implemented this strategy of strengthening its infrastructure by tweaking existing scripts for audio collections, developing a bulk editing metadata script, and re-configuring the first part of its digital preservation workflow. Throughout 2019 and early 2020 as part of the migration, Digital Services had created new workflows with most of the processes automated using Perl and Python scripts to migrate existing digital content and add new digital content to its digital collections. The migration of the audio collections represented another challenge as they epitomized the lack of standardization and variation within metadata from one audio collection to the next. Migrating these problematic audio collections meant the modification of existing scripts and the development of new scripts to pull large .wav audio files from the server and convert them to smaller .mp3 files for upload into digital collections and incorporating both the item- and track-level metadata. MODS metadata

allowed for the nesting of subelements within top-level elements, which provided the means of including delineated tracks within the metadata. The problem was finding a way to retain page-level metadata consisting of track-level relevant information while adhering to a less specific DC metadata standard. None of the existing workflows or technical processes had that capability to incorporate descriptive page-level metadata as they had been specifically designed for item-level metadata. After considering the time required to restructure item-level metadata plus the hierarchal file structure of the audio files, Digital Services determined that it was easier to upload the collection as it was, with its item-level metadata, and then export the metadata into a tab-delimited text file that contained technical metadata at the page-level added automatically by CONTENTdm during upload. Digital Services used the page-level technical metadata as a guide to place the page-level descriptive metadata within the export to prepare updated track-level information compiled into a spreadsheet. Digital Services wrote a Python script that would detect if descriptive page-level metadata is present in the spreadsheet, and then send instructions to CONTENTdm's Catcher API to add any page-level descriptive metadata. The final output was the inclusion of track names within the larger audio file to help users locate tracks and navigate within the audio collection. Dubbed the "field replacer script," this Python script could update any metadata field simultaneously for all the items rather than editing them one item at a time and by each individual metadata field. The creation of the field replacer script allowed Digital Services to make metadata additions or changes much more efficiently and at scale. Although it took time to write this new Python script, Digital Services has used this script multiple times and it will be a frequently used tool for future work.

With the transition to a new server, Digital Services created a new digital preservation workflow with the goal of depositing digital content and metadata into LOCKSS. The first step was to move files from network file storage to the new server. The prior digital preservation workflow intersected with Acumen, and Digital Services eliminated any dependencies within the Perl and Python scripts and changed locations to the new server. Digital Services normalized names and set staging areas within the network storage locations. By July 2020, Digital Services added its first digital content to the new server, moving closer to the goal of ingesting into LOCKSS.

Strategy of Enhancing Discovery

Borrowing from the famous line, "Build it, and he will come" from the movie *Field of Dreams*, suggests that if Digital Services created something that every researcher needs (in this case, amazing digital collections), users would visit the digital collections website. Increased and sustained web traffic does not happen this way for most digital collections. The strategy of enhancing discovery involves taking actionable steps to help users find and then use your digital collections. Telework provided the perfect environment to enhance the discovery of digital collections, which over time increases web traffic creating quantifiable statistics. Possible options for enhancing the discovery of digital collections include search engine optimization, social media promotion, metadata harvesting, research and subject guides, and integration with existing library systems.

Implementation of the Strategy of Enhancing Discovery

Getting users to discover a new DAM requires significant effort, and Digital Services focused on updating any resources including reference and subject guides and metadata harvesting that previously directed users to Acumen. One of the most forward-thinking parts of Acumen was the inclusion of persistent URLs (PURLs) both at the collection and item level, which allowed Digital Services to redirect PURLs to CONTENTdm upon the migration of a digital collection, creating a seamless transition for the user while still providing the opportunity to access non-migrated digital content through Acumen. Digital Services collaborated with library personnel to remove Acumen references across websites, videos, research guides, or update the resource to point to CONTENTdm. The Reference Services and Outreach Coordinator for Special Collections created a new digital collections research guide, and her work helped Digital Services to identify and fix lingering problems with PURLs for items within migrated digital collections.

With the ability to re-route PURLs to direct users to its new digital collections platform, Digital Services waited for a majority of digital collections to be migrated before transferring metadata harvesting from Acumen and enabling it in CONTENTdm. Metadata harvesters use OAI-PMH to create a metadata record and provide a URL to direct users to the item within a digital collection. Digital Services completed the transition with the statewide digital collection website, Alabama Mosaic, in January 2020, and the next harvester was EBSCO Discovery Service (EDS). The benefit of a discovery service is the creation of more of a one-stop search experience in which users can cross-search the catalog, databases, electronic journals, and digital collections. For example, a user searching for Autherine Lucy, who first integrated The University of Alabama in 1956, would see search results of published books and journal articles alongside pictures taken of her on campus during this time from two digital collections, Donn Sanford, and James Oakley, Jr. photographs collections. The inclusion of digital collections metadata in EDS exposes users who

would not have considered searching digital collections to the wealth of Special Collections and increases web traffic to CONTENTdm.

The University of Alabama Libraries plans to replace its existing integrated library system with FOLIO, an open-source library services platform jointly developed by libraries, vendors, and developers, and EDS will provide the public-facing discovery search tool for users. With this major change, the goal was to fix metadata harvesting in EDS before FOLIO implementation in fall 2020. EBSCO required the completion of an extensive institutional repository database questionnaire about a digital collection's metadata schema and a list of digital collections.[16] Digital Services collaborated with the Resource Acquisition and Discovery department, which manages the EDS, and submitted the questionnaire in January 2020. EBSCO began work in April 2020. After EBSCO harvested the metadata for the first time, Digital Services reviewed it within EDS, noting any mapping and visual display problems. Concerns were then submitted to EBSCO for resolution. This process continued until its completion in June 2020 and included the setup of an ongoing harvest schedule. The final design resulted in several improvements, with users being able to limit by the name of a digital collection and the use of the correct publication type icon displaying in the EDS interface (i.e., a periodical icon for a periodical).

In June 2020, Digital Services began switching the metadata harvesting in OCLC WorldCat's Digital Collection Gateway (DCG).[17] This tool uses OAI-PMH to transform a digital collection's metadata into MARC format and synchronizes the harvested metadata with WorldCat to create an item record. With the shutdown of the public-facing Acumen digital collections website scheduled for mid-July, users who found a WorldCat record for a digital collection of interest and clicked the hyperlinked URL were either redirected to CONTENTdm or discovered a dead hyperlink. Digital Services instructed OCLC to delete 97,501 WorldCat records harvested from Acumen and will be using the DCG to create WorldCat records from CONTENTdm over the next few months. Migration was the impetus for much of this work, but any institution can adopt the strategy of enhancing the discovery of digital collections while working remotely. By enhancing discovery and making it easier for users to locate relevant digital collections, the long-term benefit is more web traffic to an institution's digital collections.

Strategy of Try a New Approach

This strategy involves either trying something new that was never done before or significantly different from existing practices. It could incorporate the previously mentioned strategies, but a feeling of being challenged, and often apprehension, are the defining characteristics when executing this strategy. Managers should consider the size of an institution, staffing levels, and available resources in determining the level of difficulty and the scale of implementation of this strategy. Depending on the nature of the work, it may present the opportunity to collaborate with other library units that may have additional resources and staff now available with telework.

Implementation of the Strategy of Try a New Approach

At The University of Alabama, the most difficult strategy to implement was to try a new approach. The COVID-19 pandemic ushered in significant changes to everyone's daily lives, both personally and professionally, and coping with those changes was overwhelming. This pandemic is a moment in history that needs documentation for researchers now and in the future, and Special Collections collects this type of material for long-term preservation and access, but typically after an event and often many years later. Tasked with determining the parameters and logistics of collecting and preserving COVID-19 pandemic materials for Special Collections, a six-member team formed in April 2020 consisting of the Archival Access Coordinator, Reference Services and Outreach Coordinator, Institutional Records Analyst, and two Processing Archivists, with the Special Collections and Digital Initiative Librarian representing Digital Services. This paper provided more detail on the execution of this strategy specifically on the decision-making process to help other institutions emulate this type of project documenting the COVID-19 pandemic without a born-digital processing workflow in place.

The team began its work mindful of the internal resources, staffing levels, ongoing job responsibilities, and the unknown timing of a return to working on campus when the project was conceived. After consulting the Society of American Archivists' "Documenting in Times of Crisis: A Resource Kit," and looking at other institutions' projects already in progress that were created specifically in response to the COVID-19 pandemic, the team observed that institutions' scope of projects fell into three categories: physical materials only, born-digital materials only, or both physical and born-digital materials.[18] Special Collections did not accept born-digital materials without a physical media counterpart, resulting in a situation where most of these born-digital materials that existed on physical media remained unprocessed. The dilemma was that when most people routinely create born-digital content, such as taking cell phone pictures, they do not generate a physical counterpart that would be easy for Special Collections to add to its collections. The team decided that the significance

of documenting life during the COVID-19 outweighed the lack of born-digital processing workflow and proceeded with creating a project that allowed for the submission of both physical and born-digital materials. The method of delivery for born-digital content might end up in a digital exhibit, included in digital collections, or might take another form, as this was the first time Special Collections had taken this type of content.

The University of Alabama adopted the phrase and the hashtag #StillTideTogether during the pandemic, and the team named the project *Still Tide Together: Documenting Life during the COVID-19 Pandemic*. The team sought to define the project's scope by the audience, technical limitations, and internal resources. The team chose to emphasize the collection of materials from people who were connected with The University of Alabama, resided locally, or lived within the state and required all donors to be at least eighteen years old. Special Collections accepted only born-digital materials consisting of text, image, and sound files. This decision excluded video, social media, and websites due to the size of video files, the absence of a web archiving program, and the lack of experience with preserving these formats. With this type of project, it was not possible to predict the level of participation. The team created a dedicated webpage outlining the project's general information and collection policy and provided a link to a text-based questionnaire that consisted of a series of standardized questions to gather personal experiences. Respondents were asked to indicate interest in donating physical and born-digital materials to allow Special Collections to follow up with interested respondents.

Since Digital Services would be responsible for the digital preservation of the born-digital content, including the text-based questionnaire and possible inclusion within its digital collections, Digital Services held several Zoom meetings to determine what was technologically feasible and to develop a temporary born-digital workflow. Creating digital collections and preserving them demands alignment with digitization and metadata standards and structure with standardized file names, but the incoming born-digital content from the public would not adhere to these standards. The questionnaire platform needed to be able to export the data in an Excel spreadsheet or preferably a tab-delimited file to create Python scripts to parse the data. Digital Services realized the importance of creating structure and standardization within the design of the questionnaire but doing it in a way that was unobtrusive to the public filling out the questionnaire. In seeking a robust survey platform for the questionnaire, Digital Services tested several options and found that Qualtrics met its needs. It was free through the university's subscription, already branded with its colors and logo, had data validation, and offered great flexibility in creating questions and answer responses.

Digital Services adopted the mantra "Make them make the metadata for you" to reduce the time needed to normalize and transform the data to align with standards. The inclusion of the respondent's geographic area during the pandemic would be useful to future researchers since, for example, an out of state student from Oregon would have a different experience from an in-state student from Alabama. The DC field Coverage provided the geographic location in alignment with formatting and vocabulary in the Getty Thesaurus of Geographic Names. For example, the questionnaire asked for a respondent's country, state, and county (Alabama only), in a series of questions with a drop-down menu of pre-defined options except for city, which respondents enter in a free text response. By requiring respondents to enter the metadata, and only needing to verify respondents' city response for spelling and capitalization errors decreases the time spent on quality control and allows for scripted metadata transformation. Digital Services can write a Python script to parse the data to form a geographic location like United States--Alabama--Baldwin County--Mobile. The ability to script metadata transformation reduces the time-consuming work of manually creating metadata.

The most difficult decision was how to facilitate the transfer of born-digital content from the public to library servers easily and safely. Because of the anticipated short duration of this project, it was not prudent to implement a new server. Cloud storage, like Box, was an option, but the fear was that the public would submit unstructured digital content using a variety of filenames without any accompanying descriptive metadata. Abiding by the same mantra "Make them make the metadata for you," Digital Services envisioned a second questionnaire that would enable respondents to upload a file and describe the contents by answering a series of questions that Digital Services could parse and align to standards. If a respondent submitted a photograph, the respondent would need to provide the date, geographic location, and a brief caption describing who and what was happening in the image. Digital Services needed a platform that could pair descriptive metadata with the digital file and tested Google Forms, WordPress's Ninja Forms, and Springshare's LibWizard before choosing Qualtrics. This platform supports a variety of file types, enables multiple file submissions, and offers many export options. The deciding factor was Qualtrics' response id (e.g., R_1o89wqDgFbnCe0l), a unique identifier created from a random series of letters and numbers for each questionnaire submission and appended to the front of each a file name (e.g., R_1o89wqDgFbnCe0l_MyPandemicDiary). Although Qualtrics placed all the uploaded files in one zipped folder when exported, the response id allowed Digital Services to create a Python script to parse the data to match the file with its descriptive metadata. After data normalization,

Digital Services renamed the files using standardized file names and converted proprietary file formats to open file formats for long-term digital preservation.

With a draft of a born-digital workflow in place, the Special Collections and Digital Initiatives Librarian collaborated with the rest of the Still Tide Together team and advocated for Digital Services' technological needs to develop a viable project. Due to all the uncertainties associated with the pandemic, including the timing of the end of limited business operations, physical access to Special Collections, and unknown response rates, the team divided the project into three phases. Phase one was for the creation of a website and a questionnaire to gather text-based responses and to identify respondents who intended to donate physical and digital materials. In phase two, Special Collections contacted donors regarding the logistics of physical donations and would send a second questionnaire to facilitate born-digital submissions. Digital preservation and delivery of content would happen in phase three. The Still Tide Together project launched in May 2020, and by July 2020, Special Collections had received about fifty submissions to its text-based questionnaire, with donors intending to donate five to ten physical items and at least seventy-five digital files. By dividing this project into phases, the team has the flexibility to adjust its plans due to the uncertainty and fluid situation during a pandemic. As this project progresses to the next phase, Digital Services will test its proposed workflow and adapt it to meet the constraints of the data and files. Trying a new approach is challenging under any circumstances, and this project presents a practical small scale opportunity to prepare for the adoption of a born-digital program and to assess its feasibility.

Discussion

Digital Services excelled at the transition to telework through the implementation of the four strategies suggested above. Telework provided three lessons to those who led units like Digital Services. First, acknowledge the emotional toll of living through a pandemic and embrace a more flexible management style. The uncertainty of the future and unexpected changes affect everyone differently. Telework created a difficult and different situation for everyone, with employees balancing work, home, and family in a new work environment. In such situations, managers must be supportive and adjust expectations regarding how much and what type of work is appropriate to the employee's situation. Employees may have to work at night or need a few easy tasks assigned to them on days when they are not in an emotional space to remain focused on a complex project. This advice also holds true for managers. For example, writing a script requires intense focus and concentration, and that may be difficult to do when an employee is also simultaneously monitoring the online learning of a child, working at home with other family members and possibly sharing a computer, or caring for sick or elderly relatives. A manager should provide multiple work assignments at various levels of difficulty and empower employees to determine what to work on first based on how they are feeling mentally and emotionally. It gives employees a sense of control in their work when everything else feels out of control. Other employees may need more structure with the assignment of one project with small deadlines at intervals to provide a sense of accomplishment more frequently and to keep them on pace.

Second, educate stakeholders about the extent of the processes associated with your unit's work. Telework limits the ability to easily engage stakeholders, but each interaction is an opportunity to share your unit's story and what your team is accomplishing during telework. As the effects of the COVID-19 pandemic trickle down to institutional budgets, library administration will be making hard financial decisions. They must have enough information to fight devastating cuts and to preserve enough funding that supports the creation of digital collections when negotiating with higher levels of administration. In the case of digital collections, stress the entire process required to create a digital collection by shedding the production mentality of measuring success by the number of images scanned and uploaded to digital collections. Many managers have failed to convince stakeholders that productivity is more than production statistics. Instead, give a fuller account of what activities digitization units undertake to create and preserve digital content. That is not an easy task as it will have to be a continually repeated message that may slowly eradicate this misconception. Otherwise, the production statistics of teleworking make it seem as if an employee who spent time strengthening the digital infrastructure did nothing for months, while another employee uploaded thousands of pages of content to digital collections. Digital collections are more valuable than ever to researchers now when a pandemic has limited physical access to many libraries, archives, and special collections.

Lastly, accept the realities of the current situation. Library administration anticipated the end of limited business operations each month, but cases of COVID-19 continued to rise, topping 2,000 cases a day in July 2020. The ability to adapt and respond to these realities is what will allow managers to cope and lead their teams. The fact is that Digital Services will digitize fewer items this year. Within Special Collections, every manager is re-adjusting priorities as lack of access to the physical collections hinders ongoing cataloging, archival processing, and metadata creation, which all need to happen before Digital Services begins digitization. The loss of student employees further

diminishes the number of images digitized, as they digitized the majority of the content while Digital Services staff conducted quality control, executed technical processes, uploaded digital collections, and engaged in digital preservation. This acceptance of the new reality was the hardest lesson learned during telework, especially for any manager who had planned the work months in advance. The goals that managers set for their team might not be feasible during telework. Set small, achievable goals and celebrate those accomplishments. By adopting the strategies suggested in this paper, these accomplishments become realities, and a manager can be proud of the team and the work completed during teleworking.

Conclusion

Technology made telework possible, and libraries should incorporate telework into their business continuity and emergency management planning. Although not actively digitizing new content during telework, Digital Services completed projects aligned with the existing goals of the unit. The strategy of repurposing existing digital content resulted in the addition of 93,756 images and audio files to its digital collections from mid-March to July 2020. By focusing on improving the discovery of digital collections through metadata harvesting, revising subject and research guides, and removing references to Acumen throughout the website, more users are better able to discover and locate relevant digital collections more efficiently, leading to productive research and increased web traffic to digital collections. Digital Services strengthened its infrastructure through the development of additional scripts to automate parts of its workflow. The automation of the process of depositing digital content from the network drive to the new server represents the first step in the creation of a new digital preservation workflow. The Still Tide Together project forced Digital Services to try a new approach with the development of a temporary born-digital content workflow, which has the benefit of providing future researchers access to these materials that document the COVID-19 pandemic. Through the exploration of ingesting born-digital content with real-life examples, Digital Services has gained a better understanding of the logistics, resources, and technology needed to create a full-fledged born-digital program following best practices if Special Collections chooses to implement such a program in the future. The abrupt shift to telework provided many lessons, including acknowledgment of the emotional toll of living and working during uncertain times, the importance of educating stakeholders, and acceptance of a new way to work. This success was not possible without accounting for the pandemic in projects and the people who perform the work.

As the fall semester began in August 2020 with in-person classes resuming and the end of limited business operations, The University of Alabama Libraries had prepared with social distancing and altered library policies and procedures, and with Special Collections seeing university-affiliated researchers by appointment only. Digital Services returned to campus with staggered schedules to maintain social distancing while planning to work remotely the remainder of the time. With uncertainty ever-present during a pandemic and awareness that access to the physical collections may end at any time, Digital Services will add another strategy of stockpiling digital content. While on-campus, Digital Services primarily focused on digitization and left the remaining parts of the digital collection workflow for telework. This stockpiling strategy will better prepare Digital Services to transition back to full-time remote work if pandemic conditions worsen causing a return to limited business operations or if an employee has to quarantine due to COVID-19 exposure or a positive test result. The COVID-19 pandemic may continue for months or years or be replaced by something else that makes telework the new reality for Digital Services. Managers of similar units should incorporate the strategies of repurposing existing digital content, strengthening infrastructure, enhancing discovery, trying something new, and stockpiling digital content. These strategies make it possible to continue the ongoing work of creating digital collections.

References

1. Guy Robertson, "Pandemic Perspective: How an Outbreak Could Affect Libraries," *Feliciter* 52, no. 3 (2006): 111–13.
2. Will Manley, "Telework, or Watching Television," *American Libraries* 33, no. 4 (2002): 124.
3. Tricia Moe, "Personal View of Telecommuting," *Online* 29, no. 5 (2005): 25.
4. Melody Diehl Detar, "Theological Librarianship from a Distance." *Theological Librarianship* 8, no. 2 (2015): 11–15, https://doi.org/10.31046/tl.v8i2.390.
5. Jamene Brooks-Kieffer, "Librarians Considering Telecommuting, Consider This," *Library Connect Newsletter* (San Diego: Elsevier, 2004), https://libraryconnect.elsevier.com/articles/librarians-considering-telecommuting-consider.
6. Leah Black and Colleen Hyslop, "Telecommuting for Original Cataloging at the Michigan State University Libraries," *College & Research Libraries* 56 no. 4 (1995): 319–22, https://doi.org/10.5860/crl_56_04_319.
7. Richard Luce and Susan Hartman, "Telecommuting to

Work: Using Technology to Work at Home," *Library Hi Tech* 2, no. 4 (1984): 81–82, https://doi.org/10.1108/eb047576/eb047576.

8. Daniel Hickey and Neely Tang, "Theoretical and Applied Approaches to Remote Work for Academic Reference and Instruction Librarians," in *Library Staffing for the Future* (Bingley, UK: Emerald Group Publishing Limited, 2015): 181, 188–92, https://doi.org/10.1108/S0732-067120150000034008.
9. Jennifer Duncan, "Working from Afar: A New Trend for Librarianship," *College & Research Libraries News* 69 no. 4 (2008): 215–18, https://doi.org/10.5860/crln.69.4.7972.
10. Dawn Smith and Teresa Van Dyke, "A Telecommuting Interlibrary Loan Librarian's Experience: The Views of Both the Telecommuter and the On-site Supervisor," *Journal of Interlibrary Loan, Document Delivery & Electronic Reserves* 18, no. 4 (2008): 449–55, https://doi.org/10.1080/10723030802181828.
11. Manley, "Telework, or Watching Television," 124.
12. Brooks-Kieffer, "Librarians Considering Telecommuting."
13. Hickey and Tang, "Theoretical and Applied," 181.
14. Monica Rysavy and Russell Michalak, "Working from Home: How We Managed Our Team Remotely with Technology," *Journal of Library Administration* 60, no. 5 (2020): 532–42, https://doi.org/10.1080/01930826.2020.1760569.
15. Benjamin Walsh and Harjinder Rana, "Continuity of Academic Library Services during the Pandemic The University of Toronto Libraries' Response," *Journal of Scholarly Publishing* 51, no. 4 (2020): 237–45, https://doi.org/10.3138/jsp.51.4.04.
16. EBSCO, "EBSCO Discovery Service (EDS)—Institutional Repository Database Questionnaire," accessed January 2, 2021, https://connect.ebsco.com/s/article/EBSCO-Discovery-Service-EDS-Institutional-Repository-Database-Questionnaire?language=en_US.
17. OCLC, "WorldCat Digital Collection Gateway," accessed January 2, 2021, https://www.oclc.org/en/digital-gateway.html.
18. Society of American Archivists, "Documenting a Crisis: A Resource Kit," accessed July 23, 2020, www2.archivists.org/advocacy/documenting-in-times-of-crisis-a-resource-kit.

Book Reviews

Elyssa M. Gould

Sudden Position Guide to Acquisitions. Deborah Hathaway, Paul Kelsey, Stacey Marien, and Susan E. Thomas. Chicago: ALCTS Publishing, 2020. 86 p. $30.50 softcover (ISBN 978-0-8389-4849-1); $15.00 e-book (ISBN 978-0-8389-4850-7)

The third in the ALCTS Sudden Position Series tackles acquisitions and promises an easy-to-read introduction to the responsibilities covering "essential knowledge, tools of the trade, and best practices" (ix). At a slim eighty-six pages, someone "suddenly" in acquisitions, or those preparing to interview or start a new position, will still find a lot of ground covered.

For time pressed readers, chapter 1, "What You Absolutely Need to Know," covers brief overviews of ordering and types of orders, licensing language, accounting and budgeting practices, and collaboration with other departments. The librarian new to acquisitions may not be familiar with the various purchasing models that they will soon use and it may benefit these readers to have access to a glossary that included the common terms, especially while reading this first chapter. Notably, on page 3, the authors mention "approval plans" but do not define what this model encompasses until pages 44–48 in the next chapter. The section on "Common Ledger Vocabulary" has duplication of terms and it could also be confusing to those new to this language.

Management of staff may not be the first concern of someone new to an acquisitions position, but the authors' inclusion of this topic in the first chapter is practical. Though there is only so much that can be discussed in an overview, this section provides some starting points for anyone new to managing others. Getting to know staff through meetings, scheduling training, and simply being approachable are some beginning steps suggested, especially before trying to introduce new ideas and workflows.

In chapter 2, standard procedures and best practices are outlined. This chapter is dense with information and one the new-to-acquisitions librarian will probably return to as they encounter these situations and workflows on the job. The authors give an ample explanation of the basics of purchasing and subscription models, and briefly describe the patron-driven models of Demand-Driven Acquisitions (DDA) and Evidence-Based Acquisitions (EBA). However, the authors miss an opportunity to also introduce how the increasingly popular streaming video and other media fit into these models and do not do so until later in chapter 3.

Especially worth noting from the third chapter, "Things You May Encounter," is the coverage on developing a spending forecast and dealing with budget reductions and cancellations. The authors explain how forecasting is essential to planning for new resources and expected increases for current holdings. They also provide advice on where to start when experiencing decreased funding while trying to maintain quality of service. Working with reports of various types related to orders and library collections are covered, as the acquisitions librarian will at least need to consult, if not be the one to create, most of them.

The widely-used software Microsoft Excel is deservedly (though briefly) mentioned, along with other "tools of the trade" in chapter 4. This chapter includes a useful list of vendors and publishers. The library's Integrated Library System (ILS) will likely be a heavily used tool for many in acquisitions and navigating the ILS' acquisitions module is a skill the new librarian will learn primarily on the job. Since there are many ILSs, this text does not discuss the specifics of any one system, besides the inclusion of two screenshots in the first chapter of the ledger structure displays in Ex Libris' Voyager and a SirsiDynix ILS. The authors intentionally use multiple chapters to provide general overviews of how the librarian may be using their ILS, including creating purchase orders, receiving materials, paying invoices, and extracting reports.

The authors use the final chapter to suggest books, articles and journals, courses, listservs, and professional groups for continuing education and professional development. This is a good reminder for the new and possibly overwhelmed librarian that there are resources and professional activities available to increase experience in and understanding of acquisitions.

The "Sudden Position Guide to Acquisitions" is a

successful addition to the series that covers areas of library technical services. While some sections may have benefited from more clarity and a different arrangement of the information, those new to acquisitions, or even library staff interested in understanding some of the functions of this work, will be able to quickly extract the background knowledge needed for initial success. This book could also be a title the new librarian consults as they develop the skill sets and workflows for their role in acquisitions.—*Audra M. Deemer (adeemer@depaul.edu), DePaul University, Chicago*

Sudden Selector's Guide to Geography and GIS. By Carl Olson and Kim M. Ricker. Chicago: ALA, 2020. 93p. $30.50 softcover (ISBN: 978-0-8389-4771-5).

As stated in the foreword, "The ongoing purpose of the sudden selector's series is to provide current information on selection in specific subject areas in order to assist selectors in creating a manageable process in unfamiliar subject territories" (vii). This new entry in the series does just that; it provides new geography and geographic information systems (GIS) selectors with a baseline knowledge of the field. It accomplishes this by introducing the subject and a broad review of valuable tools and resources.

This book can be viewed as a blueprint for a new selector's beginner "toolbox" with three shelves of tools. The top shelf contains tools for general collection development. The middle shelf holds tools in the form of information on the subject and field of discipline. The bottom shelf includes tools for selecting materials, including advice on what software and hardware a library would need to support this discipline.

Basic collection development tools are in the top layer. Although the authors state that their goal is not to teach collection development, many resources are shared on the subject. The reader will find recommendations for books on general collection development, including six different titles and accompanying summaries. Additionally, five different review sources and their specialties are discussed, including *Choice* and *Booklist*. This layer is rounded out with descriptions and links to three electronic discussion lists and websites. This is the thinnest layer in the toolbox, but a helpful one for any new selector.

The middle layer of the toolbox contains material explaining the subject: from the basic question, "What is geography?" to the more complex question, "How do I build expertise?" The author likens geography to astronomy to explain the concepts, but this analogy falls flat. Later in the text, the authors provide a more illustrative description, stating, "geography finds its ancestry in legendary voyages and travelers' tales" (2). This leads into a review of the field with "Three Core Concepts, Two Branches, and Four Traditions" (4). The "methods and workflow of geography" are grouped into "books, boots, and benches" (7), showing how the field emphasizes learning the literature, conducting field work, and doing lab work to analyze findings. Next, the author provides examples of recent research that serve as great starting points for discussions with patrons and touchstone examples for future reference. The authors discuss examples of what GIS can do, including how "it allows us . . . to find, understand, interpret, and perhaps even question relationships and patterns that are shown based on geographic location" (21). Later in the book, readers learn how to build expertise in the field through published literature and are given further information for formal and online study. Finally, the importance of networking to stay current and to obtain assistance when needed is emphasized, accompanied by a helpful list of conferences and societies.

The bottom layer of the toolbox helps new selectors determine what is needed to support the discipline, including hardware and software, plus explanations of different types of data. The selection advice is split across several chapters, with recommendations for selecting reference materials, such as maps, journals, databases, and software. Multiple examples of each type and explanatory notes are provided. For example, the section on maps discusses cartography and usages for different types of maps. Major publishers and call numbers are listed. The section on software and vendors is invaluable. It describes the audience, price, functionality, user interface, and hardware compatibility of each title, followed by information on open-source content and how to acquire the data at the core of GIS. These chapters provide useful advice regarding how to use LibGuides and collection development policies effectively.

The material in this book is dense and wide-ranging, making it a handy reference book for selectors, and a great introduction to the subject and field. This reviewer strongly recommends this book as a starting point for selectors new to geography or GIS. Geography is a complex field, but this book gives selectors a quick foundation to jump into their new role and provides a clear structure for continued learning.—*Tamara Bozich (tbozich@ucsd.edu), University of California, San Diego*

Sudden Selector's Guide to Philosophy Resources. By George J. Aulisio. Chicago: ALCTS Publishing, 2020. $30.50, 100p. (ISBN: 978-0-8389-4843-9). ALCTS Collection Management Section Sudden Selector's Guides Series.

The *Sudden Selector's Guide to Philosophy Resources*, the ninth volume in its series, is a succinct introduction for the library professionals newly responsible for collection management and research assistance for the discipline. Its six chapters provide a broad overview of academic philosophy, issues of audience, common formats, flagship resources, and the financial aspects of effectively managing a philosophy collection.

As author George J. Aulisio notes, "Understanding philosophical concepts can be difficult even for trained philosophers" (1) and as one of the earliest and the broadest disciplines, to even know where to begin can be quite intimidating. Despite the complexity of the field, Aulisio provides an admirable breakdown of the traditions, subfields, and methodological approaches commonly used, as well as including a list of relevant professional organizations, conferences, and awards at the end of the first chapter. The latter resources provide a good reference for librarians seeking a more in-depth knowledge of philosophy.

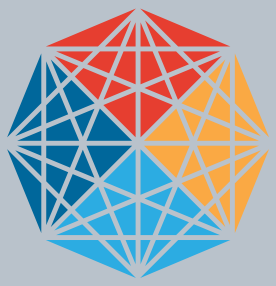
Budgetary constraints and the prioritization of certain materials are duly addressed, which will doubtless be quite valuable to readers of this book as financial stresses from the COVID-19 pandemic are unlikely to ease in the near term. Heavy focus on monographs both as research material and academic output and a proliferation of serial titles both general and specific in their disciplinary coverage mean that philosophy resources tend to be more costly than those of the other humanities and social sciences, excluding business. Discussion of package deals, major indexes, publishers, and web resources, including suggestions of resources to prioritize based on institutional demographics and collection budget are another welcome inclusion. Of one index, the author boldly encourages librarians faced with a subscription request from the department "suggest that the philosophy department should pay for this resource directly" (56) due to its lack of academic value. Such blunt honesty may not endear the selector to their department, but it indicates the useful commentary this book provides alongside the expected resource lists.

Chapter 5 specifically highlights continuing professional development and networking for librarians, including the Association of College and Research Libraries' Philosophical, Religious, and Theological Discussion Group and e-mail discussion lists and podcasts that Aulisio suggests may be well-suited to a philosophy subject guide (76).

A theme running throughout the book is the importance of understanding the characteristics of one's audience and its needs, plus the role of the philosophy department within one's institution, stating "many Catholic institutions require multiple philosophy courses" but another "may not require philosophy, but may have a highly regarded philosophy program that grants advanced degrees" (24). Readers are encouraged to get to know and build relationships with faculty and staff, to learn their preferences and research areas, as well as the broader direction of the department overall. Less emphasis is put on the need for cooperation with library colleagues due to the cross-disciplinary nature of philosophy, but subject overlap and package purchasing decisions mean that collaboration is likely unavoidable, if the philosophy and religion selectors or liaisons are not, in fact, the same individual, as may be the case at some institutions.

This book contains a few brief mentions of foreign language material, particularly primary source texts and European journals, but lacks in-depth discussion of how a new selector could approach this area of the collection, particularly if they lack the language skills to comfortably engage with these resources. It is also heavily focused on Western philosophical traditions (specifically Europe and North America), though Eastern traditions and a few major thinkers are mentioned in the overview section. Ideally more space would be given to these fields, but in-depth coverage may be viewed as out of scope for a book published by a subdivision of the American Library Association and presumed to have an audience of primarily North American and western European readers. These topics may also be better suited for discussion in a book targeted toward area- or religious-studies selectors instead of a philosophy selector.

Overall, this book is a worthy introduction to academic philosophy librarianship for those new to the role and does a commendable job providing a general introduction to the field, going beyond simply providing practitioners with a crash-course listing of relevant library resources.—*Maggie Halterman-Dess (margaret-halterman@uiowa.edu), University of Iowa Libraries*



Core

LEADERSHIP
INFRASTRUCTURE
FUTURES

A DIVISON OF THE AMERICAN LIBRARY ASSOCIATION
225 North Michigan Avenue, Suite 1300, Chicago, IL 60601 • 312-280-5038
Fax: 312-280-5033 • Toll Free: 1-800-545-2433 • http://www.ala.org/core/