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

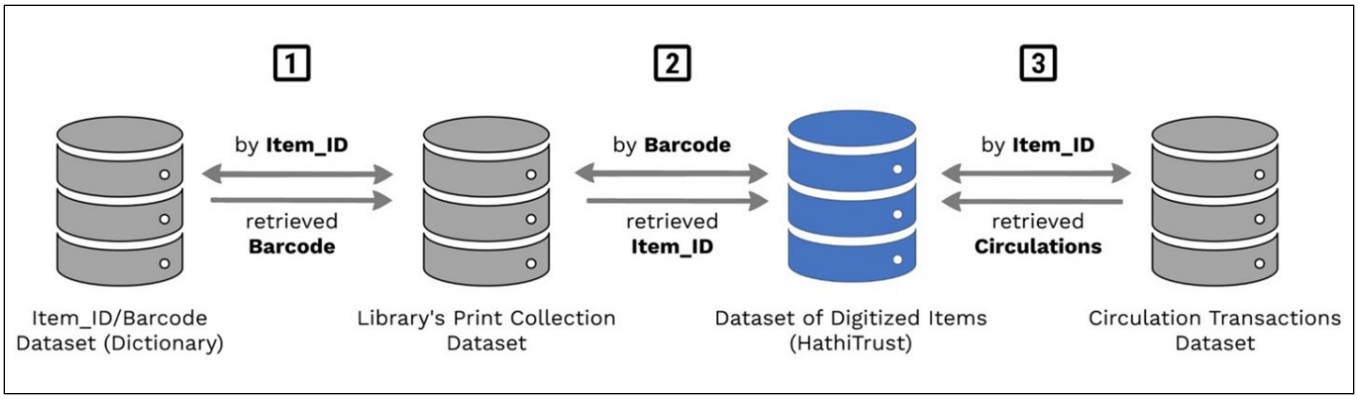

Figure 1. Retrieval of the Circulation Information for the Digitized Items

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.

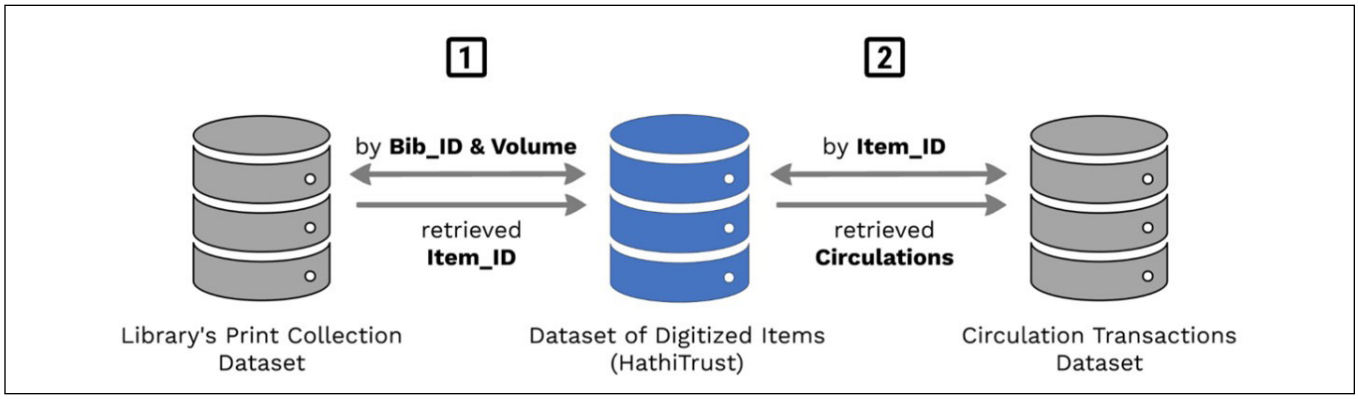

Figure 2. Retrieval of the Circulation Information for the Digitized Items by Bibliographic Identifier and Volume Information

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

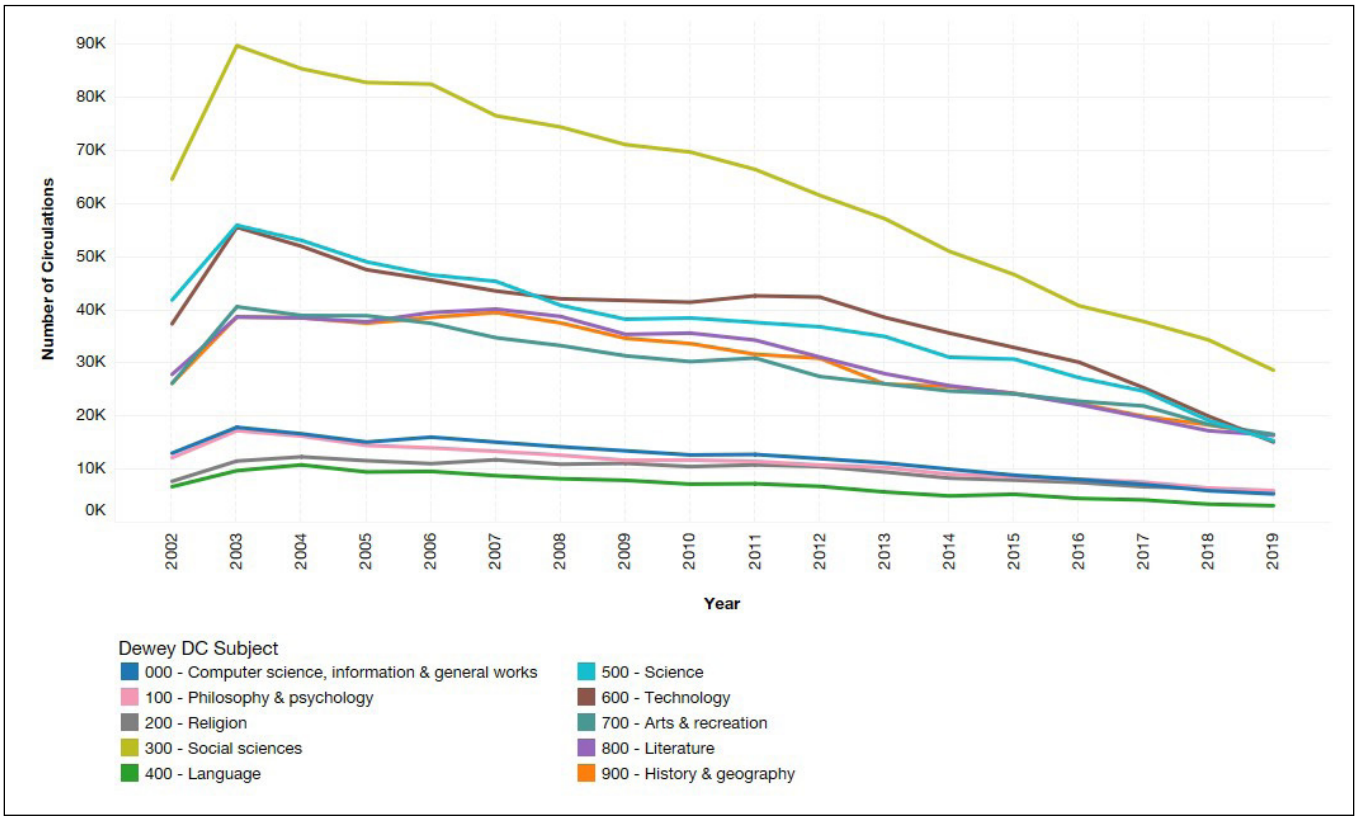

Figure 3. Circulations of the Dewey Decimal Classification Subjects by Year

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,

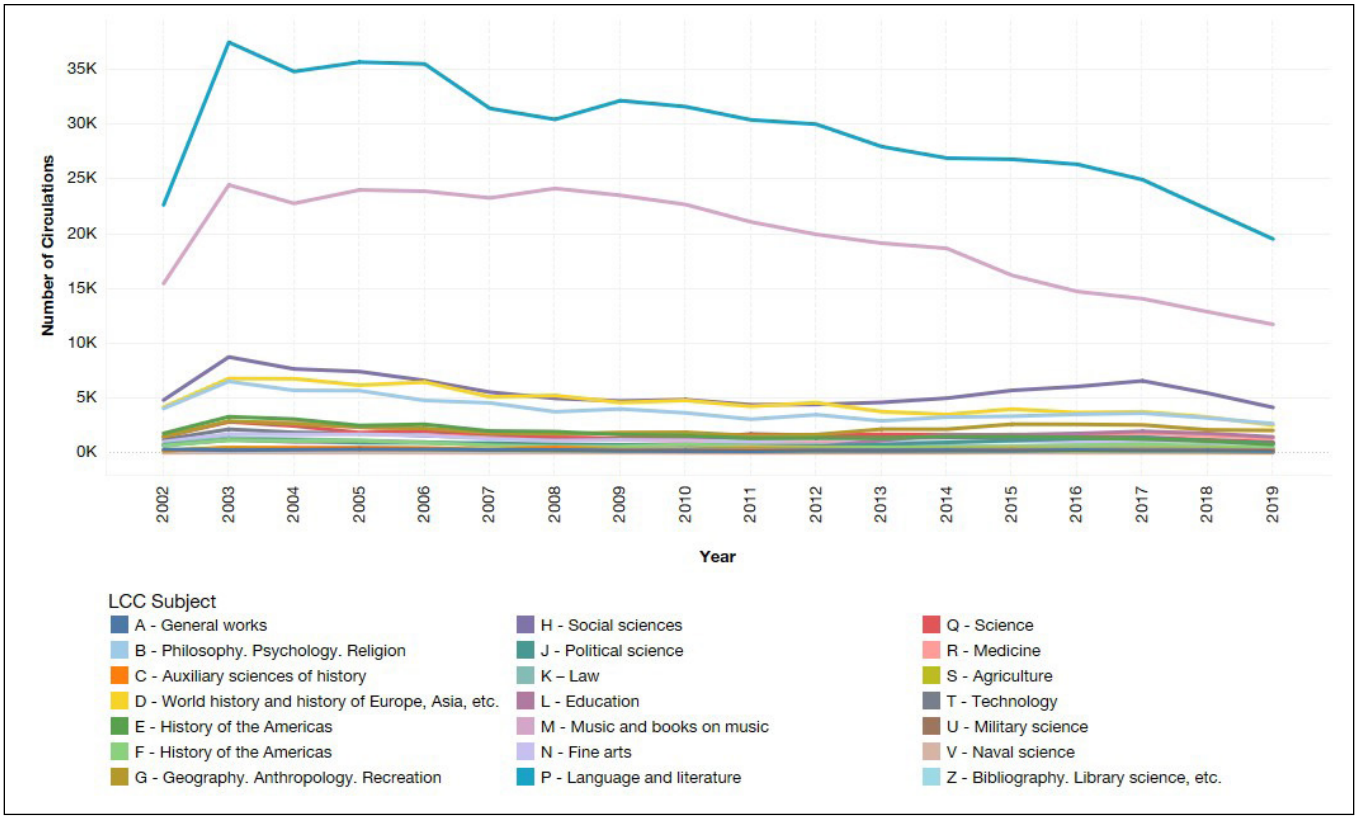

Figure 4. Circulation of the Library of Congress Classification Subjects by Year

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

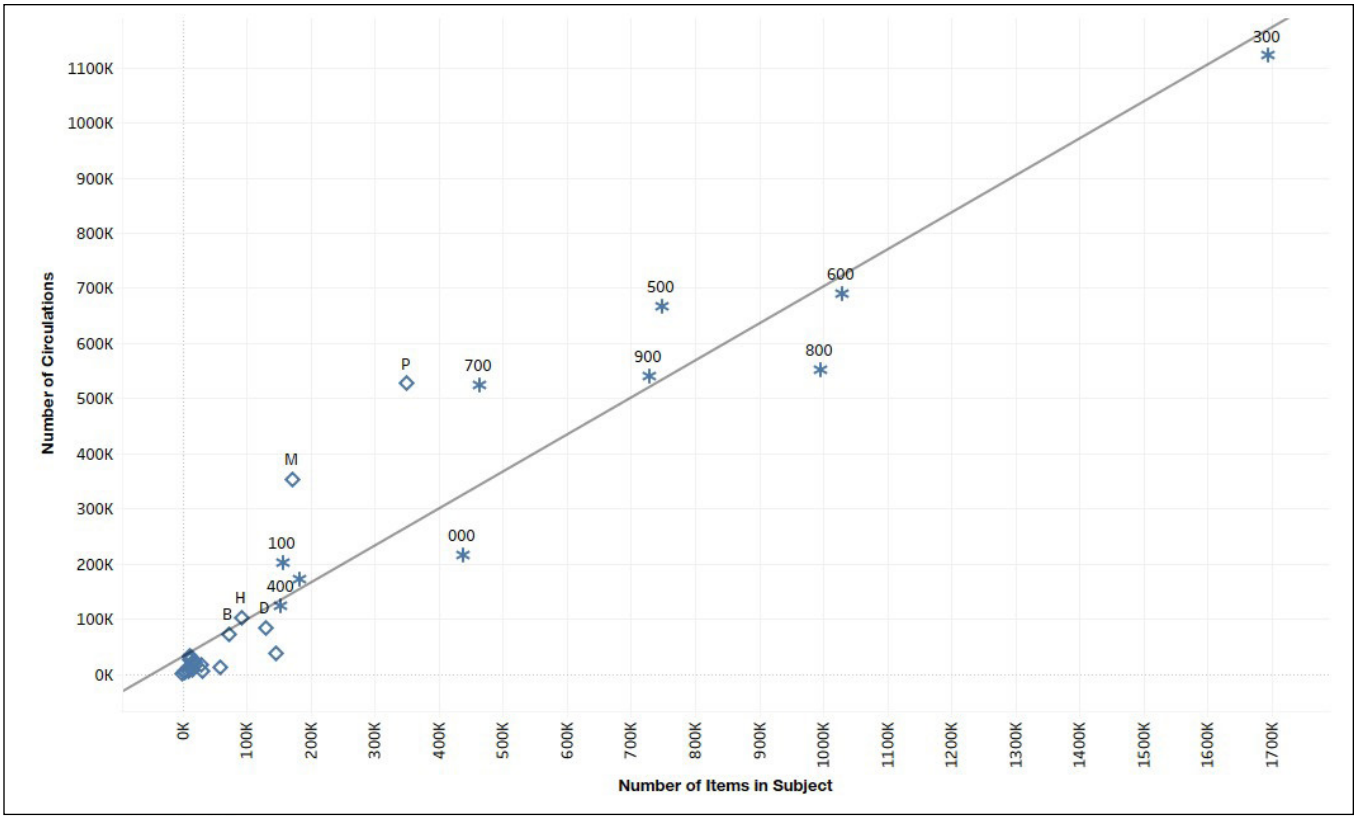

Figure 5. Correlation between Total Number of Items in the Subject and Total Number of their Circulations

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

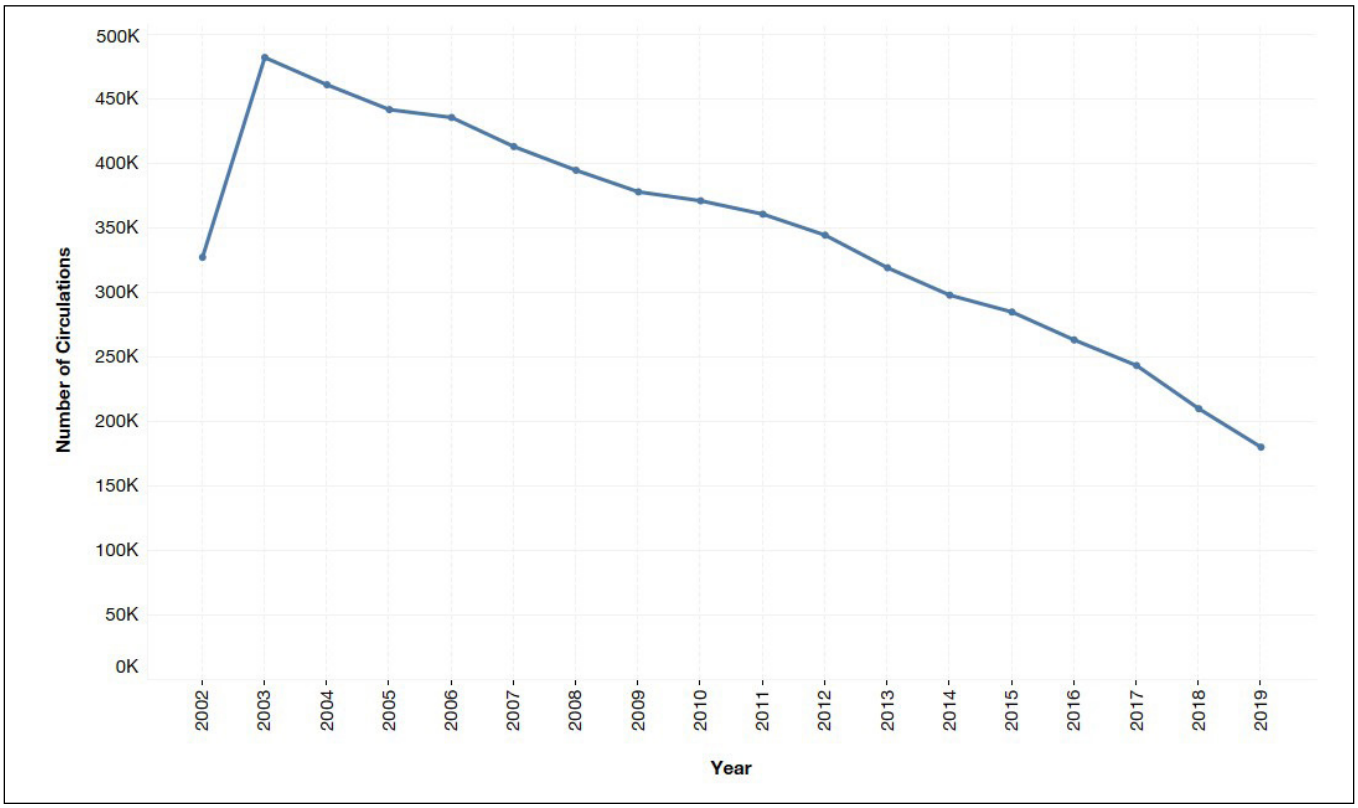

Figure 6. Annual Number of Circulations for Library's Print Collection

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.