Consortial Book Circulation Patterns: The OCLC-OhioLINK Study

Edward T. O'Neill and Julia A. Gammon

The OhioLINK consortium and OCLC Research collected and analyzed circulation data for libraries within the consortium. The study, which examines the circulation of 28,475,701 items from more than 100 academic libraries, is the largest and most diverse compilation of academic usage data for books ever collected. The authors outline the study methodology, analyze the data, and offer insights into the consortium wide collection usage. Circulation patterns analyzed include subject and language usage, as well as scattering, obsolescence, and duplication. The study results and conclusions offer valuable insights for library collection planning including library purchasing, storage considerations, and future usage prediction.

Librarians have long been collaborators. The history of cooperative collection development among libraries goes back hundreds of years. Economic, social, and technological changes during the 20th century supported significant growth in library consortia and enabled libraries to collaboratively build their collections. In today's environment, it matters less who owns resources than it does that users can quickly access them.

OhioLINK is a major player among library consortia and has been a leader in promoting resource sharing and cooperative collection development. Resource sharing is a primary goal throughout Ohio's academic libraries and is supported by a strong delivery system, a union catalog, a shared online system, and a history of cooperation.

OhioLINK may be the only library consortium created because of a space problem.[1] Due to the burgeoning of library print collections in the mid-1980s, the Ohio Board of Regents was presented with many requests for new or expanded library buildings. Knowing that collections would continue to grow, the Ohio General Assembly directed the Ohio Board of Regents to find an alternative to library construction. Subsequently, in 1986, the Library Study Committee was appointed. This committee went well beyond its charge of solving the collection storage issue and recommended a transformation of Ohio's academic libraries to coincide with the electronic information age by creating a statewide online catalog system with a delivery component.[2]

Presented with a unique opportunity to study the collection usage in a library consortium with more than 100 libraries and 600,000 potential users, the OhioLINK Collection Building Task Force and OCLC Research agreed to jointly compile and analyze the circulation data. The goal is to gain a better understanding of how the libraries'

Edward T. O'Neill, Senior Research Scientist (Retired), OCLC Research; e-mail: edwardtoneill@gmail.com. Julia A. Gammon, Professor Emeritus of Bibliography (formerly Head of Acquisitions) University Libraries, University of Akron; e-mail: jgammon@uakron.edu. ©2014 Edward T. O'Neill and Julia A. Gammon, *Attribution-NonCommercial (http://creativecommons.org/licenses/by-nc/3.0/) CC BY-NC*

resources are used and to identify how the limited resources of the libraries can be used more effectively.[3] The study is limited to analyzing the circulation of printed books and manuscripts and does not include eBooks. Although electronic resources are growing in importance and will play an increasingly significant role in supplying scholarly information, print is not dead and will continue to be an important source of scholarly information for the foreseeable future. Academic libraries have huge collections of print and continue to invest significant amounts in the acquisition of print materials. It is estimated that the OhioLINK libraries have invested over a billion dollars in acquiring, cataloging, and maintaining their print book collections and will need to maintain and manage these collections for many years. Furthermore, it is expected that many of the usage patterns observed for print will also apply to electronic resources.

OhioLINK Consortium

OhioLINK and Innovative Interfaces Inc. jointly developed the shared online system used by the consortium. By 1994, the system—which allows both local circulation and patron-initiated off-campus borrowing—was operational. Since all member libraries use a common circulation system, consistent circulation data was available. Users can use unmediated online borrowing of off-campus books and pick up the books in 2–3 days at a location of their choosing.

Virtually all academic libraries, both public and private, in Ohio are OhioLINK members. The various types of libraries in the consortium are shown in table 1. OhioLINK libraries face the same dilemma encountered by institutions in most consortia—balancing local collection needs against the collective needs. Historically, collections were developed to meet the needs of the local host institution. While academic libraries would share their resources via traditional interlibrary loan, the off-campus use was not a significant consideration for collection development.

However, unmediated borrowing with rapid delivery shifted the focus to the collective collection—the combined resources of the consortium. Emphasizing cooperative purchases enhanced the depth and breadth of the collective collection as a whole. Reductions in library funding also forced librarians to look at how they spent their money based on their collection strengths, research agenda, and usage. Unnecessary duplication within the consortium became apparent, prompting consideration of the optimum number of copies of any one title the consortium should purchase to satisfy both immediate and long-term research needs. Nevertheless, a tension remains between building strong local collections and building a strong collective collection. Effective collection management demands a balance between local and consortial needs.

Literature Review

Since this study is limited to printed books and manuscripts, the literature review focused on studies based on circulation data. Previously, the University of Pittsburgh Study was generally considered to be the most comprehensive study of collection usage.[4] In its day, the Pittsburgh Study was quite controversial after finding that 40 percent of the books added never circulated.[5] That study was done in the 1970s and covered only a single

TABLE 1 OhioLINK Institutions	
Institution Type	**Number of Institutions**
ARL Institutions	5
Universities	11
Colleges	44
Community Colleges	15
Branch Campuses	28
Depositories	5
Independent Cultural Institutions	3
Off-campus Hospitals and Medical Centers	20
State Libraries	1

institution rather than a consortium of over one hundred libraries. Joseph J. Branin, then-editor of *College and Research Libraries,* compared the findings of the Pittsburgh Study and the still-unpublished results of the OhioLINK study.[6] He observed that, after thirty years, even a consortium as well organized as OhioLINK, still had steady duplication rates in the print monographic collection, while usage continued to decline.

One early pioneer in studying library collection usage was Richard Trueswell, whose work in the 1960s identified the 80/20 rule, which was first applied to industry and warehousing, and eventually applied to libraries.[7] His theory that 80 percent of the usage can be covered by 20 percent of the collection has been a standard assumption among librarians for many years.

The 1980s seemed to be a period in libraries where reflection and developing procedures for conducting studies were developed. Broadus summarized some of the previous single institutional use studies.[8] Extrapolating from the various studies, he observed that (1) many books go uncirculated for years, (2) use of materials within the building was similar and proportional to those that circulated, (3) previous use was a good predictor for future use, (4) recent materials received the most use, and (5) Americans did not use a large quantity of non-English publications.

In 1985, Trochim outlined a procedure manual, or "cookbook form" as she called it, for small academic libraries wanting to develop their own individual use studies.[9] Not only did the book provide procedural instructions, but it described how to interpret the survey results. This made it possible for smaller libraries to take advantage of the experiences and expertise of larger institutions.

One of the longest-running circulation studies was done by the Lingnan University in Hong Kong. This fifteen-year study in a non-Western liberal arts university included a collection of books in both English and Chinese. Its results supported previous studies with the finding that 30 percent of the books never circulated and that previous circulation was a good predictor of future circulation.[10]

Few studies of collection usage across multiple libraries or in large consortia were found in the literature. The study done by the CARLI consortium in Illinois provided some similarities to the OhioLINK study, including patron-initiated borrowing, but it was limited to a span of five years of usage data based on copyright dates (2003–2008) vs. the OhioLINK study that included all volumes in the libraries' collections.[11] The CARLI study included about 25 percent fewer libraries. Findings indicated that (1) the use of the collective collection was high for the five-year period, (2) the consortium as a whole was collecting broadly, and (3) areas of highest number of purchases had the highest usage.

In any large study, it is critical to clearly define the bibliographic units and consistently use terminology. While a variety of terms were found in the literature, the terminology defined by the Functional Requirements for Bibliographic Records (FRBR) was adopted for this study.[12] The following FRBR definition of the group one entities are used:

Work: *A distinct intellectual or artistic creation.* Stieg Larsson's *The Girl with the Dragon Tattoo* is an example of a work.

Expression: *The intellectual or artistic realization of a work.* The English-language translation by Reg Keeland of *The Girl with the Dragon Tattoo* is an example of a particular expression of the work. Although important in other contexts, expressions are not explicitly used in this study.

Manifestation: *The physical embodiment of an expression of a work.* In this study, each bibliographic record in the OCLC WorldCat® database is considered to be

a manifestation. The first American edition published by Alfred A. Knopf (2008) is an example of one manifestation of the English-language translation of *The Girl with the Dragon Tattoo.*

Item: *A single exemplar of a manifestation.* An item is a particular copy of a manifestation. In this study, each circulation record is assumed to represent a single item. The copy of the first American edition of *The Girl with the Dragon Tattoo* held by Ohio Wesleyan University is an example of an item.

Extensive information on other aspects of the OhioLINK project is available on the project website and includes the project report.[13] The project report (1) presents an overview of the study, (2) provides some of the preliminary findings, (3) describes the data in detail, and (4) provides additional information on how the data were collected and processed. The website also includes extensive tables on the size and usage of the collections at each of the participating institutions. Preliminary results of the study were presented by O'Neill and Gammon at the ACRL 14th National Conference.[14] The results presented here differ from the previous publications by focusing on OhioLINK collective collections and are the first results using the publicly released data rather than the undocumented and proprietary form of the data previously analyzed.

The OhioLINK Data

The OhioLINK data consist of two distinct files; the circulation file and the bibliographic file. The circulation file is limited to records for print books and manuscripts. There are a total of 28,475,701 circulation records, one record for each item held by a participating OhioLINK library. The bibliographic file includes 6,660,250 records (manifestations) in MARC21 format from OCLC's WorldCat, one record for each unique manifestation in the OhioLINK catalog.

Circulation records from participating institutions were collected for a one-year period beginning in the spring of 2007. The raw data from the local circulation systems were extensively processed and reformatted to get the data in a form suitable for analysis. The fields in the revised circulation records are shown in table 2. The *Item No.* is an item-level identifier and is unique within the institution supplying the data. The *OCLC No.* is a manifestation-level identifier and can be used to link items at different institutions. Although works are not discussed in this paper, the *Work No.* is included in the data. In this study, works were identified using the OCLC workset algorithm,[15] which was developed at OCLC Research to cluster all manifestations of a work together.

The *Source* is the supplier of the data and the *Campus* identifies the particular location—main campus, branch campus, or other location. For the larger institutions, the *Administrative Unit* is the top hierarchical unit on the campus holding the item. For example, at The Ohio State University there are three distinct administrative units: University Libraries, Health Sciences Library, and the Law Library. When the administrative units have multiple physical locations, the *Administrative Subunit* identifies the particular library where the item is located. The administrative unit and the subunit are blank if those units are not relevant, as in a community college with a single library.

The *Accession Date* is the date the item was acquired or entered into the system. For older materials, this is generally the date the item was entered into the circulation system. The *Date of Last Use* is the last time the item circulated. A *Circulation Status* value of 1 indicates that the item can be checked out; a status value of 0 indicates that it is noncirculating. An item was considered circulating only if it was generally available for use by faculty, staff, and students. The circulation status of items can change—for instance, an item might be placed on reserve—so its status at the end of the study could

| | | TABLE 2 OhioLINK Circulation Record | |

Data Element	Sample Values	Definition
Item No.	1147735	An identifier that is unique for each item within a source file. The combination of source ID and item number is unique.
OCLC No.	1572109	The OCLC number for the corresponding World-Cat bibliographic record.
Work No.	167775	The work number as identified by the OCLC Workset algorithm. The work number was determined after the first phase of the data collection and has not been updated.
Source	Akron	Name of the institution supplying the data.
Campus	University of Akron	The name of the campus holding the item.
Administrative Unit	University Libraries	The name of the administrative unit for the libraries holding the item. This will be null if the campus does not have administrative units.
Administrative Subunit	Bierce	The name of the subunit within the administrative unit generating the circulation. This will be null if the administrative does not have subunits.
Accession Date	4/12/1993	The date the item was acquired or entered into the system, whichever is later.
Date of Last Use	4/17/2003	The date the item last circulated.
Circulation Status	1	A code indicating whether the item is circulating or noncirculating (1—item is circulating, 0—item is noncirculating)
Total Circulation	9	The total number of times the item has circulated since being entered into the system.
Annual Circulation	0	The number of times the item circulated in the year observed.

differ from its initial status. The circulation status is based on conditions at the end of the study. The *Total Circulation* is the cumulative number of times the item circulated since it was initially entered into the system, and the *Annual Circulation* is the number of times the item circulated in the year observed. In some cases, noncirculating items may have circulated if their status changed or if only users such as faculty were allowed to check out the item.

Analysis

The analysis presented here is limited to the annual circulation. In computing circulation rates, only circulating materials are considered to obtain more reliable estimates. All of the software used for the analysis was developed specifically for this project, and the analysis was performed on Windows®-based computers. Microsoft Excel® was used to format the tables and create the figures.

TABLE 3
Distribution of Resources

	Number of Items	Percent Noncirculating	Number of Manifestations	Unique Manifestations	Annual Circulation	Circulation Rate	Average Age
ARLs	7,445,927	18.8%	1,980,146	1,245,012	868,277	0.144	37
Universities	7,466,865	19.1%	1,616,656	740,382	872,741	0.144	33
Colleges	6,849,675	8.4%	1,439,421	572,093	706,125	0.113	37
2-yr Campuses	1,679,366	8.2%	174,835	90,138	239,746	0.156	29
Depositories	4,181,392	4.5%	1,299,211	866,085	101,177	0.025	46
Others	852,476	12.7%	149,981	95,485	33,969	0.046	27

Distribution of Resources

Library resources are widely distributed across all types of libraries, as shown in table 3. The non-ARL universities held 7,466,865 items, the largest number for any type of library. However, the combined holdings of the five ARLs are almost as large and the college libraries collectively held nearly as many items. The depositories held a large number of the lesser-used items. The ARLs have been the biggest contributors to the depositories; and, if the items in the depositories had been counted with their contributing libraries rather than the depositories, the ARL collections would have appeared to be much larger and their circulation rate significantly lower. The total circulation of the ARLs and the universities was almost identical and only a little higher than that of the colleges libraries. However, the ARLs held the most manifestations and the most unique manifestations. A unique manifestation as used in table 3 refers to a manifestation that is only held by that type of library. The ARLs held 1,245,012 manifestations that are not held by any non-ARL libraries.

Most libraries held a substantial number of noncirculating items: reserve materials, rare books, and the like. The percent of noncirculating items in the ARLs and university collections, at nearly 20 percent, was particularly high. The *circulation rate* is the average number of times an item circulated during the year observed. Only the items that are allowed to circulate are considered in computing the circulation rate. The very low circulation rate of items in the depositories validates the selection policies—low use was the primary consideration in selecting the items for the depositories. The age of the collections was also very similar, except for the two-year campuses (which have much newer collections) and the depositories (which have much older collections).

Duplication Rates

Duplication is an important and frequently discussed issue. How much duplication is necessary; how much is excessive? Every dollar spent on duplicating resources is a dollar less that is available to acquire unique materials. Figure 1 shows the level of duplication, based on the date of publication, for the last hundred years. The duplication level was defined as the average

number of items per manifestation. The definition of item used within OhioLINK differs slightly from the FRBR definition. In the OhioLINK records, each individual volume of a multiple volume manifestation is considered an item for circulation purposes. As a result, it is difficult to distinguish between manifestations with multiple copies and multivolume manifestations. To avoid overestimating the level of duplication, multivolume manifestations, as identified from the bibliographic record, were not considered when estimating the duplication level.

For books published in the early 1900s, the average number of copies is about 2 and grew steadily until 1970 when it reached a high of 5.2 copies. However, the level of duplication dropped to 4.0 in the early 1970s and has remained relatively steady for the last thirty years. What caused the drop in the 1970s? There is no clear explanation, but it may be more than a coincidence that the drop coincided with the founding of OCLC. Most Ohio academic libraries joined OCLC in the early seventies. OCLC's Online Union Catalog (now WorldCat) made it possible to easily identify the libraries in Ohio that had acquired a particular book, and the availability of this information is likely to have influenced acquisition decisions.

There are several reasons for acquiring multiple copies of a manifestation, including (1) having a sufficient number of copies to supply local needs, (2) when the book is not available locally, to provide it promptly from the collective collection, or (3) the result of local approval plans. Being able to provide the book locally when requested depends on two conditions: having acquired the book and having it available on the shelf. Acquiring multiple copies of high-demand books is often necessary to ensure that at least one copy is available when requested. In other cases, the duplication was not planned; instead, it resulted from the use of similar approval plan profiles.

Even the strongest local collections will not be able to immediately supply every book requested and will need to rely on the collective collection to fill some of their requests. Although OhioLINK is exceptionally fast at providing the material from off-campus, it still takes a couple of days. The collective collection is subject to the same limitations as local collections; the book had to have been acquired by at least one of
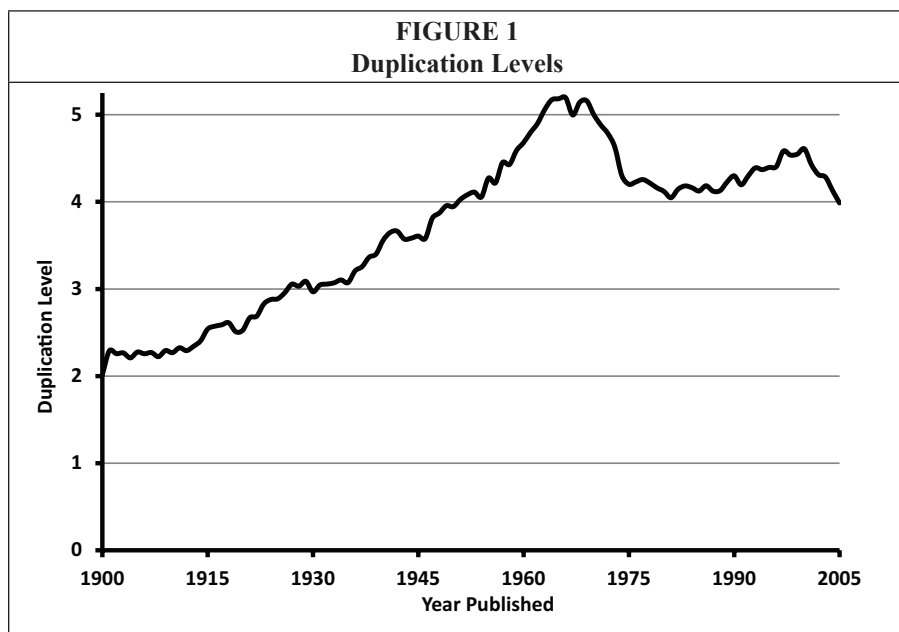

FIGURE 1
Duplication Levels

TABLE 4
Subject Collections

Subject	Number of Items	Percent Non-circulating	Number of Manifestations	Duplication Level	Annual Circulation	Circulation Rate	Average Age
Genre & Unclassed	2,943,640	32.7%	1,064,109	2.46	80,208	0.040	40
Arts and Recreation	2,033,325	11.0%	475,965	4.06	342,160	0.189	34
Business & Economics	2,075,157	7.2%	440,007	4.60	150,392	0.078	30
History & Geography	3,527,028	10.3%	757,818	4.26	344,679	0.109	44
Language and Literature	5,362,041	12.7%	1,429,058	3.48	520,986	0.111	44
Science and Technology	3,859,076	10.4%	879,913	4.11	385,197	0.111	32
Social Science	5,656,431	8.0%	1,131,597	4.79	751,880	0.145	35
Medicine	1,069,280	9.9%	222,595	4.66	174,651	0.181	24
Law	1,949,723	25.3%	259,188	6.37	71,882	0.049	25

the member libraries and at least one copy must be available to circulate.

It is possible to estimate the duplication level necessary to achieve a given level of availability of books in the collective collection using queuing theory. Even with generous assumptions about circulation periods, with the observed demand pattern, the level of duplication is far in excess of that necessary to achieve a high level of availability. Therefore, it can be assumed that the primary justification for the duplication is either to satisfy local needs or as the result of using similar approval plans. However, even a modest reduction in the duplication would have major benefits. Reducing the duplication level from 4.0 to 3.5 copies could allow for a 50 percent increase in the number of unique manifestations acquired annually or permit acquiring more digital resources.

Subjects

The analysis of OhioLINK subject collections is divided into nine broad subject groups: (1) Genre & Unclassed, (2) Arts & Recreation, (3) Business & Economics, (4) History & Geography, (5) Language & Literature, (6) Science & Technology, (7) Social Science, (8) Medicine, and (9) Law, based on their Library of Congress Classification (LCC). Materials lacking a valid LCC were included with genre in the Genre & Unclassed group. These broad subject groups were derived from those used in the North American Title Count.[16] The genre includes books that are categorized primarily by what they are rather than by their subject. This subject group includes encyclopedias, indexes, almanacs, bibliographies, and other similar materials. The Genre & Unclassed group comprises about 13 percent of the collection.

The characteristics of the subject collections are shown in table 4. The

collective collection for Language & Literature, with 1,429,058 manifestations, was the largest collection in terms of the number of manifestations, but the Social Science collection was larger in terms of the number of items. Medicine and Law were the smallest collections, as not all the institutions have these professional schools. Arts & Recreation had the highest annual circulation rate, at 0.189 circulations per item, while Genre & Unclassed and Law had the lowest, at 0.040 and 0.049 respectively. In terms of collection age, History & Geography and Language & Literature were the oldest, with an average age of 44 years each. Medicine was the youngest, with an average age of 24 years.

Languages
The data shed light on the use of non–English-language books compared to the use of those in English. Table 5 shows the general characteristics of the books by language. Not surprising, English-language books dominated the collections. English-language books also had the highest circulation rate, nearly double that of Spanish, the second most used language collection. Russian-language books were the least used. The relative circulation rate compares the circulation rates of the books in the major languages to that of English. Finally, the average age of the non–English-language books was much older than that of English-language books.

The proportion of non-English books in the collection by their year of publication is shown in figure 2. The proportion of non-English books peaked at almost 18 percent of the collection after the First World War but has declined since then. Over the last hundred years, the non-English proportion of the collections declined from approximately 15 percent to about 5 percent. The impact of the two World Wars is particularly notable in the figure. The use of non-English also saw a significant decline during the
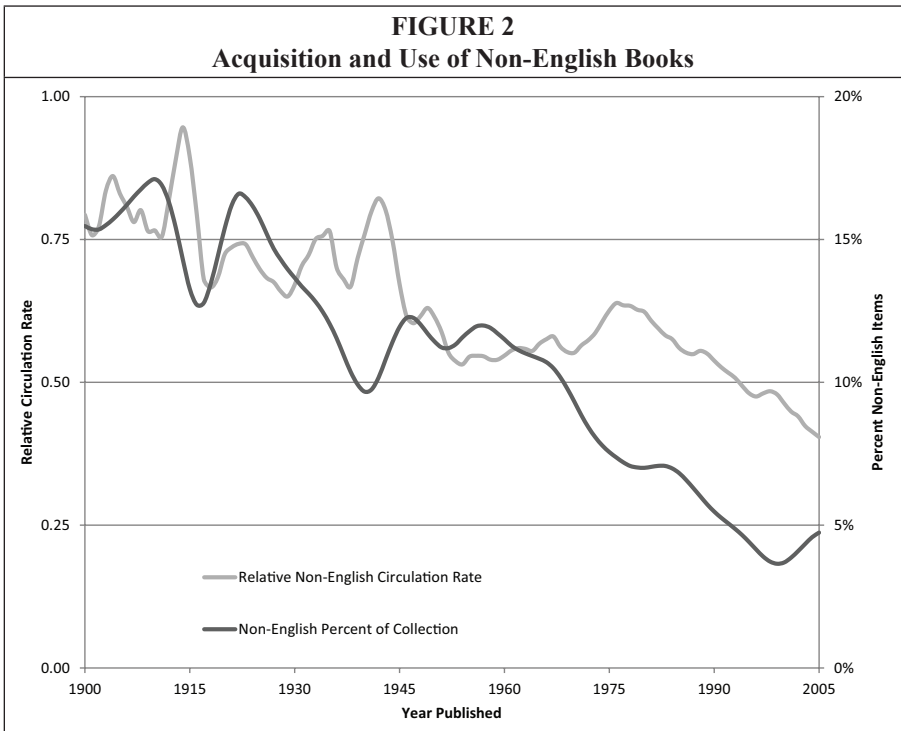

FIGURE 2
Acquisition and Use of Non-English Books

| | | | TABLE 5 | | | | | |
| | | | Language Collections | | | | | |
Language	Number of Items	Percent Non-circulating	Number of Manifestations	Duplication Level	Annual Circulation	Circulation Rate	Relative Circulation Rate	Average Age
English	25,873,893	13.7%	5,237,297	4.64	2,731,106	0.122	1.000	35
German	664,523	12.7%	318,106	1.67	19,063	0.033	0.269	56
French	543,256	10.6%	245,694	1.76	19,177	0.039	0.323	63
Spanish	335,927	6.2%	164,594	1.77	19,974	0.063	0.518	41
Russian	198,579	9.5%	123,764	1.35	3,906	0.022	0.178	39
Chinese	109,004	3.3%	68,050	1.19	5,731	0.054	0.445	27
Italian	114,215	11.7%	72,933	1.32	5,234	0.052	0.425	57
Japanese	84,295	16.6%	48,518	1.29	3,613	0.051	0.420	31
Arabic	30,174	2.9%	25,076	1.03	1,012	0.035	0.282	26
Greek, Modern	41,275	11.5%	30,545	1.14	1,238	0.034	0.277	46
Hebrew	53,929	14.0%	41,280	1.14	1,223	0.026	0.216	48
Indonesian	77,384	0.8%	66,332	1.08	1,040	0.014	0.111	25
Latin	100,325	43.7%	44,387	1.57	2,808	0.050	0.407	159
Polish	39,172	5.4%	31,672	1.11	422	0.011	0.093	39
Portuguese	30,030	5.9%	21,384	1.20	678	0.024	0.196	44
All Others	179,720	11.7%	120,618	1.27	5,810	0.037	0.299	46

same period. Non–English-language books published in the early 1900s circulated only slightly less than English; but, for recent publications, the circulation rate for non-English books dropped to about 40 percent of that for English. The combination of decreasing acquisitions and declining use of what was acquired has significant implications for language and area studies collections.

Scattering

Collection usage was widely distributed across manifestations; a few of these were heavily used, but most only circulated a few times if at all. The annual circulation for OhioLINK manifestations is shown in figure 3. The 100 most used manifestations each circulated over 170 times during the year observed. These heavily used manifestations generally were widely held, often with multiple copies, so that the annual circulation per item would be far lower. For example, Robert McCloskey's popular children's story *Make Way for Ducklings* circulated 127 times, making it the 190th most used manifestation. However, since there were 36 copies available in various OhioLINK libraries, the average annual per item circulation was only 3.53. This pattern of a few heavily used books and many little-used books was identified by Anderson as the *Long Tail*.[17] The long tail described by Anderson is another example of the use of power laws to explain the use of the library resources. Other power laws, such as Bradford's law, have also been widely used to describe the use of library resources.[18]

Trueswell's 80/20 rule—that 80 percent of the usage is generated by 20 percent of the collection—is widely accepted.[19] However, with the OhioLINK data, the pattern observed varied significantly from that predicted by the 80/20 rule. Usage was far less scattered with 7.2 percent of manifestations accounting for 80 percent of the circulation. Of the 5,899,520 manifestations that were permitted to circulate, only 1,041,405 (17.7%) circulated; less than 20 percent of the manifestations accounted for all of the circulation. This observation contrasts with Ettelt's findings that the 80/20 rule did accurately predict book circulation patterns at Columbia-Greene Community College. He reported that "19% [of the books] accounted for 79% of the circulation" and that 75% of the books did not circulate.[20] One possible explanation for the contrasting findings is that community colleges have newer, smaller, and more active collections than do academic libraries in general.

FIGURE 3
Scattering of Usage
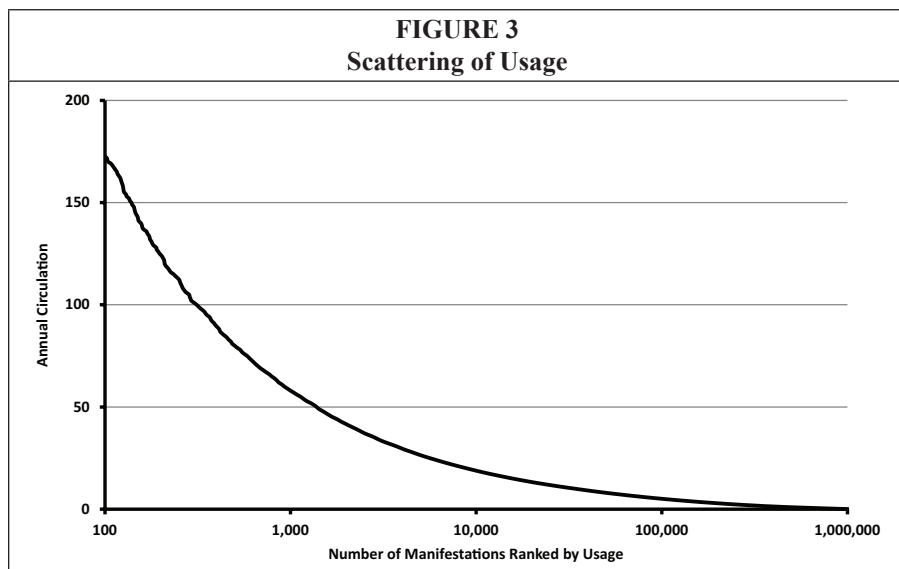
To better understand the usage patterns, the collections were grouped into four categories based on the observed usage:

- Noncirculating: Manifestations that are not allowed to circulate. A manifestation was considered noncirculating only if all of its items were noncirculating.
- Unused: Manifestations that are allowed to circulate but did not circulate during the year observed. For manifestations with multiple items, none of the items were used.
- Single Use: Manifestations that circulated once during the year observed.
- Multiuse: Manifestations that circulated multiple times during the year observed.

The relative size of these groups for the major subjects is shown in table 6. Overall, 11.4 percent of the manifestations were noncirculating. Most disciplines had a similar proportion of noncirculating books. The Genre & Unclassed category, which included many reference materials, had the largest proportion of noncirculating books, while Medicine had the smallest.

Overall, 72.9 percent of the manifestations did not circulate, and 8.5 percent circulated only once. The individual subject collections exhibited similar usage patterns, with close to 80 percent of the books either not circulating or circulating only once. The Arts & Recreation, Social Science, and Medicine collections were the most active, with over 20 percent of the books circulating at least once. In spite of the widely held belief that science and technology relies primarily on journals, the Science & Technology collection was more active than Language & Literature. Arts & Recreation and Medicine were the most active subject collections, with 24.2 percent of their collection circulating.

The 7.2 percent of manifestations that circulated more than once accounted for more than 80 percent of the total circulation. It should be emphasized that, whether a manifestation circulates or not in a given year is somewhat random. Except for the manifestations that are not permitted to circulate, every manifestation has a non-zero probability of circulating. For many of the books in the long tail, that probably will be very small—but not zero. Collectively, even manifestations with expected circulation rates as low as 0.001 (one circulation expected every 1,000 years) will be used periodi-

TABLE 6 Scattering of Usage					
		Manifestations			
Subject	**Number of Manifestations**	**Noncirculating**	**Unused**	**Single Use**	**Multi-use**
Genre & Unclassed	1,064,109	16.6%	78.8%	3.4%	1.2%
Arts & Recreation	475,965	10.1%	65.7%	12.0%	12.2%
Business & Economics	440,007	9.3%	76.4%	8.3%	6.1%
History & Geography	757,818	10.7%	71.8%	9.5%	8.0%
Language & Literature	1,429,058	11.6%	74.3%	7.8%	6.2%
Science & Technology	879,913	9.2%	73.8%	9.5%	7.4%
Social Science	1,131,597	10.2%	67.9%	10.7%	11.3%
Medicine	222,595	8.0%	67.8%	11.7%	12.5%
Law	259,188	13.2%	75.6%	7.1%	4.1%
All Subjects	6,660,250	11.4%	72.9%	8.5%	7.2%

cally. From a million such manifestations, a thousand would be expected to circulate in any given year. While the number of manifestations that do not circulate, circulate once, or circulate multiple times will remain fairly constant from year to year, the individual manifestations will move between the three groups in a somewhat random process.

Obsolescence

Obsolescence, the decrease in demand over time, is widely assumed to be a major factor affecting circulation and has been extensively studied. Buckland provides an extensive background on the impact of obsolescence on the use of library collections.[21] He identifies the two ways to viewing obsolescence:

  a. *In diachronous view, one is concerned with the use of a given document in successive years—'through time'.*
  b. *In a synchronous view, one is concerned with distribution of use made during a given span of time of documents of different ages.*[22]

The OhioLINK data is synchronous—the data cover the circulation occurring in a single year for books published in different years. The information in the WorldCat bibliographic record was used to determine the publication date. Since the corresponding circulation information is also available, the obsolescence rates can be computed.

The OhioLINK consortium is a nearly ideal source for synchronous data to examine the influence of age on book usage. There is detailed information in the bibliographic record to determine the age, the subject, and other relevant characteristics. A common problem with synchronous studies is estimating the size of the population—how many resources were available for use? The circulation of books published ten years ago may be half of that of recently published books. However, that does not necessarily imply that ten-year-old books are obsolete or even that their use has decreased over time. Meaningful usage rates must be based on per-item use—if the library had only half as many ten-year-old books, even though the total circulation was less, the per-item usage would be unchanged. To accurately estimate the circulation rate, it is necessary to know both the circulation and the number of items available for circulation.

The majority of the research on the obsolescence of library resources has focused on the journal literature and has used citations as the measure of use. Brookes proposed that obsolescence was a continuous exponential decay process.[23] That is, usage each year declined by a constant proportion following the negative exponential function. However, more recent studies have concluded that obsolescence is more complex. Alvarez, Escalona, & Pulgarin found that there is "no way to find an aging factor that is time-independent…."[24] Egghe and Egghe and Ravichandra Rao also report similar findings.[25] The observed obsolescence rate for OhioLINK books also did not follow negative exponential behavior. The overall obsolescence rate for these resources is shown in figure 4 along with the obsolescence rates for major subjects. The circulation of new books declines at about 20 percent per year; that is, the circulation for a recent book can be expected to be 20 percent less than what it was the previous year. However, the obsolescence rate slows as the books ages. After ten years, the obsolescence rate decreases to 6 percent. The usage rate for books published prior to the Second World War (70+ years old) did not show any significant further decline with age; the observed circulation rate for the prewar material was relatively stable at approximately 0.035 (35 uses per 1,000 items). Although low compared to new material, the use of the older material still is significant. Collectively, books over 50 years old generated as much circulation as did the material published in the most recent year.

Science & Technology, Medicine, and Business & Economics generally exhibited the fastest obsolescence and eventually stabilized at a relatively low circulation rate. Arts & Recreation exhibited the slowest obsolescence rate, followed closely by Literature

## FIGURE 4
Obsolescence Rates
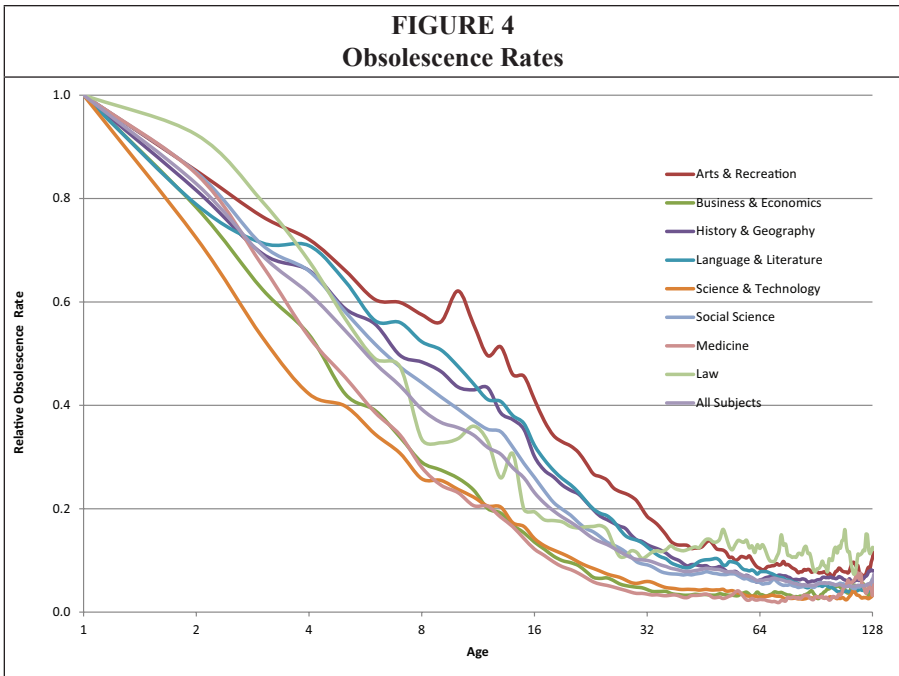
and History & Geography. The subjects with lower obsolescence also stabilized at a higher circulation rate than did those with the higher obsolescence rates. Law actually showed a small increase in the usage of its older books.

While age is a good predictor of use, it is not the best predictor—date of last use is a much better predictor of future use. The probability of an item circulating as a function of how long it has been since it last circulated is shown in figure 5. The probability of an item circulating if it has been idle for ten years or more stabilizes at approximately 0.0075 and does not change significantly for items unused for longer periods. Excluding the resources already in the depositories, there are over 7 million items that have not circulated for ten years or more. Although the usage rate is very low, it is not insignificant; collectively these 7 million items can be expected to generate over 50,000 circulations during the next year.

Limitations and Recommendations

From strictly a statistical perspective, the results presented here apply only to printed books and manuscripts held in participating OhioLINK libraries. However, considering the number, size, and diversity of the OhioLINK libraries, many of the findings are expected to be very similar to those of American academic libraries generally and, to a lesser extent, to those of academic libraries in English-speaking countries. Because of the strong influence of language on usage, the results are unlikely to apply to libraries serving primarily non–English-speaking users. Since OhioLINK is dominated by academic libraries, these results are unlikely to apply to other types of libraries.

A very relevant question is to what extent these findings apply to eBooks. Although print books will continue to play an important role in academic libraries, the importance of eBooks is rapidly increasing; and a better understanding of their usage patterns would assist in the effective development of eBook collections. At this point in the evolution of eBooks, the material available in digital form is very similar to

FIGURE 5
Last Use as Predictor of Future Use
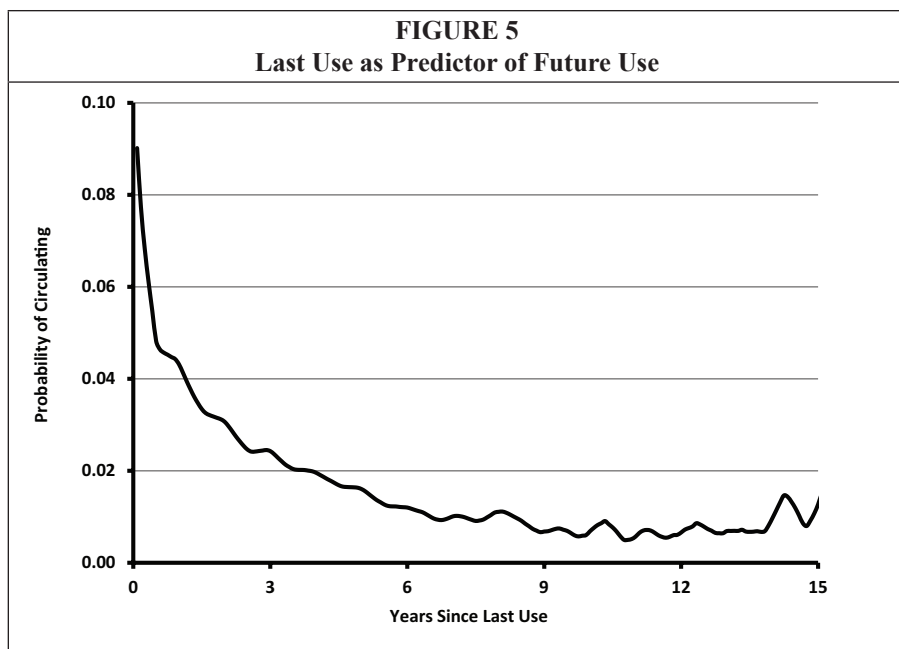
material that either is or has been published in print. Therefore, the authors expect that eBooks, at least in the short term, will exhibit similar usage patterns to print. If equivalent measures of usage were available for eBooks, the authors believe similar usage patterns would occur particularly as they relate to subject and language usage as well as scattering and obsolescence. However, until consistent and reliable usage data is available for eBooks, extrapolating from print usage may provide the best estimates of eBook use available.

This study compiled a very large and diverse set of circulation records. It includes circulation records for 28,475,701 different items from more than 100 different academic libraries. This paper only addresses general issues and key findings. However, this rich set of circulation data can support additional research, and further analysis of the data is strongly encouraged. OhioLINK and OCLC have made both the circulation and the bibliographic records publicly available under the Open Data Commons Attribution license for study and research.[26]

Conclusions

The OhioLINK libraries collectively have a very comprehensive and diverse collection. While the ARLs hold the largest collections, the combined collections of the other library types were similar in size and contained an equivalent number of unique resources. Except for the depositories, which hold the lesser-used items, the circulation rates were similar across the different types of libraries, with an average annual circulation rate of 0.11 circulations per item. The sciences showed stronger than expected usage, with both Science & Technology and Medicine having average or above-average circulation rates.

The level of duplication increased steadily from about 2 items per manifestation published in 1900 and peaked at over 5 in the late 1960s. Duplication began to drop for manifestations published in the early 1970s, and the current duplication level is now just over 4 items per manifestation. Much of the duplication appears to be the result of local need, since the level of duplication exceeds that required to ensure high

consortial availability. Reducing duplication will free resources for the acquisition of unique material or eBooks. Adding additional eBooks that can be shared consortially would be an alternative to duplicating resources, particularly for less-used books.

There has been a pronounced decrease in the use of non–English-language books. The overall circulation rate for non-English items is less than a third of that of English language items and has been dropping. The proportion of non-English items has also declined steadily. Non–English-language material published in the early 1900s comprise over 15 percent of the collection but comprise less than 5 percent of recently published material. The OhioLINK libraries are buying fewer non-English books, and those that are acquired are little used.

Most usage was generated by a relatively small proportion of books: 7.2 percent of manifestations produced 80 percent of circulation. Over 80 percent of the manifestations did not circulate in the year observed. The scattering pattern showed that a few books were heavily used but that a huge number were rarely, if ever, used. Many of the rarely used books were old; all subject areas exhibited a pronounced decline in use over time, with the sciences showing the most rapid decline. However, the obsolescence rate slowed as the books aged and appears to stop after about fifty years. While these older books have a much lower circulation rate than new books, they continue to be used and are an essential part of research collections.

The combination of scattering and obsolescence has resulted in a collection where less than 10 percent of items circulated more than once during the year observed. Collectively, these little-used books are an important and valuable resource that must be retained at least until digital copies are available. Many of these books are ideal candidates for the depositories where they can be kept in a controlled environment with greatly reduced storage costs. The books currently in the depositories were well selected and have low circulation rates. However, a huge number of books remain on open shelves that have circulation rates as low or lower than those currently in the depositories. Previous use is the best predictor of future use and could be effectively used to identify additional books suitable for the depositories.

Notes

1. William J. Studer, *OhioLINK, a Concise History* (Columbus, Ohio: 2002), 1.

2. Ohio Board of Regents, Library Study Committee, *Academic Libraries in Ohio: Progress through Collaboration, Storage, and Technology* (Columbus, Ohio: Ohio Board of Regents, 1987).

3. OhioLINK Collection Building Task Force, Julia Gammon, and Edward T. O'Neill, *Ohio-LINK–OCLC Collection and Circulation Analysis Project 2011* (Dublin, Ohio: OCLC Research, 2011), available online at www.oclc.org/research/publications/library/2011/2011-06r.html [accessed 29 March 2013].

4. Allen Kent et al., *Use of Library Materials: The University of Pittsburgh Study* (New York: Marcel Dekker, 1979).

5. Jasper G. Schad, "Missing the Brass Ring in the Iron City," *Journal of Academic Librarianship* 5 (1979): 60–70.

6. Joseph J. Branin, "Use of Library Material" (editorial), *College and Research Libraries* 70, no. 4 (July 2009): 311–12.

7. Richard L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43 (Jan. 1969): 458–61.

8. Robert N. Broadus, "Use Studies of Library Collections," *Library Resources & Technical Services* 24, no. 4 (Fall 1980): 317–24.

9. Mary Kane Trochim, *Measuring the Book Circulation Use of a Small Academic Library Collection* (Washington, D.C.: Association of Research Libraries, Office of Management Studies, 1985), available online at http://babel.hathitrust.org/cgi/pt?id=mdp.39015014328689;q1=trochim%2C%20mary%20kane#view=1up;seq=17 [accessed 12 July 2013].

10. Sheila Cheung and Terry Chung, "Monograph Circulation over a 15-Year Period in a Liberal Arts University," *Library Management* 32 (2011): 419–34.

11. Lynn Wiley et al., "A Domestic Monograph Collection Assessment in Illinois Academic Libraries: What Are We Buying and How Is It Used," *Interlending and Document Supply* 39 (2011): 167–75.

12. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, Sept. 1997, as amended and corrected through Feb. 2009 (International Federation of Library Associations and Institutions, 2009): 17–24, available online at www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf [accessed 8 August 2013].

13. OhioLINK Collection Building Task Force, Julia Gammon, and Edward T. O'Neill, *Ohio-LINK–OCLC Collection and Circulation Analysis Project 2011* (Dublin, Ohio: OCLC Research, 2011), available online at www.oclc.org/research/publications/library/2011/2011-06r.html [accessed 29 March 2013].

14. Edward T. O'Neill and Julia A. Gammon, "Building Collections Cooperatively: Analysis of Collection Use in the OhioLINK Library Consortium," ACRL 14th National Conference, Seattle, Washington, March 12–15, 2009.

15. Thomas B. Hickey and Jenny Toves, *FRBR Work Set Algorithm* (Dublin, Ohio: OCLC Research, 2005), available online at www.oclc.org/research/activities/frbralgorithm.html [accessed 29 March 2013].

16. Association for Library Collections & Technical Services, and University of Illinois, *North American Title Count 1997: Titles Classified by Library of Congress and National Library of Medicine classifications* (Chicago: Association for Library Collections & Technical Services, 1998).

17. Chris Anderson, *The Long Tail: Why the Future of Business Is Selling Less of More* (New York: Hyperion, 2006).

18. S.C. Bradford, "On the Scattering of Papers on Scientific Subjects in the Scientific Periodicals," *Engineering* 137 (1934): 85–86.

19. Richard L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43, no. 5 (1969): 58–61.

20. Harold Ettelt, *Does the 80/20 Rule Apply to Books?* (ERIC Report No: ED 298 963, Washington, D.C: U.S. Department of Education, Office of Educational Research and Improvement, 1988): 4.

21. Michael K. Buckland, *Book Availability and the Library User* (New York: Pergamon Press, 1975).

22. Ibid., 12–13.

23. B.C. Brookes, "The Growth, Utility, and Obsolescence of Scientific Periodical Literature," *Journal of Documentation* 26 (1970): 46–54.

24. Pedro Alvarez, Isabel Escalona, and Antonio Pulgarin, "What Is Wrong with Obsolescence," *Journal of the American Society for Information Science* 51, no. 9 (2000): 812–15.

25. L. Egghe, "On the Influence of Growth on Obsolescence," *Scientometric*s 27, no. 2 (June 1993): 195–214; L. Egghe and I.K.R. Rao, "Citation Age Data and the Obsolescence Function: Fits and Explanations," *Information Processing & Management* 28, no. 2 (1992): 201–17.

26. Egghe and Rao, "Citation Age Data and the Obsolescence Function."