

Building Collections Cooperatively: Analysis of Collection Use in the OhioLINK Library Consortium

Edward T. O'Neill and Julia A. Gammon

Introduction

Librarians have long been collaborators. The history of cooperative collection development among libraries goes back hundreds of years.

The foundation for cooperative collection development was laid in the early 1800s, with international exchange program begun by European universities. A century later conventions for resource sharing such as union catalogs and interlibrary loan, were established. The mid-20th century saw the introduction of a number of cooperative collection development programs: the Farmington Plan; the Foreign Newspaper Project of the Association for Research Libraries; and the Midwest Inter-Library Center, later known as the Center for Research Libraries.¹

Economic, social and technological changes during the 20th century saw the growth in consortia and enabled libraries to work collaboratively in building collections. In today's resource sharing environment, it doesn't matter who owns the materials, but that the

right materials are available to meet the needs of users. Determining what the *right* materials are is the key. Collection analysis can provide that information; changing selector's behavior has to follow.

Background on OhioLINK

OhioLINK is a major player among library consortia. Resource sharing is a primary goal throughout the state and was further enhanced by an infrastructure including a strong delivery system, a union catalog, a shared online system and an established history of cooperation. Today, OhioLINK includes most academic libraries in the state.

The easiest way to describe the consortium is to reference the information in its website.

"The Ohio Library and Information Network, OhioLINK, is a consortium of 89 Ohio college and university libraries, and the State Library of Ohio, that work together to provide Ohio students, faculty and researchers with the information they need for teaching and research. Serving more than 600,000 students, faculty, and staff at 90 institutions,

Edward T. O'Neill is Senior Research Scientist, OCLC Research; Julia A. Gammon is Head, Acquisitions Department at the University Libraries, University of Akron

OhioLINK's membership includes 16 public/research universities, 23 community / technical colleges, 50 private colleges and the State Library of Ohio.

Together, OhioLINK and its member libraries provide access to:

- 47.6 million books and other library materials
- Millions of electronic articles
- 12,000 electronic journals
- 140 electronic research databases
- 40,000 e-books
- Thousands of images, videos and sounds
- 17,500 theses and dissertations from Ohio students”²

Ohio has a long history of library cooperation going back to the 1967 founding of Ohio College Library Center, which eventually became the OCLC Online Computer Library Center. OCLC's cataloging products and services found a national audience and it now provides information services globally.³

OhioLINK may be the only academic library consortium created because of a space problem.⁴ 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 the collections would continue to grow, the Ohio General Assembly directed the Ohio Board of Regents to find an alternative solution 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 proceeded to recommend a transformation of Ohio's academic libraries to coincide with the electronic information age.

In 1987, the Library Study Committee made three key recommendations⁵:

- Creation of an offsite high density book depository system,
- Formation of a statewide electronic catalog system with a delivery component,
- Appointment of a steering committee.

In response to these recommendations, the Steering Committee was formed and prepared a Planning Paper (1988); a Request for Information (1989); and a Request for Proposal (1989) to initiate this system—the blueprint for OhioLINK was created.

Working as partners, OhioLINK and Innovative Interfaces Inc. developed the current online system used by the consortium. By 1994, a system (Innovative Interfaces' Inn-Reach product) was allowing patron initiated borrowing to function between the OhioLINK libraries. Users are able to use unmediated online borrowing and pick up books at the location they select. Deliveries are usually 2-3 days with loan periods of three weeks with the potential for up to four renewals. In 2007, Ohio libraries filled 751,200 borrowing requests from other libraries—an increase of 1128% since borrowing began in 1994.

Building Collections: Buying Books as a Consortium

With the technical infrastructure in place, the practical planning work of resource sharing of monographs could begin. As in many library organizations, OhioLINK is built on a voluntary committee structure. Headed by an executive director, with a Library Advisory Council (LAC) consisting of directors of member libraries, much of the consortia's work is done by four committees consisting of representatives of member libraries. One of the committees is the Cooperative Information Resources Management Committee (CIRM) which deals primarily with resource management and increasing access to information resources for OhioLINK members.

To better control the *unnecessary* duplication of titles within the state, under CIRM's charge, the OhioLINK Approval Plan Task Force was formed to select a common book vendor for the state and a consortial approval plan. Subsequently, YBP Library Services was selected as the vendor. OhioLINK libraries that used the common vendor now had the ability to know, at point of purchase, how many copies of a title were expected to be purchased by other libraries and what titles were expected to be shipped on approval plans. This groundbreaking project is documented in an article by Diedrichs.⁶ Once the statewide vendor was selected, and coordinated firm ordering and approval plan profiling began, the infrastructure for making informed collection decisions was available via YBP's GOBI online product available to all the consortial libraries. The Approval Plan Task Force's charge was expanded, and it was subsequently renamed the Collection Building Task Force (CBTF). Its new charge was "...to expand the amount spent on cooperative purchases through identifying specific information

resources of value to the OhioLINK community and encouraging the reduction of duplication in current purchases, thereby freeing funds to be reallocated to increase the depth and breadth of the collection; to investigate and foster specific methods to achieve the former; and to foster increased involvement in and coordination of local collection development activities with those of other OhioLINK member libraries.”⁷

Selling cooperative collection management is not easy. Various methods were discussed, planned and implemented by the CBTF to encourage cooperative collection development activities throughout the state.⁸ Some examples included: fostering collaboration among the 28 subject groups of collection specialists; presenting ‘road shows’ in different geographic locations to market cooperative ideas and to promote new vendor services; showcasing different library workflows for improved efficiency; and promoting the value of a diversified consortial collection. One example in reducing unnecessary duplication and increase diversity in the OhioLINK collection was the “Not Bought in Ohio” feature developed in collaboration with YBP Library Services’ GOBI database. This feature helped identify titles not purchased within the state and was a method to diversify the collection as a whole if used by selectors.

What Was Missing

One of the problems in “selling” or “marketing” cooperative collection development is that the CBTF lacked data to demonstrate to selectors how many copies were actually needed within the consortium to meet OhioLINK’s circulation needs. Where did we have too many copies? Where did we need more copies? Some duplication is certainly needed with 600,000 potential users, so the CBTF was careful to refer to their charge as reducing *unnecessary* duplication.

To guide library selectors in making purchasing decisions, CBTF provided a “suggested” optimum number of consortium-wide copies, 3-8 circulating copies depending on the subject area, alternatively, if 25% of the total number of copies were “available” for borrowing (on the shelf), then additional copies are probably not needed. Of course, one has to evaluate one’s local needs first before applying these suggested purchasing number guidelines. Some individual libraries set their own purchasing limit based on these guidelines. For example, Ohio State University has a ceiling of 5 copies, and John Carroll University uses

8 copies. These guidelines are called “soft caps” within the consortium and can be overridden when the need becomes clear.

One of CIRM’s objectives was to “Develop models for analyzing collections and use in order to make informed decisions.”⁹ CBTF knew from practical observation that there were too many copies of some books, not enough copies of others, and that they were probably missing some titles all together. The CBTF thought the best way to approach collection analysis was to create a list of the questions we wanted answered in the analysis.¹⁰ We quickly learned that the most important questions we wanted answered on usage of materials could not be supplied by our central Innopac circulation system. OhioLINK looked at several off-the-shelf collection analysis products but, at that time, none of the products were scalable to the number of libraries and the number of circulation transactions required. In 2005, OhioLINK was approached by the OCLC Research about working on a study of circulation of books within the consortium. OCLC Research had recently done circulation analysis at the University of Colorado and wanted to apply the methodology in a consortial setting.¹¹ After some discussion, there was agreement that a joint project would be mutually beneficial.

Goal of the Study

The goal of this research is to better understand the usage patterns of books in academic libraries; what books are being used, what books aren’t being used, how many copies are needed, the ideal size of subject collections, etc. This study is limited to books and manuscripts since these materials typically circulate and circulation is a significant element in evaluating book collections. It is expected that by providing detailed usage data, selection behavior will change and resource dollars can be allocated more effectively. In discussions, parallels were drawn to the Pittsburgh Study of the 1970s,¹² one of the most comprehensive studies of collection usage patterns. However, that study covered only single institution’s usage—not a that of a large consortium.

Designing the Study

The CBTF identified the data needed, and worked closely with OCLC Research to plan the study. During the spring of 2007 and again in the spring of 2008, the library systems managers at each of the

OhioLINK libraries were asked to generate a file of circulation records for all their print books and manuscripts. The records included the fields shown in Table 1. Table 1 also shows the contents of a typical circulation record. Detailed instructions were written, rewritten and tested by member libraries to make sure the instructions were clear so the results generated provided the necessary data for the analysis.

Data Elements	Required/ Optional	Example Values
Source of Data	Required	University of Akron
Item Number	Required	i25878591
OCLC Number	Optional	45207959
Brief title	Required	The infinite / A.W. Moore
LCCN	Optional	00051722
Location code	Required	bc
Circ. status code	Required	-
Total circulations	Required	5
Total renewals	Required	1
Accession date	Required	8/3/2001
Date of last use	Required	8/3/2001
ISBN:	Optional	0415252857 (pbk.)

Marketing the Research Study

Months before the data was to be requested from the libraries, a plan was devised to promote the research study. Because the project was going to require much time, effort and goodwill among the consortium members, especially the systems librarians who had to do most of the work, the CBTF had to promote the value of the research project to the OhioLINK membership to ensure widespread participation. Because top-down buy-in was important, directors were informed about the study, and encouraged to have their libraries participate. The research study was discussed at various meetings and announced on the listservs. Publicity began many months before the data collection was due and reminders were sent periodically. As in most, if not all, OhioLINK endeavors, participation was purely voluntary, but to get statistically valid results, both groups involved in the study knew a high level of participation was necessary.

Collecting the Data

In 2007 and 2008, a one month window was set for data gathering and delivery. The instructions for collecting the data can were posted on the OhioLINK website¹³ and additional details on the planning phase were provided by Gilliland.¹⁴ Technical assistance was provided to individual libraries to ensure a good representation was achieved for the study. Each time, as the deadline approached, any institution whose files were not yet received were called or e-mailed from a committee member offering assistance and encouraging participation. All OhioLINK libraries participated during the first phase except for two community colleges; another community college dropped out during the second phase. All of the other OhioLINK libraries continue to participate in the study.

Once the records were received, they underwent an extensive process to match them to the corresponding bibliographic record in OCLC's WorldCat. As part of its duplicate detection effort, OCLC merged numerous records so many of the incoming OCLC numbers had to be replaced with their current OCLC number. When possible, records lacking OCLC numbers were matched to WorldCat records using either the LCCN or the ISBN. All matches were validated by comparing the brief title in the circulation record with the title in the matching WorldCat record. Records lacking any standard number and those records where the titles were significantly different were excluded from the study. After the second phase of the data collection, the records for the two years were combined into a single record. In addition to the information in the original records, the combined record includes the annual circulation for 2008 (spring of 2007 to the spring of 2008).

Data Analysis

Extensive analysis has been undertaken but, due to time and space limitations, we will focus here on the analysis of the collective collection—the combined collection of the OhioLINK libraries. This study and the analysis has been heavily influenced by FRBR¹⁵ and we will explicitly use the following group 1 FRBR terminology:

Work: A distinct intellectual or artistic creation. For this study, the working definition of works is that used by Hickey and Toves for the OCLC work set algorithm.¹⁶

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

Item: A single exemplar of a manifestation. Items are a particular copy of a manifestation. In this study, each circulation record is assumed to represent a single item.

The fourth FRBR group 1 entity, expression, is not used in this study due to the difficulty of identifying expressions algorithmically.

The initial step in the analysis was identification of the manifestation and works. Since each of the circulation records had been matched to a WorldCat bibliographic record, the OCLC number from the bibliographic record was used to identify the manifestation. The OCLC work set algorithm was then used to identify the corresponding work. The results are shown in Table 2.

Collectively, the OhioLINK libraries hold an extensive collection of non-English language books; a total of 2,383,462 items, almost nine percent of the collective collection. One of the questions was how

FRBR Resource Types	Frequency
Items	29,570,205
Manifestations	6,955,505
Works	5,686,173

much usage is made of these non-English resources. The average circulation rate for circulating books is 0.109 circulations per year. By contrast, the average circulation rate for the 14 most widely held languages is only 0.019—significantly below the average. The circulation rates for the 14 most widely held languages are shown in Figure 1. Spanish language materials are the most used but their circulation rate of 0.065 is still significantly below the average.

The analysis by subject is a key part of the study. All of the items were grouped into 24 broad subject areas based on the Library of Congress Classification in the corresponding WorldCat bibliographic record. A 25th category was used for books that were either unclassified or had an invalid classification. The collective collection size and the annual circulation for these subject areas are shown in Table 3. Not sur-

Figure 1. Circulation Rate for Non-English Language Books

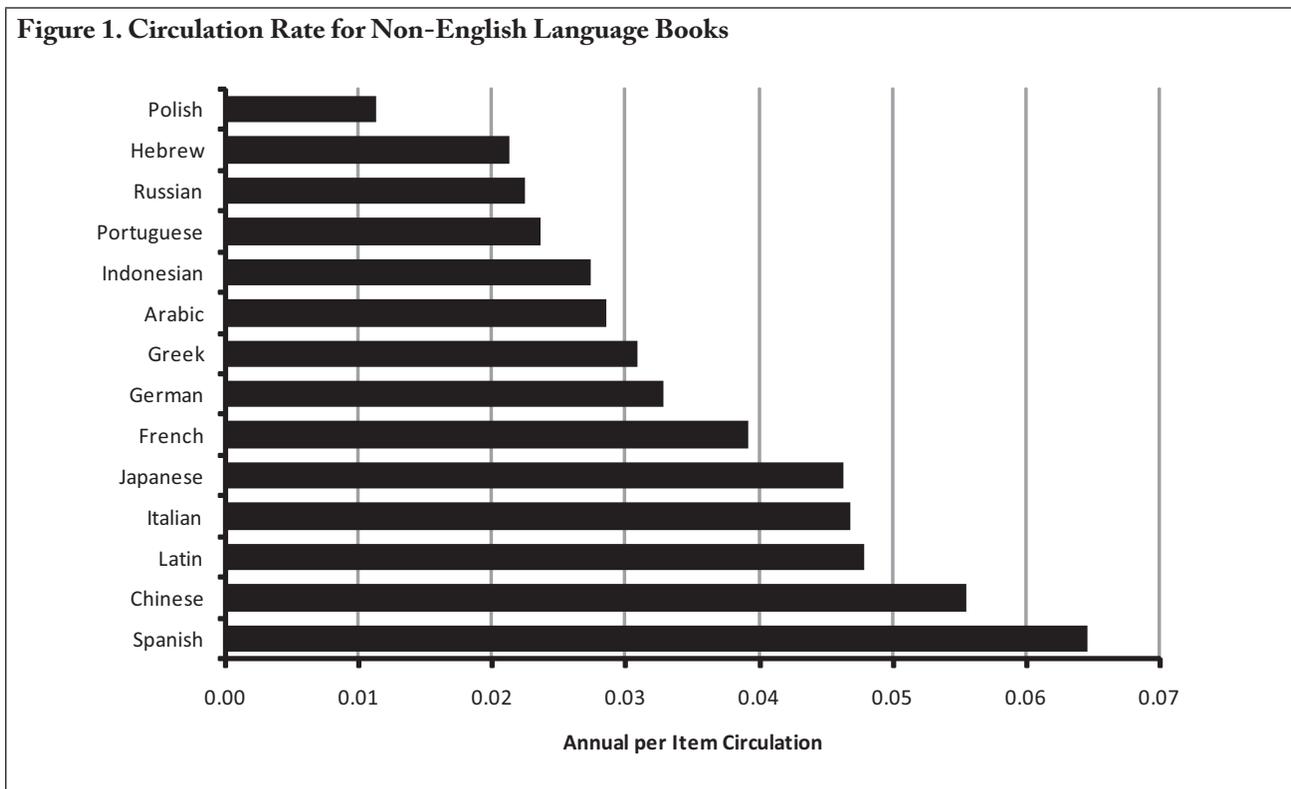


TABLE 3
Size and Circulation of Subject Collections

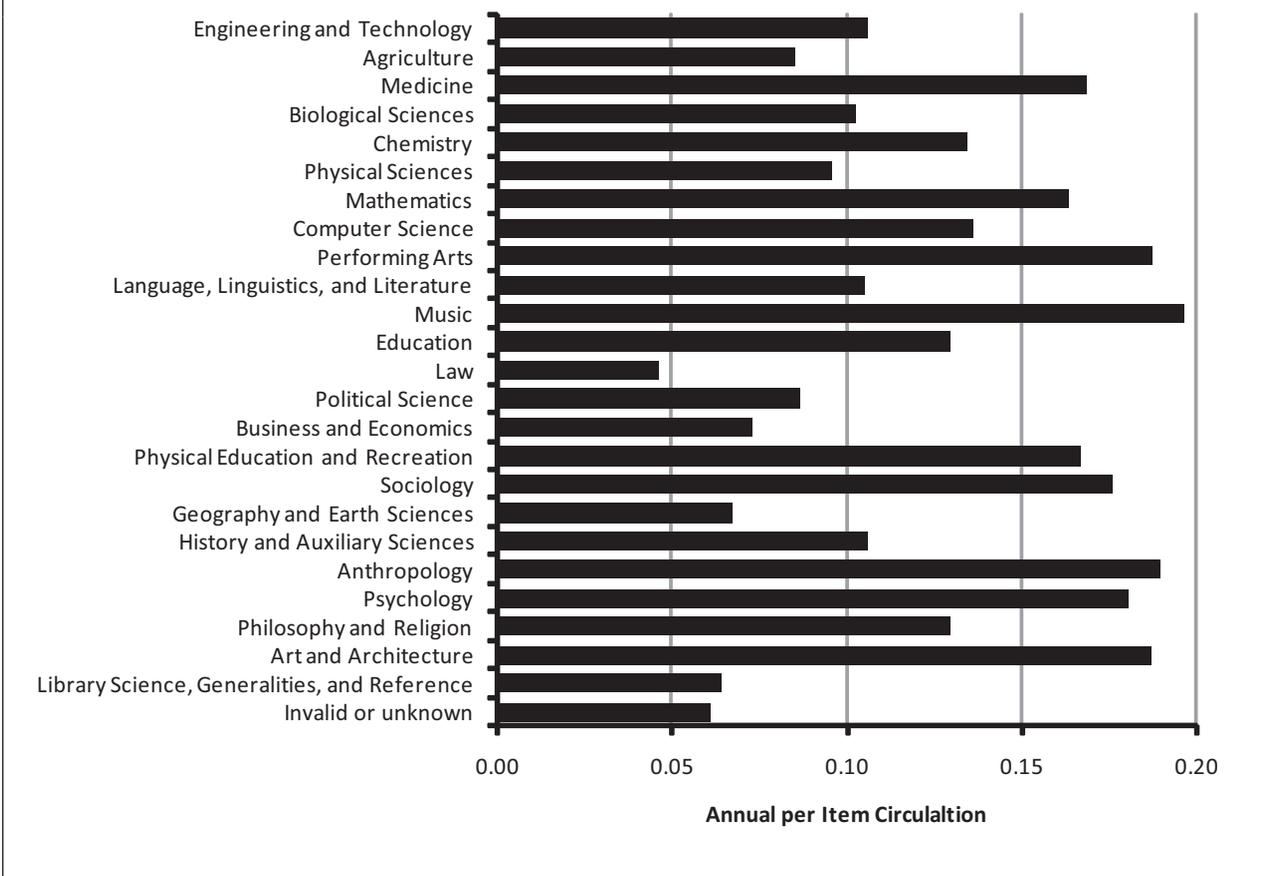
Subject	Collection Size (Items)	Collection Size (Manifestations)	Collection Size (Works)	Circulating Collection Size (Items)	Annual Circulation
Invalid or unknown	3,861,671	1,510,864	1,258,428	3,110,852	189,383
Library Science, Generalities, and Reference	673,777	127,889	114,198	574,904	36,815
Art and Architecture	1,088,032	269,095	231,165	1,006,539	188,132
Philosophy and Religion	1,927,591	454,981	356,611	1,785,808	231,790
Psychology	354,318	54,125	41,655	336,618	60,794
Anthropology	170,882	35,354	29,046	161,360	30,605
History and Auxiliary Sciences	3,369,936	701,134	568,703	3,191,611	339,060
Geography and Earth Sciences	464,741	97,626	84,713	436,898	29,508
Sociology	1,297,313	203,040	173,229	1,246,877	219,518
Physical Education and Recreation	204,204	52,102	44,781	192,921	32,217
Business and Economics	2,073,908	424,579	369,588	1,988,304	146,007
Political Science	851,292	151,541	128,475	811,780	70,683
Law	1,743,673	207,645	177,308	1,400,923	64,704
Education	990,440	166,856	148,308	958,805	124,787
Music	457,763	85,824	68,507	410,848	80,778
Language, Linguistics, and Literature	5,284,971	1,384,718	1,013,712	4,885,495	513,816
Performing Arts	258,892	46,932	39,238	241,345	45,277
Computer Science	222,486	51,553	44,651	216,490	29,492
Mathematics	377,970	66,464	53,026	368,555	60,352
Physical Sciences	483,603	87,498	73,203	453,032	43,364
Chemistry	203,644	30,875	25,222	183,010	24,668
Biological Sciences	570,530	106,878	89,414	538,728	55,169
Medicine	1,264,780	245,522	205,609	1,200,159	202,248
Agriculture	224,616	72,216	64,510	213,133	18,220
Engineering and Technology	1,149,172	320,194	282,873	1,075,672	114,511

prisingly, literature and history are the most widely held subjects, accounting for nearly thirty percent of items.

There were considerable differences in the circulation rates by subject as shown in Figure 2. Music books received the heaviest use while law books were the least used. As with the other circulation rates, the circulation rates are based only on the items that can circulate. Therefore, the fact that some disciplines make heavy use of reserves and other non-circulating material should not be a factor. The sciences had

higher circulation rates than expected considering the widely held belief that books are not as important in the sciences as in other disciplines. Medicine was heavily used and mathematics, computer science, and chemistry also had above average circulation rates.

Obsolescence, the decrease in demand over time, is widely assumed to be a major factor effecting circulation and has been extensively studied. Buckland¹⁷ provides an extensive background on the impact of obsolescence on the use of library collections. He

Figure 2. Circulation Rate by Subject

identifies the two ways of viewing obsolescence¹⁸:

(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.

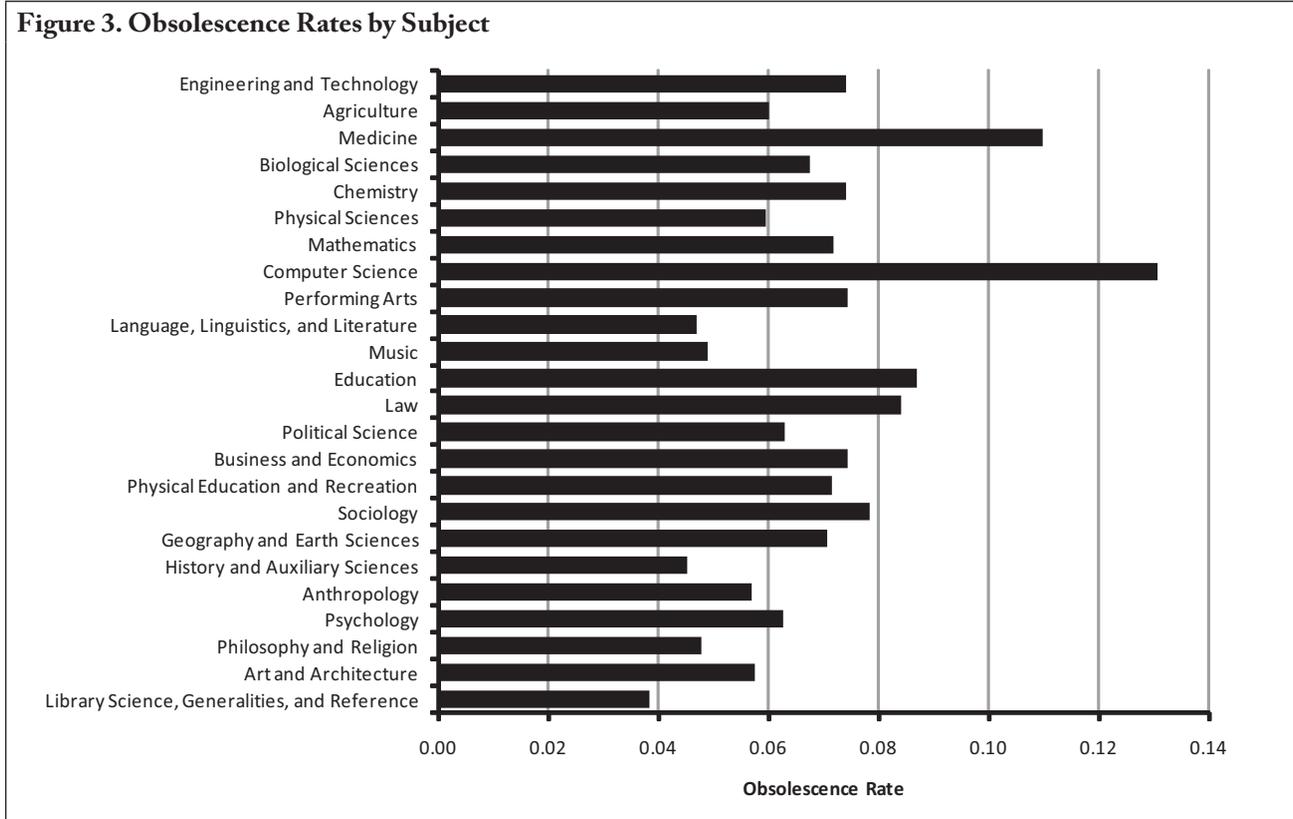
Buckland also discusses many of the problems related to estimating the obsolescence including changes in the size of both the collections and the user populations. Many of these issues have limited the practical application of obsolescence to monographic collections.

The OhioLINK data provides a nearly ideal synchronous view to use to examine the influence of age on book usage. There is detailed information in the bibliographic record and the age of the material can be determined from the publication date. The classification provides the necessary subject information permitting accurate estimates of the number of books in the collective collection in a given subject published

in any given year. Since the corresponding circulation information is also available, the obsolescence rates can be computed.

The obsolescence rate is defined as the rate at which the demand for an item or collection of items decreases over a given time period, usually a year. For example, if a particular book has a 10% obsolescence rate and was used 100 times when first acquired; it would be expected to be used 90 times in the second year, 81 times in the third year, 73 times in the fourth year, etc.

The obsolescence rates for the different subject areas are shown in Figure 3. Computer science, with an obsolescence rate of 13.1%, had the highest rate; Library science, generalities, and reference at 3.8% had the lowest. History also had a very low obsolescence rate of 4.5%. Books in subjects with low obsolescence rates such as history or literature should continue to be used for many years. By contrast, older books in computer science or medicine will experience only limited use. This certainly is not an unexpected result—a ten year old Word 97 book is of little value except as a



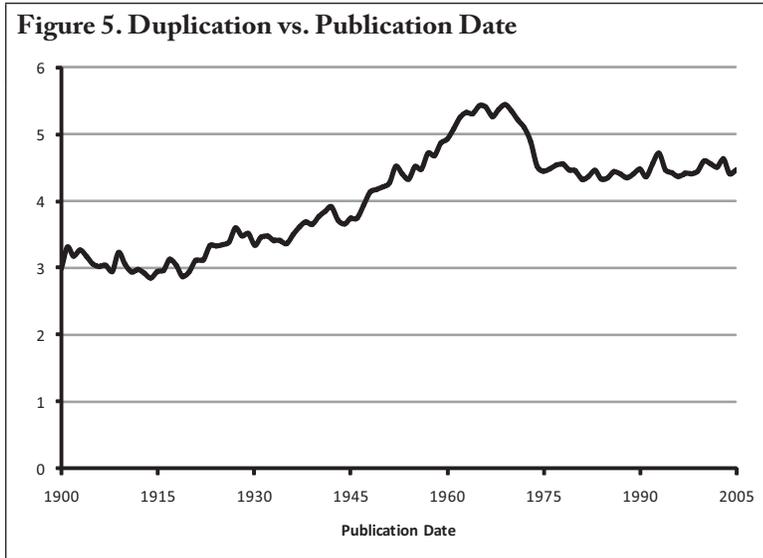
historical document while a John Steinbeck novel will continue to be read for many years.

The observed obsolence for history and computer science is shown in Figure 4. It is interesting to note that the average computer science book circulates

1.5 times in its first year, almost twice the average circulation as that for a new history book. However, after five years the circulation rates of the two subject areas have equalized and after twenty years, the circulation rate for the average history book is nearly triple that of the typical computer science book.



Duplication is an important and frequently discussed issue—how much duplication is necessary; how much is excessive. Every dollar spend on duplicating resources is a dollar less that is available to acquire unique materials. Our approach to the duplication question was to start by examining current duplication within OhioLINK. The first question is how much duplication exists and has it changed over time. Figure 5 shows the level of duplication, based on the date of publication, for the last hundred years. For books published in the early 1900s, the average number of copies is about 3 and grew steadily until 1970 when it reached a high of 5.5 copies. However, the level of duplication dropped in the early 1970 to 4.5 copies and has remained fairly steady at that level 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 it coincided with the founding of OCLC. OCLC's bibliographic database made it possible to identify the libraries in Ohio that had acquired a particular book and this information may have influenced acquisition decisions.

As shown in Figure 6, duplication varied from a low of 3.1 in agriculture to a high of 8.4 in law. Much of this variation can be easily explained. Ohio State University has the only comprehensive agricultural program in the state so one would not expect to find high duplication in this area. The high duplication in law is harder to explain. There are nine law schools in Ohio and the high duplication implies that the law schools have nearly identical collections. Chemistry, sociology, and psychology also had high duplication levels but these subjects are included in the curriculum at many institutions.

What Next? Using the Information

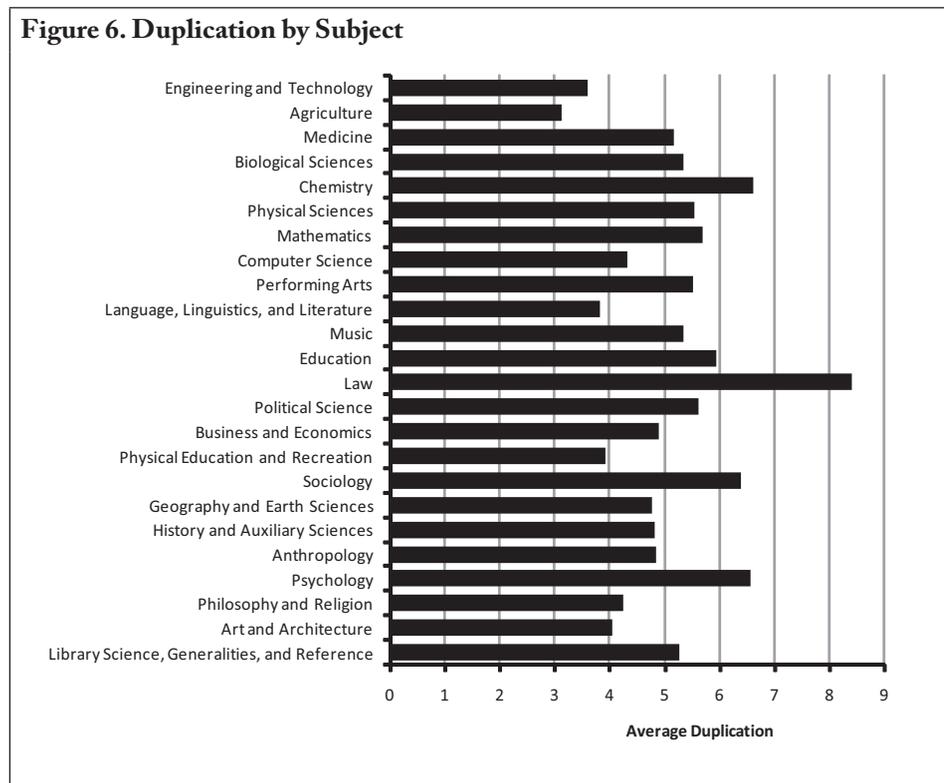
In addition to the analysis of the collective collection, the collection of the individual libraries is also being analyzed. A workshop will

be held in the spring of 2009 at OCLC to discuss the results and get feedback from the libraries. OhioLINK also plans to make the data available to OhioLINK members in a statistical package so that individual libraries will be able to run reports as needed for various projects.

Some of the potential uses of the data include:

- Changing the buying behavior in selectors
- Diversifying the collection
- Assisting in making purchasing decisions and approval plan profiling
- Providing guidelines on how many copies of a book in a particular subject area are needed based on circulation patterns
- Identifying items that can be discarded or moved to offsite storage
- Showing the strengths and deficiencies in the collective collection
- Showing the strengths and deficiencies of the individual collections
- Allowing money not used in duplicating materials to be used to diversity the collection or purchase other types of materials

Figure 6. Duplication by Subject



Conclusions

The OhioLINK circulation data is the largest and most diverse set of academic usage data for books ever collected. From a statistical perspective, the results are limited to the Ohio academic libraries. However, because of the number and diversity of the OhioLINK libraries, most of the findings, certainly the general trends, are expected to apply to most academic libraries.

The project is still ongoing—while the data collection phase is complete, the analysis phase has just begun. We have already answered some questions and improved our understanding of the books usage patterns in academic libraries but the results are raising new questions. We now know how much duplication exists but not how much is necessary. The study will continue to utilize the expertise of collection development librarians, systems librarians, and library administrators throughout Ohio to interpret the findings and identify additional analysis to ensure that the needs of the library users are met, our collections are efficiently managed, and our limited resources are effectively allocated.

Notes

1. Milton T. Wolf & Marjorie E. Bloss, "Introduction," *Collection Management* 24 (no. 1/2, 2000)
2. "This is OhioLINK," <http://www.ohiolink.edu/about/what-is-ol.html> (accessed December 1, 2008).
3. William J. Studer, *OhioLINK, a Concise History* (Columbus, Ohio; 2002), 1.
4. *Ibid.*
5. Ohio Board of Regents. Library Study Committee. *Academic Libraries in Ohio: Progress Through Collaboration, Storage, and Technology*, Columbus, Ohio, Ohio Board of Regents, 1987.
6. Carol Pitts Diedrichs, "Designing and Implementing a Consortial Approval Plan: The OhioLINK Experience," *Collection Management* 24 (no. 1/2, 2000) 15-44
7. <http://platinum.ohiolink.edu/cbtf/index.htm> (accessed December 9, 2008)
8. Julia A. Gammon & Michael Zeoli, "Practical Cooperative Collecting for Consortia: Books-Not-Bought In Ohio," *Collection Management*, 28, no. 1/2, 77-105
9. OhioLINK Policy Manual <https://www.ohiolink.edu/ostaff/policyhbk/cirm/1.assumptions.html> (accessed December 1, 2008)
10. Collection Building Task Force. "Collection Analysis Questions by Categories", <http://platinum.ohiolink.edu/cbtf/oclcres/questions.pdf>.
11. Jennifer D. Knievel, Heather Wicht & Lynn Silipigni Connaway, "Use of circulation statistics and interlibrary load data in collection management," *College & Research Libraries* 67 (no. 1, 2006) 35-49.
12. Allen Kent, *et al*, *Use of Library Materials: The University of Pittsburgh Study*, (New York: Marcel Dekker, 1979)
13. Collection Building Task Force, OCLC Research Project Checklist, 2007. Available at <http://platinum.ohiolink.edu/cbtf/oclcres/oclcchecklist.pdf>. (accessed Dec. 17, 2008)
14. Anne T. Gilliland, "The OhioLINK OCLC Collection Analysis Project: A Preliminary Report," *Collection Management* 33 (no. 1/2, 2008) 161-172
15. IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic records: Final Report*. München, K. G. Saur, 1998.
16. Hickey, Thomas B., and Jenny Toves. *Work Set Algorithm*, 2003. Available at <http://www.oclc.org/research/projects/fibr/algorithm.htm>. (accessed Dec. 16, 2008)
17. Buckland, Michael H., *Book Available and the Library User*. Pergamon Press, 1975.
18. *Ibid.*, p. 12-13.