Chapter One

GLOBAL COLLECTIVE RESOURCES

Introduction

In 2001, OCLC staff and member libraries, and many librarians and information specialists worldwide, celebrated the 30th anniversary of WorldCAT. OCLC was founded in 1967 as the Ohio College Library Consortium. The Union Catalog, later renamed WorldCAT, was founded in 1971. WorldCAT is now the largest and most comprehensive database of bibliographic information in the world. Through the merged catalogs of libraries around the world, resources are made available to libraries and their users that no single library could provide alone.

For the first 20 years of its existence WorldCAT was primarily a cataloging database also heavily utilized for interlibrary loan. Beginning in the latter 1980s, reference and other database services were added. By the year 2001, there were 37,297 libraries in 76 countries participating in the OCLC Library Cooperative. As of June 30, 2000, there were 42,476,614 total records in WorldCAT, with 767 million location listings for those materials. Of these bibiographic records, 35,570,633 or 83.7% were for books with the remaining 16.3% for other formats such as serials, maps, scores, sound recordings, computer files, and mixed materials. The language distribution at that time was 26,869,669 records for English language materials with the remainder divided among 376 languages.[1] These records are for materials spanning four millennia of recorded knowledge, from 2,000 BC to the present.

In 1999, WLN became a division of OCLC and the collection assessment and analysis programs of WLN became OCLC products. This study profiles the monographic

bibliographic contents of WorldCAT utilizing the OCLC/WLN iCAS software. Analysis is by variables of subject, imprint year, and language, for separate groupings of research, academic, public, school and special/other libraries. Public and school libraries are also analyzed with juvenile and adult audience levels. The results of the study will contribute to a greater appreciation and understanding of WorldCAT as an international resource including bibliographic records on the universe of publication available for research and resource sharing worldwide.

The Problem

Users of WorldCat view bibliographic records one at a time. The user sees the Atrees@ individually but has little idea of what the Aforest@ contains. Subject and key word searching do allow the construction of bibliographies so that subsets of the database can be viewed. But for the majority of users, there is no concept of WorldCat as a coherent whole or a single aggregated resources base that reflects the global universe of publication. The scope of the resources under bibliographic control in WorldCat has not been analyzed in a multi-dimensional mode, that is single analyses have been performed of imprints by year or number of records by language, but not analyses with more than one variable combined together.

The problem of defining a universe of publication for subject areas has long plagued collection development. It is difficult to judge the strengths or weaknesses of library collections when there are no benchmarks to use for relative measurement. Viewing WorldCat as reflecting an international universe of publication offers at least a partial solution to the problem of defining a universe of publication for subject analysis of library collections.

Another aspect of collection analysis is that of identifying publishing and collecting patterns over time. While statistics have been published annually for U.S. book production and a few Western European countries, statistics for international book production are still unreliable for comparative analysis. Can WorldCat, as the largest international bibliographic database be used as a surrogate for studying publishing trends and library collecting patterns?

The next section reviews the existing body of research on collecting patterns using WorldCat and OCLC collection analysis products.

Review of Related Research

Prior to the development of the OCLC Online Union Catalog, there were a number of automated collections analyses, primarily overlap studies, conducted by groups of libraries to determine the feasibility of establishing regional network consortia. The research literature of overlap studies has been reviewed by Potter[2] and Medina.[3] Both Perrault[4] and McNeely[5] have reviewed studies on the methodology and feasibility of using database records to analyze library holdings and collection analysis studies using automated data analysis. Nisonger has provided a useful annotated bibliography of *Collection Evaluation in Academic Libraries*[6] which has been supplemented by bibliographies appearing in *Collection Building*.[7] The bibliographies provide a broad overview of the research in collection evaluation.

This review only covers research that has utilized WorldCat or the OCLC/AMIGOS CACD to study aggregated collections at the national level. The majority of these studies have been on academic library collections.

Four studies by Schwartz have been reported which utilized both the OCLC Union

Catalog database and the OCLC/AMIGOS CACD for data collection.[8] Schwartz used the two systems in combination to measure literature loss in anthropology, Judaic studies, international relations and psychology. Book production for a specified time period was measured by subject searching the OCLC Union catalog. Those records retrieved were assumed to represent the universe of publication in the disciplines under study. Book production was compared to the aggregate holdings of the ARL libraries for the same time period derived from the OCLC/AMIGOS CACD database. The methodology is posited as a model for determining the growing gap between total book publication output and the aggregate holdings of a peer group of libraries.[9] Schwartz found that in international relations overall, "ARL holdings dropped from 75% of book output in the late 1970s to 65% of book output in the late 1980s." In anthropology for the same time period, holdings dropped from 66% of book production in the late 1970s to 53% in the late 1980s. For anthropology, 40% of the total book output was not held by any of the 70 ARL libraries and 10% of the titles were held by only one or a few of the libraries. The remaining 30% Schwartz characterizes as "mainstream literature" of which the average was 10% for any one of the 70 libraries. He concluded that literature loss defined as the "growing gap between book publication output and the aggregate holdings of 70 institutions in the Association of Research Libraries ...grows rather steadily".[10]

Research by Perrault was designed to determine the effects of the decline in book acquisitions on the collective resources base of 72 ARL libraries.[11] The study utilized the 1991 OCLC/AMIGOS CACD to compare the two imprint years 1985 and 1989. This was the first reported research to study the ARL library collections as an aggregated resources base by both subject and language grouping parameters. The findings corroborated the trends in monographs acquisitions that had been tracked in the *ARL Statistics* series for

those years. The 1990/91 *ARL Statistics* showed a decrease of 14% in monograph purchases for the previous five years, with 1988 having the lowest point at a 20% decrease.[12] Among the findings of Perrault=s research were that, as of the time of data extraction in 1991, overall there was a 27.76% decline in the total number of books acquired by the libraries in 1989 as compared to 1985. English language titles had declined 12.34%, while foreign language books declined more steeply by 43.33%. A shift in collecting patterns occurred between 1985 and 1989 within the three broad knowledge groupings with the humanities declining in proportional share of total by 2.35 percentage points and the sciences increasing share by 2.99 percentage points. The social sciences remained constant in percentage share of total as well as percentage of unique titles to total and in the mean number of holding libraries.[13]

In an article on the methodology used in the ARL study, Perrault provided examples of collection analysis utilizing the OCLC/AMIGOS CACD product, updating the findings of the ARL study with data extracted from the 1996 OCLC/AMIGOS CACD.[14] This article reported on three peer groups of academic libraries, with all records in the CACD database composing a fourth peer group. The size of the collections for the four peer groups were compared for the time frame 1987-1995. The analysis concentrated on the trends in collecting for six major subject areas over the nine years. The conclusions of the study were that the aggregated resources of academic libraries in the United States are becoming increasingly composed of English language publications; the budget crisis of the 1980s is still visible in reduced numbers of titles for those years across all subject areas in 1996; smaller academic libraries have more stable collecting patterns indicating that the larger research institutions were harder hit by the serials crisis in the 1980s; and the number of titles for all categories rises in the early 1990s indicating somewhat of a

recovery from the decline in the 1980s.[15]

A broader study of the decline in foreign acquisitions in academic libraries on the 1980s was conducted for the AAU/ARL Foreign Acquisitions Project. The project was Adirected toward developing a clear understanding of the forces influencing North American research libraries= ability to build and maintain collections of publications produced outside the United States and Canada.@[16] The study utilized database scans for foreign publications in OCLC WorldCat with imprint dates of 1988-1993. The data were extracted for publications by country of origin but not by language parameters so that foreign English language publications are included in the counts. These data are not comparable to data from the CACD database in the research by Perrault. No subject analyses were conducted for the AAU/ARL Foreign Acquisitions Project. The emphasis of the investigation was on materials published in eight different geographic areas outside the United States. The results of the four-year study were analyzed in *Scholarship, Research Libraries, and Global Publishing*, authored by Reed-Scott.[17]

The trend analysis for foreign publishing included in *Scholarship, Research Libraries, and Global Publishing*, was partially replicated and updated in an ARL AGlobal Resources Report@ in 1999.[18] The update showed that publishing output outside of the U.S. had continued to increase in the latter 1990s. The data also show that the number of foreign titles with imprint dates from 1988-1994 had considerably increased in the four years after the initial data extraction. The increases in the number of imprints on the older end (1988) which were added are smaller than the increases beginning with 1991 imprints. The largest increase was for 1994 titles, the last imprint year of the earlier study, an increase of 67% being added in the 1995-1999 time frame. From the updated data, it appears that the time lag in adding foreign imprints to WorldCat is in the three to five year

range, with the number added beginning to slow after five years. Nonetheless, the updated data substantiated the earlier study in 1995, that North American Aresearch libraries collection coverage of global publishing output is declining.@[19]

Both Holleman[20] and McNeely[21] have conducted studies updating the Perrault studies. Holleman performed a small scale study with only the world history LC subject divisions using the 1995 edition of the OCLC/AMIGOS CACD. He found that by 1995 the number of 1989 imprints showed a .56% increase over the number of 1985 imprints as found by Perrault in the 1991 edition of the CACD. In the four years between the two editions of the CACD, a decline of 27.76% in the number of titles from 1985-1989 had become a modest increase as libraries continued to add titles from the latter 1980s. Holleman=s findings called attention to a problem with the reliability of the numbers of records in the last few imprint years in each edition of the CACD due to a time lag in library acquisitions and cataloging.

The McNeely study utilizes the Perrault methodology to analyze the changes in book collecting patterns of three peer groups: (1) members of the Association of Research Libraries (ARL); (2) libraries included in the OCLC/AMIGOS CACD, 1998 edition; and (3) TexShare Libraries. Data were analyzed for the period 1985-1995 by 104 Library of Congress classification ranges and for seven language groupings formulated for the Perrault studies. Unique titles by the three peer groups and subject categories were aggregated into humanities, social sciences, and sciences. Yearly changes, rate of change, and yearly average rate of change were calculated for LC subject categories, language groups, broad subject groups, and unique title counts. Findings were that the ARL and OCLC peer groups had an average yearly rate of growth between 2-3% for the ten year period. TexShare libraries had between 1-2%. OCLC had the largest growth in

the humanities; ARL had the largest in the social sciences, and TexShare had the largest growth in the sciences. The number of unique titles(diversity) in the collections increased as measured by yearly average for all peer groups: OCLC 3.76%, ARL 3.l8%, and TexShare, 1.64%. There was a 3% average yearly decline in percentage of non-English titles for this period. McNeely also found that the annual rate of change went into a negative mode for the last two imprint years in the study, as had the earlier studies by Perrault, further substantiating Holleman=s findings of the number of titles by subject divisions by imprint years continuing to increase over time. Overall, McNeely=s longitudinal study shows that Athe >national collection= continued to grow at an average yearly rate of 2-3 percent for the period studied (1985-1995).@[22]

Hardesty and Mak conducted a study of college library collections using both the CACD and data extracted from WorldCat. The collection size range was 100,000 to 299,000 volumes. The purpose of the study was to ascertain if there was an identifiable core of materials being collected by the group of college libraries. The findings were that there was not an identifiable core, but that the holdings were scattered.[23]

Senkevitch and Sweetland studied adult fiction collections in over 4,000 OCLC member public libraries under the auspices of an OCLC research grant.[24] The study examined the extent of change over a one-year period in the most widely held adult fiction titles and sought to answer the question AIs there a consensus core of adult fiction?@ Findings were that the most widely held adult fiction titles were Arecent, popular works that form a stable core from one year to the next. Results also suggest that such a listing of widely held titles might be suitable as an evaluation tool in smaller public libraries.@[25]

Two studies were reported which utilized data from OCLC or the CACD to compare library holdings with *Choice* AOutstanding Academic Books@ lists. Budd and Craven[26]

investigated the holdings of four peer groups of academic libraries for the intervals 1984/85, 1990, and 1995. Data were extracted from two editions of the OCLC/AMIGOS CACD. The researchers took a purposive sample of titles from the *Choice* AOutstanding Academic Books@ lists for those years. Titles were chosen for major academic disciplines representing the humanities, social sciences, and sciences. Findings demonstrated that Aacross all subjects and for broad subject areas, there have been significant declines in holdings over time and for all groupings of libraries.@ The researchers maintain that the findings of other studies with regard to the Ashrinking of library resources is corroborated in this study.@[27]

Although not a study of academic library collections in the aggregate, a survey by Sweetland and Christensen of selection practices for language and literature collections in Wisconsin academic libraries used OCLC data to ascertain the number of holding libraries for *Choice* titles.[28] The researchers used *Choice* 1993 AOutstanding@ titles in languages and literature. The list was published in January 1993. The data were extracted from the OCLC Online Catalog in March, 1995, two years later. Findings were that, on the average, about 10 libraries held each title with the range being from 4 to 23 libraries per title. The authors observe, A...considering that these titles are supposedly the most outstanding of all those recommended by *Choice* for four-year college libraries, the small number of holdings is curious, to say the least.@ Part of the problem is that from the study data it appeared that Aa substantial number of libraries did not add books to their collection until sometime in 1994 (or very late in 1993).@[29] The researchers give as a possible explanation for the delay in adding material the almost total reliance upon faculty recommendations with faculty using scholarly reviews to identify new titles. The researchers conclude that Afaculty still control selection in most [college] libraries;@ Athe libraries in the study still consider current,

immediate needs, and rarely consider other library holdings in selecting material to purchase;@ Aeven though libraries feel they are buying current material, their definition of ›current= means within the last couple of years;@ and Arespondents are currently buying very little outside the English language.@[30]

The majority of the research utilizing data from WorldCat or the OCLC/AMIGOS CACD has been on national collecting patterns in North American libraries. It has been heavily focused on the ARL research library group. There have been fewer reported national studies of smaller academic or public library collections. This review of related research shows that there is still much to be learned about the collective resources of the libraries contributing to WorldCat and the state of those collections in the latter years of the 20th century. While broad profiles of the database have been produced by the OCLC research office, the monographic bibliographic records in WorldCat have not been formally studied for collecting patterns by subject, language, and type of library and level of audience parameters. This study is the first reported research on WorldCat as reflective of global collective resources and the international universe of publication. The methodology for the study is outlined in the next section.

Methods

The study is a bibliometric content analysis of OCLC WorldCat, a merged, electronic catalog of the bibliographic and holdings information of OCLC member libraries, founded in 1971. The more than 45 million records in WorldCat as of 2001 represent more than 750 million items held in libraries around the world. The study was designed to take advantage of the data analysis capabilities of the OCLC/WLN iCAS software product which is a CD-ROM tool for collection analysis. (For information on iCAS go to

http://www.oclc.org/western/products/aca/icas.htm)

The iCAS software performs basic age and content analyses. Data are shown in a grid with the customer library=s choice of publication date ranges across the top and WLN conspectus division, categories, and subjects down the side. Each cell contains the number of titles held for that subject area and publication date/date range, as well as its percentage of the division, category or subject. Graphing, printing and export options are included.

Data Collection and Analysis

A 10% systematic random sample of the monographic bibliographic records in WorldCat was extracted in the OCLC Office of Research and forwarded to the Lacey Product Center of OCLC in February 2001. The number of MARC bibliographic records in the sample was 3,555,907. Each MARC record in the sample carried an identifier for a type of library grouping as research, academic, public, school, or special\other. The ICAS software was utilized to profile the bibliographic records in the WorldCat sample dataset according to the following parameters:

$ imprint year analyses -- Dates and date ranges are: by century until 1800; 1800 - 1849,1850-1899; by decade until 1980; 1980-1984; then annual through 2000.

$ type of library categories -- Research, academic, public libraries, special/other, and school

$ subjects WorldCat, research grouping and the academic group holdings are analyzed by WLN Library of Congress Conspectus categories; the public and school library groups holdings are analyzed as WLN Dewey Conspectus categories

$ juvenile and adult subtotals B for the public and school library groupings, analyzed as Dewey Conspectus categories and publication year

$ language groupings BEnglish, Chinese, French, German, Japanese, Russian and

Spanish for WorldCat, the research and the academic library groupings analyzed by publication year and WLN LC Conspectus categories

The iCAS data analysis for WorldCat was completed in June, 2001. A number of different analyses were conducted by staff at WLN on the monographic bibliographic records in the 10% systematic sample of the WorldCat database. The total number of records in the sample is 3,555,907. Of those records a small number, 4% of the records, had no library holdings symbol or were not owned by any of the libraries in the sample. These records were not used at all. Of the remaining records, 3,376,272 records were used in the language analysis. The language analysis is performed by sorting on the language codes in the fixed field of the record. The same dataset was used to run an analysis of the records to determine the number of records with ISBN numbers.

With regard to publication dates, an Aother@ category in the analysis by imprint years contains the number of records that had no publication date. Approximately 1% of the records did not have publication dates, too small a number to have any effect on the findings.

In order to perform subject analyses the bibliographic record has to have a Dewey, Library of Congress, or NLM call number. Of the 3,376,272 usable records, 1,179,107 did not have call numbers. Hence 35% of the sample records could not be used for subject analysis. This set of records formed a second dataset that appears as ANo call number present@ in many of the tables in the study. Throughout this report it is indicated which of the datasets have been used for the particular analysis under discussion.

After preliminary analysis of the data and conferring with Lacey Product Center staff, the researchers requested that a number of additional analyses be run. These included an analysis of WorldCat, the research libraries, and the academic libraries with the

ANo Call Number@ records included as one line so that a total number of records from the sample could be accounted for in each of the three largest data groupings. Separate analyses were run for unique records for these three groupings and an analysis of records by defined ISBN number groupings for WorldCat. The second round of iCAS analyses were delivered in October 2001.

It is important to bear in mind that the data in this study form a Asnapshot@ of the contents of WorldCat at one point in time. Bibliographic records are continually added by OCLC member libraries for imprints across the entire time span defined for the study. The analysis in this study concentrates on proportions, percentages of total, and trends. These proportions should represent the profile of the population of monographic bibliographic records in WorldCat from which the sample was drawn. Since the absolute numbers are from the sample, these numbers cannot be utilized for benchmarks, only for the findings of proportions, percentages, and trends. Were another sample to be extracted in several years, the absolute numbers would be different, but the proportions or trends should be nearly the same. While discernible shifts over time can occur in the profile of the database, such shifts would necessarily involve a large number of bibliographic records in a database of over 45 million records. There are a small number of studies in the literature review in this chapter with which the findings of the WorldCat study can be compared. But because the data for these previous studies were extracted from the WorldCat database at different times, again, the findings can only be compared in proportions or trends and not in exact numbers of records.

The proportions found in the data analysis in this study can be influenced by the collecting patterns of the library membership of OCLC. In the separate analysis for each of the type of library groupings, the composition of member libraries for each group is

described in general terms. A listing of library membership in OCLC with the type of library code can be accessed at http://www.oclc.org/contacts/libraries/.

The report of findings begins with Chapter two which profiles WorldCat as reflecting the international universe of publication over time. Subsequent chapters analyze the profiles of the aggregated collections of the type of library groupings, the language groupings, unique titles, and an ISBN analysis.

Click here for Chapter Two.

References

1. OCLC Annual Report, 1999/2000: 46-47.

2. William G. Potter, "Studies of Collection Overlap: a Literature Review," *Library Research* 4 (1982): 2-21.

3. Sue O. Medina, "Duplication and Overlap Among Library Collections: A Chronological Review of the Literature," *Advances in Collection Development and Resource Management* v.1, ed. Thomas W. Leonhardt (Greenwich CT: JAI Press, 1995): 1-60.

4. Anna H. Perrault. AThe Changing Print Resource Base of Academic Libraries in the United States: a Comparison of Collection Patterns in Seventy-two ARL Academic Libraries of Non-serial Imprints for the Years 1985 and 1989.@ Ph.D. dissertation, The Florida State University, 1994.

5. Bonnie Michaels McNeely. ALongitudinal Study of the Collection Development Patterns of Libraries in the Association of Research Libraries and the Online Computer Library Center, and Texas Academic Libraries Using the OCLC/AMIGOS Collection Analysis Compact Disc to Assess Strengths, Weaknesses, and Overlap for 1985-1995.@ Ph.D. dissertation, Texas Woman=s University, 2001.

6. Thomas E. Nisonger. *Collection Evaluation in Academic Libraries: A Literature Guide and Annotated Bibliography*. (Englewood, CO: Libraries Unlimited, 1992)

7. Thomas E. Nisonger. AThe Collection Development Literature of 1996: a Bibliographic Essay.@ *Collection Building* 17, no. 1(1998): 29-39; AA Review of the 1997 Collection Development and Management Literature.@ *Collection Building* 18, no.2(1999): 67-80.

8. Charles A. Schwartz, "Literature Loss in Anthropology," *Current Anthropology* 33 (June 1992) 315-317; "Empirical Analysis of Literature Loss," *Library Resources and Technical Services*, 38, no.2 (April 1994), 133-138; "Literature Loss in International Relations," *PS: Political Science & Policy* 25 (December 1992), 720-723; "Gap Analysis of Book Publication Output and Aggregate Research Collections in Psychology," *American Psychologist* 48 (1993): 1151-52.

9. Schwartz, AEmpirical Analysis of literature loss.@

10. Ibid.

11. Anna H. Perrault, "The Shrinking National Collection: a Study of the Effects of the Diversion of Funds from Monographs to Serials on the Monograph Collections of Research Libraries," *Library Acquisitions Practice & Theory*, 18, no.1 (1994): 3-22.; "The Changing Print Resource Base of Academic Research Libraries in the United States," *JELIS* (*Journal of Education for Library and Information Science*) 36, no.4 (Fall 1995): 1-16.

12. *ARL Statistics*, 1990/91.

13. Perrault, AThe Shrinking National collection...@ p.

14. Anna H. Perrault. ANational Collecting Trends: Collection Analysis Methods and Findings.@ *Library & Information Science Research* 21, no.1(1999): 47-67.

15. Ibid., p.65.

16. Jutta Reed-Scott, *Scholarship, Research Libraries, and Global Publishing*, (Washington, DC: Association of Research Libraries, 1996), xvii.

17. Ibid.

18. Deborah Jakubs. AThe AAU/ARL Global Resources Program: Both Macrocosm and Microcosm.@ *ARL* No. 206 Special Issue on Global Resources (October 1999).

19. Ibid., p.8.

20. Curt Holleman. AThe Study of Subject Strengths, Overlap, and National Collecting Patterns: The Uses of the OCLC/AMIGOS collection Analysis CD and Alternatives to It.@ Collection management 22, no. 2: 57-69.

21. McNeely. ALongitudinal Study of the Collection Development Patterns....

22. Ibid. v-vi.

23. Larry Hardesty and Collette Mak. ASearching for the Holy Grail.@ *The Journal of Academic Librarianship* 19, no.6 (1994): pp. 362-371.

24. Judith J. Senkevitch and James H. Sweetland. APublic Libraries and Adult Fiction: Another Look at a Core List of Classics.@ *Library Resources & Technical Services* 42, no.2 (1998): p.102-112.

25. Ibid., p.102.

26. John M. Budd and Catherine K. Craven. AAcademic Library monographic Acquisitions: Selection of *Choice*=s Outstanding Academic Books.@ *Library Collections, Acquisitions, & Technical Services* 23, no.1(1999): 15-26.

27. Ibid., p.122.

28. James H. Sweetland and Peter G. Christensen. ADeveloping language and Literature Collections in Academic Libraries: a Survey. *Journal of Academic Librarianship* v (March 1997):119-125.

29. Ibid., p.122.

30. Ibid.,p.123.