Changing Global Book Collection
Patterns in ARL Libraries

Mary E. Jackson, Auto-Graphics, Inc. (formerly Association of Research Libraries)
Lynn Silipigni Connaway, OCLC Online Computer Library Center, Inc.
Edward T. O’Neill, OCLC Online Computer Library Center, Inc.
Eudora Loh, University of California, Los Angeles

2006
Changing Global Book Collection Patterns in ARL Libraries

Mary E. Jackson
Lynn Silipigni Connaway
Edward T. O’Neill
Eudora Loh

Association of Research Libraries
21 Dupont Circle, NW, Suite 800
Washington, D.C. 20036

© 2006 Association of Research Libraries

The Global Resources Network (GRN) is a voluntary and collaborative initiative of higher academic research institutions to expand and enhance access to international scholarly resources. The Center for Research Libraries (CRL) provides administrative, technical, legal, financial management, and communications support for the regular activities of the GRN and its related projects. The Association of Research Libraries (ARL) and the Association of American Universities (AAU) maintain a presence on the GRN Advisory Committee and many ARL member libraries are active participants in GRN and its projects.

For current information on GRN, see http://www.crl.edu/grn/index.asp or contact James Simon, GRN Director, Center for Research Libraries, at simon@crl.edu.
Changing Global Book Collection Patterns in ARL Libraries

Introduction

The overlap of global book collections in member libraries of the Association of Research Libraries (ARL), as exemplified by an analysis of nearly 28.5 million records in a July 2004 snapshot of the OCLC WorldCat database, is not as extensive as expected.

For all countries combined, excluding North America, the average number of ARL library holdings per bibliographic record is fewer than five, that is, fewer than five ARL libraries own copies of the book represented by the bibliographic record.

For each of ten world regions, excluding North America, the average number of ARL library holdings per bibliographic record ranges from a low of 2.99 for East Asian books to a high of 6.28 for Latin American books. That is, on average fewer than three and only slightly more than six ARL libraries hold any book that was published outside of North America.

These key findings emerged from a study initiated by the Association of American Universities (AAU)/Association of Research Libraries (ARL)/Center for Research Libraries (CRL) Global Resources Network (GRN) Advisory Committee to understand the character and nature of foreign materials held by ARL member libraries and to identify the extent of gaps and overlaps in research library collections. Background information on GRN and additional details about the snapshot of WorldCat, the OCLC Online Computer Library Center, Inc. bibliographic database, and the methodology used in this study are included in the appendices.

This report highlights the major findings from an analysis of book records and holdings of ARL member libraries that contribute cataloging to the OCLC WorldCat database. For this analysis, books were defined as print books including more than forty-eight pages; government documents and manuscripts were excluded. In certain instances, comparisons are made to records and holdings of all OCLC cataloging member libraries. The findings are presented as answers to a series of questions relating to bibliographic records, holdings, countries of publication, and publication dates.

How Many Global Book Publications are Owned by ARL Libraries?
A July 2004 snapshot of the OCLC WorldCat database included the total number of “book” publications cataloged by OCLC member libraries (“records”) and the total number of holdings for each country of publication. The snapshot consisted of over 29 million bibliographic records and more than 725 million holdings. Analysis of the snapshot led to the elimination of some records, resulting in 28,432,888 usable bibliographic records and 719,012,619 useable holdings. The snapshot also included comparable subsets of ARL libraries’ records and holdings.¹

¹See Appendix B for a more detailed explanation of the snapshot.
The holding symbol for at least one ARL library was attached to 19.7 million records, or 69 percent of the useable records. Though these records represent a variety of book publications, from research-level, scholarly, and specialized titles to romance fiction, how-to manuals, children’s literature, and other items outside the scope of collection development policies of most ARL libraries, it was not possible to exclude records for non-research materials from the sample. With the understanding that the holding symbols of ARL libraries were found on more than two-thirds of the records, the investigators concluded that the snapshot was sufficiently representative of research-level materials and other scholarly resources relevant to faculty, students, and scholars using libraries in ARL member institutions.

What is the Distribution of Records by Year of Publication?
The useable bibliographic records were sorted into three broad date ranges: pre-1600 through 1954 (including records with no or unknown dates), 1955 to 1979, and 1980 to 2004. These date ranges were selected to reflect major trends in publishing and library collecting patterns. Just under half (46 percent) of the 28.4 million records were published between 1980 and 2004, with the other half split fairly evenly between the two older date ranges. Figure 1 summarizes the date ranges of publications held by ARL libraries and by other OCLC member libraries. The ARL libraries’ holdings are not included in the number of other OCLC member libraries’ holdings.

Figure 1: Distribution of WorldCat Records, by Publication Date Ranges

![Figure 1: Distribution of WorldCat Records, by Publication Date Ranges](chart.png)
What is the Distribution of Records by World Regions?
ARL libraries have contributed records to WorldCat for books published in 245 countries. For these libraries, the number of records per country ranges from one (East Timor, Johnson Atoll, and Midway Islands) to 1.3 million, 1.6 million, and 2.2 million for books published in France, Germany, and the United Kingdom (UK) respectively. The average number of ARL library records per country is 80,512 and the median is 2,196.

The 245 countries were grouped into 11 world regions. Most of this report presents findings for ten of the eleven world regions. North America was excluded as one of the world regions in order to eliminate US publications. As a consequence, publications from Canada and Saint Pierre and Miquelon were also excluded. Figure 2 details the number of records owned by ARL libraries by the ten world regions.

ARL libraries own more titles published in the 27 Western European countries than the other nine world regions combined. Publications from the Western European countries represent 58.47 percent of all ARL libraries’ book records. This strong representation reflects current and historical library collecting patterns, institutional programmatic strengths, and publishing patterns within a country. Publications from the six East Asian countries represent the second largest cluster of records owned by ARL libraries (11.95 percent). Publications from the 27 Eastern European countries account for another 11.10 percent of records, and the remaining seven world regions account for the final 18.48 percent of records.
Appendix C lists the countries included in each of the regions.

What do ARL Libraries Hold?

OCLC Research developed an interactive prototype system that provides a visual tool for managing and representing geographically based library collections and library data. The collection data were mined from WorldCat, which contained more than 58 million records in July 2004. The WorldMap, as shown in Figure 3, can be accessed at http://www.oclc.org/research/researchworks/worldmap/prototype.htm.

Figure 3: OCLC ResearchWorks WorldMap
What is the Distribution of Holdings by Year of Publication?
ARL libraries (main and branch/departmental) are represented by nearly 166 million holdings, or just 23 percent of the 719 million useable holdings in the sample (including North America). The number of holdings sorted by the three date ranges, in Figure 4, compares the holdings of ARL and other OCLC member libraries. ARL libraries account for 36 percent of all holdings for books published through 1954, 22 percent of all holdings for books published between 1955 and 1979, and 20 percent of all holdings for books published between 1980 and 2004. The ARL libraries’ holdings are not included in the number of other OCLC member libraries’ holdings.

Figure 4: Comparing ARL Libraries’ WorldCat Holdings to Other OCLC Libraries’ WorldCat Holdings, by Publication Date Ranges
What is the Distribution of Holdings by World Regions?
Figure 5 represents ARL library holdings in WorldCat by region of publication, excluding North America. The nearly 56.8 million ARL library holdings for books published in Western European countries underscore the historical strength and dominance of collecting patterns by ARL libraries. Holdings for the other world regions are relatively proportional to the number of records for each region. (See Figure 2 for comparison.)

Figure 5: ARL Libraries’ Holdings in WorldCat, by Region of Publication
How Extensive are the Holdings of ARL Libraries?
Another way of understanding the comprehensiveness of the collections in ARL libraries is to look at the percentage of total holdings for these libraries. ARL holdings account for more than half of the total OCLC holdings in the ten world regions outside North America. The percentages range from a little more than half of the total holdings of Western European publications to more than three quarters of the total holdings for books published in Eastern European countries. Figure 6 summarizes the percentage of total ARL library holdings in WorldCat by world region.

Figure 6: Percentage of Total ARL Library Holdings in WorldCat, by World Region
How Do Collecting Patterns Vary by Publication Date and World Region?
The investigators explored whether collecting patterns and holdings varied for more recent publications. They created a subset of books with publication dates between 1980 and 2004. For each of the ten world regions, the investigators calculated the total number of titles and holdings in ARL member libraries for each five-year period between 1980 and 2004. Figure 7 presents the total number of ARL libraries’ records and Figure 8 presents the average ARL library holdings by world region for 1980 to 2004 publications.

*Figure 7: Total Number of ARL Libraries’ WorldCat Records, by World Region: 1980–2004*
For most of the ten world regions, ARL libraries hold, on average, fewer titles of books published between 2000 and 2004 than books published between 1980 and 1984. The decrease is illustrated in Figure 8. This may be attributed to cataloging backlogs reflected in WorldCat, as reported by Anna Perrault.3

To test the effect of cataloging and processing backlogs, a comparison of 1980–1984 and 1995–1999 publications was undertaken. In that comparison, there is an increase in the average number of ARL library holdings for publications from Africa, Caribbean, Middle East, Oceania, and Southeast Asia. That is, ARL libraries hold more copies of 1995–1999 publications than hold copies of 1980–1984 publications. However, in each of the cases the increase is rather modest: for example, from 4.04 to 4.05 average holdings for the Caribbean region to 3.11 to 4.28 average for the Middle Eastern region.

Figure 8: Average ARL Libraries’ WorldCat Holdings, by World Region: 1980–2004

How Extensive is the Overlap among ARL Libraries?

ARL libraries hold fewer titles of books from all ten world regions than might be expected. The average holdings range from 2.99 for East Asian books to 6.28 for Latin American books. On average, ARL libraries have fewer than five holdings in WorldCat for books published in seven of the ten world regions. Figure 9 compares the average ARL library holdings with the holdings of other OCLC member libraries. The ARL libraries’ holdings are not included in the number of other OCLC member libraries’ holdings.

Figure 9: Comparing Average ARL Libraries’ WorldCat Holdings and Other OCLC Libraries’ WorldCat Holdings, by World Region
How Do ARL Holdings Compare with Total Holdings by Country of Publication?
ARL libraries represent 100 percent of the total OCLC holdings for all years of publication for five countries: East Timor, Heard and McDonald Islands, Johnston Atoll, Midway Islands, and Tokelau. At the other extreme, ARL libraries hold just 24.67 percent of Kiribati books, 25.43 percent of Nauru books, and 25.87 percent of UK books.
However, these percentages may be misleading, as the number of book records for these countries varies significantly. ARL libraries reported 2,375 holdings for the 611 records of Kiribati books and 369 holdings for the 81 records of Nauru books. On the other hand, Tokelau had only 12 ARL holdings attached to 7 records; Heard and McDonald Islands had 3 ARL holdings for 2 records; Johnston Atoll and the Midway Islands had 2 ARL holdings each for 1 record each; and 1 ARL library’s holding symbol was attached to the single record for a book published in East Timor.

Are There Variations in ARL Holdings Within a World Region?
The average number of ARL library holdings per record varies by world region and by countries within a region. Analysis by the ten world regions provides a high-level view of collecting patterns by ARL libraries and by other OCLC libraries. Another way of exploring ARL libraries’ coverage is by reviewing records by country of publication. Since there is no accurate calculation to determine the annual publishing output of each country, it is impossible to confirm whether WorldCat records for a country represent comprehensive or minimal collecting by ARL libraries or other OCLC member libraries.
Given that caution, reviewing ARL library collecting patterns on a country-by-country basis is illustrative. Appendix D includes four charts that illustrate the range of average ARL library holdings for countries in the world regions of East Asia, Western Europe, Eastern Europe, and Latin America. Western Europe, East Asia, and Eastern Europe have been selected because they represent regions with the largest number of ARL records. Western Europe, Eastern Europe, and Latin America are the three regions with the largest number of ARL holdings, while East Asia and Latin America represent the world regions with the smallest and largest average number of holdings respectively.
What Countries Represent the Largest Number of ARL Records?
Excluding records from the three North American countries and records with no or unknown publication locations, the top ten countries in terms of number of ARL libraries’ records are: UK, Germany, France, Japan, Italy, China, Spain, Russian Federation, India, and the Netherlands. The number of ARL libraries’ records ranged from just under 2.2 million for UK books to just under 295,000 for books from the Netherlands.

This subset of records was reviewed by five-year time periods between 1980 and 2004. Figure 10 highlights the change in the total number of records during the twenty-five-year time period. Six countries (China, France, Italy, Netherlands, Spain, and the UK) have more records for 2000–2004 publications than for 1980–1984 publications. These increases may reflect increases in overall publishing output or increases in the number of books purchased by ARL libraries. Four countries (Germany, India, Japan, and the Russian Federation) have fewer records for books published between 2000 and 2004 than for books published between 1980 and 1984. These decreases may reflect drops in publishing output, fewer purchases by ARL libraries, or cataloging backlogs.

*Figure 10: Top Ten Countries Represented by ARL Libraries’ Records in WorldCat: 1980–1984 to 2000–2004*
Are Average ARL Holdings Increasing or Decreasing for These Ten Countries?
For each of the ten countries, the investigators calculated the average ARL library holdings for each five-year period between 1980 and 2004. The average holdings decreased in each five-year increment between 1980 and 2004 for all countries except the UK, which showed a slight increase—from an average of 16.81 holdings for 1980–1984 publications to 17.61 holdings for 2000–2004 publications. Figure 11 highlights the average number of ARL libraries’ holdings for the ten countries between 1980 and 2004.

Figure 11: Comparing Average ARL Libraries’ Holdings in WorldCat: 1980–1984 to 2000–2004

The decreases in average ARL libraries’ holdings for titles published between 1980 and 2004 in the nine countries, especially for 2000 to 2004 publications, may reflect the overall growth in publishing output, increasing frequency of significant currency fluctuations, stable or shrinking library material budgets, shifts in research library collecting strategies, fewer purchases by ARL libraries, or a combination of two or more of these factors.

The decreases may also be indicative of delays in cataloging and adding records and/or holdings to WorldCat. To test the impact of those delays, records with publications dates of 2000 through 2004 were eliminated from the subset of records. When the 1980 to 1984 and the 1995 to 1999 time periods were compared, all countries, again with the exception of the UK, showed decreases in the average number of ARL holdings. These decreases suggest that fewer ARL libraries are acquiring copies of books from these nine countries than they purchased a quarter century ago.

For example, the drop in the average number of ARL libraries’ holdings of titles published in India likely illustrates the effect of the cessation of the Library of Congress’s PL-480 program in

A second example is publications from Spain. It is reasonable to assume that catalogers fluent in Spanish are found in more ARL libraries than catalogers with expertise in less familiar languages. Even discounting the challenges of cataloging Spanish books, the average number of holdings for books from Spain slipped from 7.23 for 1980–1984 publications, to 6.57 for 1985–1989 publications, to 5.85 for 1990–1994 publications, to 5.22 for 1995–1999 publications, and, finally, to 4.97 for 2000–2004 publications. The overall average number of holdings for Spanish publications published between 1980 and 2004 is 5.84. Excluding the 2000 to 2004 time period because of cataloging backlogs, ARL libraries hold an average of 6.21 copies per title published in Spain between 1980 and 1999, representing just 5 percent of the total ARL membership. This downward trend in the average number of ARL holdings of books published in Spain between 1980 and 1999 provides additional evidence to suggest that ARL libraries are purchasing fewer copies than two decades previously.

The final example may illustrate the results of collaborative action. The GRN Latin Americanist Research Resources Project’s Distributed Resources initiative is aimed at strengthening the collective coverage of monographs and other resources produced in Latin America. Participating libraries have voluntarily agreed to reallocate at least seven percent of their Latin American monographic budgets to deepen their collections in established areas of local emphasis. This reallocation may be reflected, in part, in the increase in the number of books published in Bolivia, Mexico, Paraguay, and Peru held in WorldCat by ARL libraries between 1980 to 1984 and 2000 to 2004. However, over that same timeframe, the average number of ARL holdings for those four countries decreased, suggesting that project participants are purchasing more unique titles and reducing duplication of more commonly held titles. Figure 12 illustrates this trend for the four Latin American countries.

Figure 12: Possible Impact of Collaborative Purchasing

<table>
<thead>
<tr>
<th>Country</th>
<th>Change in Average Number of Titles held by ARL Libraries, 1980 to 2004</th>
<th>Change in Average Number of Holdings in ARL Libraries, 1980 to 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>+1,539</td>
<td>-0.09</td>
</tr>
<tr>
<td>Bolivia</td>
<td>+440</td>
<td>-1.30</td>
</tr>
<tr>
<td>Paraguay</td>
<td>+170</td>
<td>-3.47</td>
</tr>
<tr>
<td>Peru</td>
<td>+60</td>
<td>-1.74</td>
</tr>
</tbody>
</table>

**How Do the Findings from this Investigation Compare To Previous Studies?**

Jutta Reed-Scott’s *Scholarship, Research Libraries, and Global Publishing* characterized growth patterns of foreign language materials using a January 1995 snapshot of the OCLC WorldCat database. That snapshot included unique bibliographic records from all OCLC members with publication dates for the period 1988 through 1994.

Reed-Scott noted that as “cataloging backlogs are reduced, the total number of titles cataloged will increase over time in each of the world areas.” The investigators tested that statement by
comparing the number of records with 1990 to 1993 publication dates cataloged by all OCLC members using the 1995 and 2004 snapshots. The 2004 snapshot did not break down publication totals by individual years as the 1995 snapshot did. As a result, the number of total records per region for publication years 1990–1994 in the 2004 snapshot was reduced by 20 percent to have comparable four-year periods. This reduction was made with the awareness that the number of books published each year is not constant. Figure 13 presents the comparison.

However, Figure 13 may also illustrate the effect of records contributed by new OCLC members, particularly those from outside North America. In fiscal year 1994–95, OCLC had 6,877 governing members (i.e., libraries that contribute cataloging records), of which 25 percent were academic libraries. In 2003–04, OCLC reported 8,597 governing members, 31 percent of which were academic libraries.\(^5\) The globalization of the OCLC membership is reflected in the increased internationalization of records in the WorldCat database. For example, Japan’s Waseda University Library contributed 750,000 MARC records and 1.4 million holdings to the WorldCat database between 1995 and 2005. The National Library of Chile contributed 410,000 records, including 230,000 original records, in June 2004. It is highly likely that some of the records contributed by Waseda University Library and the National Library of Chile are for titles not previously included in WorldCat. These new records may have contributed to the increase in the total number of records for the East Asian and Latin American world regions respectively.

Figure 13: Possible Effects of Cataloging and Processing Backlogs for 1990–1993 Publications


Conclusion
Analysis of nearly 28.4 million records from a July 2004 snapshot of the OCLC WorldCat database sought to identify collection patterns of international resources as documented by WorldCat. Analysis was undertaken using country and year of publication. It is important to remember that WorldCat does not include records for all publications nor does it include all holdings of all libraries in the world. However, WorldCat is the largest bibliographic database in the world and its records and holdings are indicative of the extent of coverage by ARL member libraries and other OCLC member libraries.

The analysis provides a baseline aggregate description of research library collections by country of publication and publication date. The study documented changes in collecting patterns between 1955 and 2004. The findings are one effort to describe the current state of international and foreign language publications available in ARL libraries. Comparing future analyses with this baseline will provide one way of measuring progress toward the goals of the Global Resources Network.

This analysis identified several world areas where expanded coverage may be needed or desired (i.e., regions with low average number of records and holdings) and world areas where access via interlibrary loan and document delivery may be preferable or is now feasible (i.e., regions with high average number of holdings).

The relative low average number of holdings per record for many countries and for many time periods suggests that libraries build and maintain collections that reflect local academic programs and faculty research interests or areas of historical strength. On the other hand, the low average number of holdings may also suggest that libraries have been fairly successful in formal or informal collaborative collection development efforts. Finally, the low number of holdings argues for the need for effective interlibrary loan and document delivery services as faculty and students are not able to find all materials in their local or consortia libraries. The downward trajectory in the average number of holdings identifies a decrease in the number of titles research libraries are adding to WorldCat and possibly to their collections.

This analysis provided some data on the overlap at a country and publication year level. However, it did not confirm whether the adequacy of coverage as measured by the number of titles and average holdings is sufficient in relation to the number of titles published in a country or world region. Nor did the analysis attempt to determine whether records included in the snapshot represented core or supplementary material from a specific country.

This study raises a number of new questions: Should ARL libraries have more than an average of four holdings per title published by Caribbean countries? Conversely, is the average of six holdings per title published in Latin American countries more than sufficient to meet the needs of research libraries? Is the coverage of publications from individual countries sufficient? That is, are eight holdings per title published in Peru or Argentina adequate to meet local as well as North American needs? Are three holdings per title published in Greece too few? In addition, this study does not recommend whether there is sufficient redundancy within a country, world region, or publication year.

Additional Research
This investigation raises several areas that warrant additional analyses of the 2004 snapshot or new research studies. Examining the OCLC snapshot record-by-record to document the number of titles held by only one ARL library would result in a deeper understanding of the modest overlap by country, region, and publication date. However, examining holdings at the individual record level was out of the scope of this investigation.
Further analysis should be undertaken with comparable samples from OCLC-PICA (which includes extensive library holdings from Germany, the UK, and the Netherlands), the Deutsche Bibliothek, and OCLC WorldCat once the holdings from the RLG Union Catalog have been integrated. (On July 1, 2006, OCLC and RLG became a combined organization and currently are in the process of integrating the two union catalogs—the RLG Union Catalog and OCLC’s WorldCat.) Such studies would place the findings of this study into a more global context, confirm or negate some or all of the findings of this study, or suggest new areas for further analysis of this snapshot. A number of research libraries report holdings to both WorldCat and the RLG Union Catalog. RLG’s Union Catalog includes 47 million titles, an unknown number of which are also represented in WorldCat. However, a 2004 RLG analysis of the records in the RLG Union Catalog showed that nine countries have over one million title holders: United States, United Kingdom, Germany, France, Italy, Japan, Canada, Switzerland, and China and Hong Kong. The first five are also the top five in WorldCat, with the positions of Japan and Italy switched in WorldCat. Switzerland ranks as 17th in the WorldCat snapshot.

The findings also highlight several countries and regions that would benefit from more detailed analysis and possible targeted cooperative action by consortia or organizations such as the Center for Research Libraries or one or more of the GRN global projects.

Next Steps: Local and Collaborative
An institution-specific version of this analysis may aid a library in analyzing its collections and assessing how they relate to overall patterns of collecting, partnerships, and collaborations. A library might be able to compare area-specific collections with all ARL library titles in WorldCat and determine how to prioritize local processing and cataloging backlogs.

GRN may wish to engage the scholarly societies to react to the results of this study and to gather additional data (quantitative or impressionistic views) of areas not held in WorldCat by ARL libraries.
Appendix A: Background and Impetus for the Study

The AAU/ARL/CRL Global Resources Network (GRN) seeks to strengthen North American access to international and foreign language resources in support of research, teaching, and learning. GRN supports a series of discrete projects that focus on expanding the depth, breadth, quantity, range of formats, variety, and access to international information resources available to students and scholars. GRN also works to coordinate international acquisitions more effectively in order to expand overall coverage and reduce unnecessary duplication by making dispersed collections more interdependent and complementary. Finally, GRN encourages inter-institutional cooperation, non-duplication, and reliance on rapid and reliable interlibrary loan and document delivery services among libraries. The ARL membership has expressed differing perspectives on what combination of collection building and access strategies are needed to pursue and achieve this goal, and thus, the current global projects are testing different combinations. The global projects include:

- Collaborative Initiative for French and North American Libraries
- Cooperative African Newspapers Project
- Digital South Asia Library Project
- German-North American Resources Partnership
- Latin Americanist Research Resources Project
- Southeast Asia Indexing Project

In seeking to advance the concept of the distributed network of complementary collections, the GRN Advisory Committee identified the need to collect baseline data on collections and acquisitions. Necessary background data includes statistical assessments of duplication (overlaps) and also of materials that research libraries are not currently collecting (gaps). There was broad support for graphic representations of areas of particular weakness, and redundancy, using maps and GIS-based tools.

To advance the investigation, the GRN Advisory Committee identified a wide range of questions, but agreed to focus on just one or two. The initial questions included:

- How do we define the problem and need for global resources? What is the nature of foreign materials not being acquired by ARL libraries? What percentage of global publications is held in North America? How do these percentages break down by the regions of the world? What is the shape and nature of materials not represented in OCLC WorldCat? How does this information compare to previous studies? How do changes in teaching, research, and service in ARL/AAU institutions inform the approach to global resources in libraries? What topical areas will make the greatest difference in transforming global research and education at AAU institutions?

The GRN Advisory Committee sought and received the cooperation of the OCLC Office of Research, which provided a copy of the snapshot and prepared draft collection and collecting “maps.” The OCLC Office of Research’s Lynn Silipigni Connaway, Edward T. O’Neill, and Clifton Snyder collaborated with GRN Director, Eudora Loh, and ARL’s Mary Jackson on this investigation.
Although the original idea was to document gaps in collections, obtaining reliable publication data by country was extremely difficult, if not impossible. Identifying the quantity of publishing output from an individual country is impossible in countries that do not publish national bibliographies or regularly report statistics to UNESCO. Even for countries that publish national bibliographies or report the quantity of their publishing output to UNESCO, a simple comparison between the number of titles published and the number of titles held by ARL libraries might be misleading. For example, the numbers supplied to UNESCO might not include publications from scholarly societies, governmental and non-governmental organizations, dissertations and theses, or non-print resources that research libraries collect. Thus, this investigation narrowed its focus to the level of duplication and overlap in ARL libraries as represented by records and holdings in the OCLC WorldCat database.
Appendix B: The Data Set and Methodology Used

The OCLC Office of Research contributed two snapshots of the OCLC WorldCat database—one taken in January 2004 and the second in July 2004. The snapshots represented all WorldCat “book” records by country of publication and publication date, reported in five-year increments. In addition, the snapshot included holdings data of all OCLC and ARL libraries and a subset of ARL libraries.

The sample represents 8,597 OCLC governing member libraries that contributed cataloging to WorldCat during this time period. College and university libraries represent nearly one-third of that total.

The July 2004 snapshot of WorldCat included 29,375,454 bibliographic records for books with 725,756,130 holdings. In all analyses the investigators eliminated 64,294 records with publication dates of 2005 and beyond as these records represent items that were not actually held by an OCLC or ARL library as of July 2004. In addition, 878,638 records with unknown or multiple publication locations were eliminated, resulting in 28,432,888 usable records and 719,012,619 useable holdings. The 787,253 records with unknown publication dates included in the 28.4 million usable records were included in the “all years” analysis and also in the country of publication analyses.

It should also be kept in mind that the snapshot includes a small, but unknown, number of records that are publisher or vendor records, multiple records for the same item, and acquisition and pre-publication records. However, given the size of the sample, and based on other research by the OCLC Office of Research, it was determined that including these incorrect or duplicate records would not significantly skew the overall results. The OCLC Office of Research estimated that the duplicate records are likely to make up well under five percent of the sample, and that the rate of duplicate records is proportional to the number of holdings.

Although WorldCat is the world’s largest database of bibliographic records and holdings, it does not represent the universe of publication output for a particular country or year. The available data did not identify the proportion of total publishing output from a specific country or world region owned by ARL or OCLC libraries. As a result, the average number of holdings and percentages of total holdings may be incomplete or underrepresented for certain world regions or publication years. One example of the possible under-representation relates to books published in non-Roman scripts. Not all libraries (ARL and other OCLC members) have contributed all of their non-Roman script holdings to WorldCat.

In addition, not all libraries contribute cataloging at the same rate and thus the sample reflects local processing and cataloging practices. In an analysis of the 1995 and 1999 snapshots of the OCLC database, Jutta Reed-Scott identified several reasons for delays in adding records or holdings to WorldCat. These delays are a result of cumulative delays in acquisition, shipment, and processing foreign language materials, as well as backlogs in processing and cataloging. In addition, ARL libraries may prioritize their processing and cataloging based on local needs, thus creating varying delays in contributing records to WorldCat. Reed-Scott observed: “it takes more than five years for many foreign imprints to be added to the [OCLC] database.” Thus, the data reported for the 2000 to 2004 time period in this study should be interpreted with caution.

Data Elements Included in the Study

The publication date was pulled from either the date1 or date2 portion of the fixed field of each MARC record (008/7–10, 008/11–14 respectively). The date type field determined which of the two dates was more relevant to a particular record. If a record included multiple dates, the earliest one was chosen, as it was the closest date to the original publication date.
The country of publication was pulled from the fixed fields (008/15–17). The record was attributed to a given country if and only if all of the records associated with that work had the same country of publication. Otherwise, the record was put into the “multiple countries” category. Records with multiple or unknown countries were excluded from the country and world region analyses but included in the publication date analyses. A small number of records contained incorrect country codes in the fixed fields of the MARC record. In those cases, incorrect data for the country of publication were used, as it was impossible to review manually each of the 28.4 million records in the snapshot.

The number of holdings per record was derived from the number attached to book records in WorldCat. The holdings data include all holding symbols for an institution. For example, if the University of Texas, Austin has three holding symbols and each of the three holding symbols was attached to a record, the three holding symbols were counted as three holdings rather than one holding from one institution. Thus the average holdings may over-represent an institutional view of holdings. If those three holding symbols were aggregated into one institutional symbol, the average number of holdings per record in this study would be lower. (In other words, there is not a one to one correlation between a holdings symbol and an ARL member library.)

The average number of holdings per bibliographic record was calculated by dividing the number of records by the number of holdings. This average does not imply that all titles represented in OCLC have exactly the average number of holdings. However, in the aggregate, the average holdings capture overall collection strengths by country of publication and publication date.

The Analysis

OCLC and ARL staff undertook a detailed analysis of the more than 28.4 million usable records in the July 2004 snapshot. Countries were organized into 11 world regions and publication dates were sorted into three date ranges.

The aggregation of countries into 11 world regions enabled a focus on regional trends, rather than more variable publishing patterns within a single country. The 11 regions included: Africa (58 countries), Caribbean (26), East Asia (6), Eastern Europe (27), Latin America (24), Middle East (19), North America (3), Oceania (34), South Asia (8), Southeast Asia (13), and Western Europe (27). These 11 regions are similar, but not identical, to the ten regions used in Reed-Scott’s 1996 Scholarship, Research Libraries, and Global Publishing (Africa, Australia, East Asia, Eastern Europe, Latin America, Middle East, South Asia, South East Asia, Western Europe, and Other).

Records were aggregated into three date ranges: all years (pre-1600–2004, including records with unknown publication dates); 1955–1979; and 1980–2004.

The investigators also tested the sample using the Functional Requirements for Bibliographic Records (FRBR) to bring together similar works and minimize “duplicate” records from the sample. However, the FRBR version of the sample was not used because of the importance of specific manifestations (i.e., those specified by language, edition, country of publication, etc.) for the collections of research libraries.
Appendix C: Countries by World Regions

Sub-Saharan Africa
Angola, Benin, Botswana, Bouvet Island, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo (Brazzaville), Congo (Democratic Republic), Cote d’Ivoire, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, South Africa, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe

North Africa
Algeria, Djibouti, Egypt, Eritrea, Ethiopia, Libya, Morocco, Somalia, Spanish North Africa, Sudan, Tunisia, Western Sahara

South Asia
Bangladesh, Bhutan, British Indian Ocean Territory, India, Maldives, Nepal, Pakistan, Sri Lanka

Southeast Asia
Brunei, Burma, Cambodia, Christmas Island (Indian Ocean), East Timor, Indonesia, Laos, Malaysia, Philippine, Singapore, Spratly Island, Thailand, Vietnam

East Asia
China, China (Republic: 1949- ), Japan, Korea (North), Korea (South), Mongolia

Oceania
American Samoa, Antarctica, Australia, Cocos (Keeling) Islands, Cook Islands, Fiji, French Polynesia, Guam, Heard and McDonald Islands, Johnston Atoll, Kiribati, Marshall Islands, Micronesia (Federated States), Midway Islands, Nauru, New Caledonia, New Zealand, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Paracel Islands, Pitcairn Island, Samoa, Solomon Islands, Terres australes et antarctiques, Tokelau, Tonga, Tuvalu, U.S. Misc. Pacific Islands, Vanuatu, Wake Island, Wallis and Futuna

Latin America
Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Falkland Islands, French Guiana, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, South Georgia and the South Sandwich Islands, Surinam, Uruguay, Venezuela

Caribbean
Anguilla, Antigua and Barbuda, Aruba, Bahamas, Barbados, Bermuda Islands, British Virgin Islands, Cayman Islands, Cuba, Dominica, Grenada, Guadeloupe, Haiti, Jamaica, Martinique, Montserrat, Netherlands Antilles, Puerto Rico, Saint Helena, Saint Kitts-Nevis, Saint Lucia, Saint Vincent and the Grenadines, Trinidad and Tobago, Turks and Caicos Islands, United States Misc. Caribbean Islands, Virgin Islands of the United States
Western Europe
Andorra, Austria, Belgium, Denmark, Faroe Islands, Finland, France, Germany, Gibraltar, Greece, Greenland, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Monaco, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, United Kingdom, Vatican City

Eastern Europe & Eurasia
Albania, Armenia (Republic), Azerbaijan, Belarus, Bosnia and Hercegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia (Republic), Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Poland, Romania, Russia (Federation), Serbia and Montenegro, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan

Middle East
Bahrain, Cyprus, Gaza Strip, Iran, Iraq, Iraq-Saudi Arabia Neutral Zone, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, West Bank of the Jordan River, Yemen

North America
Canada, Saint Pierre and Miquelon, United States
Appendix D: Average ARL Member Library Holdings by Selected World Regions

D1: East Asia

D2: Eastern Europe
D3: Latin America

D4: Western Europe