

Chapter Seven

Summary and Conclusions

In February 2001, the Boston Public Library entered the 46 millionth record into WorldCat.¹ In that same month the 10% systematic random sample of monographic bibliographic records for this study was extracted from WorldCat. This study presents a snapshot in time of the WorldCat database as it existed in February 2001. The study is hindsight or *ex post facto*, in that we don't know what the profile looked like in earlier time frames. A record is added to WorldCat every 15 minutes, so that literally, a database scan is only accurate for a few minutes. We are viewing the universe of publication as it was reflected in WorldCat early in 2001.

The tables in this study were all constructed for the report from the iCAS analysis of the 10% systematic random sample of monographic bibliographic records from WorldCat extracted in February, 2001. The subject divisions, categories, and most detailed subject descriptors follow the construction of the iCAS product that is based upon the WLN Conspectus subject breakdowns. The unique title and title overlap, language, and the adult/juvenile audience analyses follow the iCAS product templates. The ISBN analysis is not a regular feature of iCAS, but was run especially for this study. The major findings from the study are summarized below by chapter.

Findings

WorldCat (Chapter Two) <http://www.oclc.org/dummyaddress.html>

\$ The 10% systematic random sample of monographic bibliographic records in WorldCat contained 3,378,272 usable records.

¹. "The Year in Review, July 1-June 30, 2001," *OCLC Annual Report, 2000/2001*, p.6. (The 46 million is not the actual total of records that is less, at 45.1 million when duplicate records are removed.)

\$ 2,199,165 of those records had call numbers and could be analyzed by subject.

\$ An additional 1,179,107 records did not have call numbers (shown as ANo call number present@ in the tables).

\$ A pattern of decline in the number of records annually beginning in the early 1990s is evident The number of records with call numbers declines 72 percentage points between 1992 and 2000, while the number of records without call numbers declines 88 percentage points in the same time frame.

\$ Subject analysis on the set of records with call numbers shows that the language, linguistics, literature division has the largest number of records comprising 21% of all records with call numbers. The history division is second with 13.3%. The business and economics division comprises 9.6%; philosophy/religion has 8.2% and engineering/technology comprises 6.8% of all records with call numbers.

Library Groupings (Chapter Three) <http://www.oclc.org/dummyaddress.html>

\$ The research libraries grouping has the largest number of records with call numbers in the dataset, 1,745,034.

\$ The profile of the research libraries by imprint year and subject divisions closely parallels the profile of WorldCat.

\$ The decline in the number of records annually in the 1990s is most acute in the research libraries aggregated collections.

\$ The total number of records with call numbers for the academic libraries grouping is 1,323,165.

\$ The patterns of increase/decrease over time in records for the academic libraries are similar to those of the research libraries.

\$ The major differences between the research libraries and the academic libraries are in the records without call numbers.

\$ The concentration of records by subject division in the academic libraries is the same for LLL and history, but the third and fourth ranked subject divisions are reversed in the academic library collections from the research library collections. In academic libraries philosophy/religion is ranked third, but it is fourth in the research libraries. In the academic libraries business/economics is fourth, whereas it is third in the research libraries.

\$ On a percentage basis, the decline in the number of records annually for the academic libraries is less severe than in the research libraries, but follows the same pattern.

\$ In the special/other libraries grouping, the pattern of decline in the number of records does not begin until 1993 and it is less severe than in the research and academic libraries.

\$ The subject concentrations in the special libraries are not as similar to the research and academic libraries by percentage of total for subject divisions. There are differences that can easily be attributed to the nature of the professional and discipline emphases in the special libraries. The large number of corporate, legal, medical, and technical libraries account for higher percentages of total in business, legal, and scientific subject divisions.

\$ The bibliographic records for the public and school library groupings were analyzed by adult and juvenile audience levels, a feature of the iCAS analysis. The public libraries, with adult collections comprising over 90% of total, have a ratio of 21 to one, adult to juvenile. For school libraries the ratio of adult to juvenile is 5 to one.

\$ In the school library grouping the percentage of juvenile titles steadily increases from 1985 to present, the only measure to show a pattern of increase in the number of records in the latter 1990s.

Unique Records (Chapter Four) <http://www.oclc.org/dummyaddress.html>

- \$ The data analyses on number of unique records and title overlap demonstrate that the universe of materials under bibliographic control in WorldCat show a high level of diversity of resources with 53% of records in the analysis having only one library location symbol.**
- \$ The unique records in WorldCat are almost evenly divided between records with call numbers and records without call numbers.**
- \$ Both WorldCat and the research libraries have close to the same absolute numbers in the datasets for unique records, both those with call numbers and without call numbers.**
- \$ The research libraries have 63.51% of total records as unique records.**
- \$ The research libraries grouping contributes 62% of the unique records in the sample database, while the other four library groupings contribute the remaining 38 percent.**
- \$ The other four types of library groupings have low percentages of unique titles and higher overlap among the four.**
- \$ The academic library records naturally have a high overlap with the research libraries.**
- \$ The core set of records (101,378) shared by all of the library groupings (overlap) represents only 3% of the total records in the sample. These represent 5% of the records used in the subject analysis of which the 101,378 records is a subset.**
- \$ The number of records in common among the five types of library groupings (101,378) is governed by the school libraries grouping in that it has the smallest number of total records and thus the overlap cannot exceed the number of records the school libraries have in common with the other four groupings.**

Language analysis (Chapter Five) <http://www.oclc.org/dummyaddress.html>

- \$ Of all records in the study, approximately 65% are for English language monographs.
- \$ Of the total number of records with call numbers, 63.6% are for English language materials and 36.3% are for non-English titles.
- \$ The six foreign language groupings analyzed in the study B Chinese, French, German, Japanese, Russian, and Spanish B account for 67.5% of all foreign language records in the sample.
- \$ In the analysis by subject divisions, the language, linguistics, and literature division has the largest number of records in all language groupings, followed by the history division. After these two subject divisions, the ranking of subjects varies by language grouping.

ISBN Analysis (Chapter Six) <http://www.oclc.org/dummyaddress.html>

- \$ Only 27% of all records with call numbers in the sample have ISBN numbers.
- \$ Fifty-seven percent of all records with call numbers in the 1970-1999 time frame have ISBN numbers.
- \$ Less than 10% of all records without call numbers in WorldCat have ISBN numbers.
- \$ For English language records, 92% of the records with ISBN numbers also have call numbers.
- \$ For foreign language records, 72% with ISBN numbers also have call numbers.
- \$ The number of foreign language records with ISBN numbers but without call numbers (71,240) is larger than the number of English language ISBN records without call numbers (32,824).
- \$ Sixty-eight percent of ISBN records without call numbers are foreign language

records.

§ It appears that titles from mainstream publishers are being consistently added to WorldCat within a current time frame, and the decreases in numbers of records in the latter 1990s are not in the English language titles with ISBN numbers.

§ The decline in the foreign language ISBN records is responsible for the overall decrease in records with ISBN numbers in WorldCat, as the English language records have slight decreases in 1997 and 1998, with an actual 2% gain in 1999.

§ Records with ISBN numbers have a lower rate of decrease than records without ISBN numbers and the rate of decrease in the late 1990s is lower for those records with ISBN numbers. This supports the speculation that the ANo call number present@ records are largely comprised of foreign language titles and titles which are not from mainstream U.S. and European publishers.

§ In subject areas in which research libraries collect globally, the differential between the English language and foreign language totals reflects the foreign language collecting in those libraries. Both LLL and the history division have a considerable number of foreign language records with ISBN numbers.

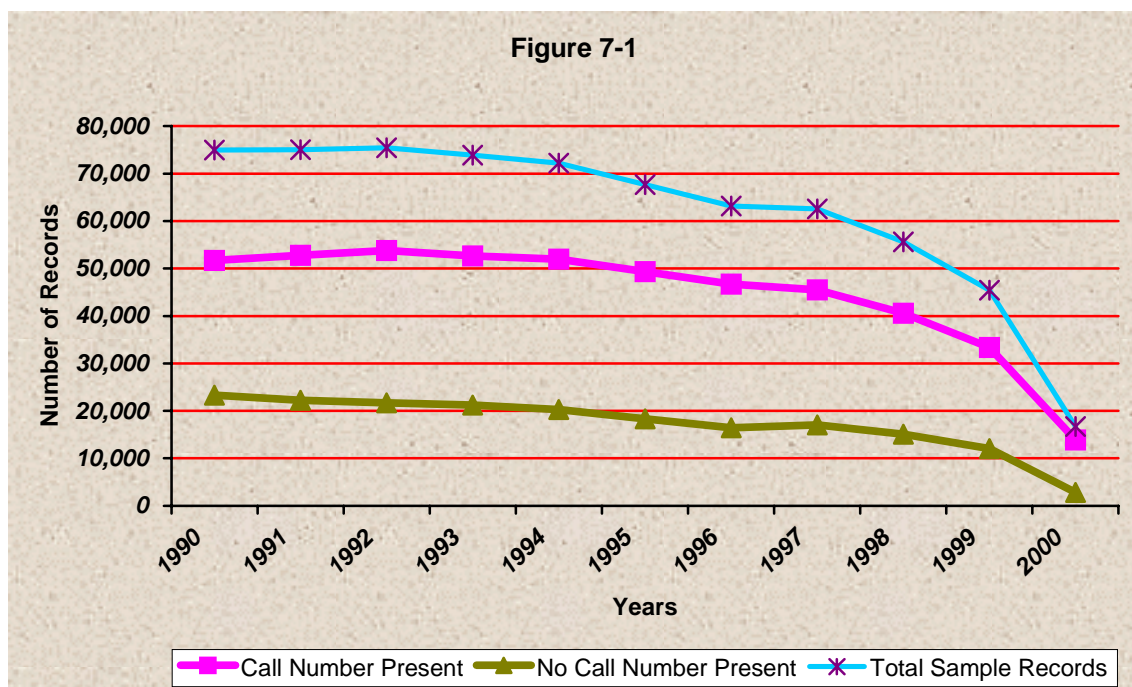
One of the most troubling of the findings is that of the decline in the number of records in the most current years of the analysis. This aspect of the findings is examined more closely in the next section.

Decline in the Number of Records for Current Years

The findings by several measures show a decline in the number of records for the most current imprint years in the study. Beginning before 1995, in most instances, the absolute number of records decreases steadily and annually for

unique records, language records and certain categories of ISBN records. Tables 7-1 through 7-6 trace the decline by these different measures for the imprint years 1990-2000. Figures 7-1 through 7-6 are derived from these tables.

In Table 7-1 and Figure 7-1 the *Annual Increase/Decrease in Number of WorldCat Records 1990-2000* is shown. The rate of change is shown for both those records with call numbers and those records without call numbers and the two combined as total records. It can easily be seen that the number of records without call numbers is considerably less than the number of records with call numbers. But the decrease is much higher for the records without call numbers, causing the overall decrease to be higher than the rate of change for the records with call numbers.



The decline also begins to occur earlier for the records without call numbers. The peak year for total records and those with call numbers is 1992, after which the rate of change is negative. Oddly, there is a reversal in 1997 of the decline in records without call numbers, but the pattern reverts back to a decrease in the following year. The absolute number difference between 1997 and 1996 for the no call number records is only 600 records. In that same year, 1997, the records with call numbers also have less of a decrease than the preceding year. The total number of records is nearly the same from 1996 to 1997. We can only speculate as to the reason(s) for 1997 interrupting the downward pattern. 1997 is four years removed from the year of extraction of the sample. It is possible that it represents the point at which acquisitions and cataloging lag actually commence, whereas those years prior to 1997, the numbers of records are a closer reflection of true collecting patterns.

The downward trend is most severe for all three categories beginning in 1995. While the records without call numbers have the steepest percentage decline, as seen in Table 7-1, the lower number of those records shows as a flatter curve in Figure 7-1. It will be seen in the remaining tables and graphs that the other measures show similar downward trends. As Table 7-1 and Figure 7-1 show total records for WorldCat, the other measures have to follow the overall trend.

Table 7-2 and Figure 7-2, *Annual Increase/Decrease in Number of Records 1990-2000 Academic and Research Libraries*, show the totals for the two largest library groupings. From Table 7-2 and figure 7-2 it can be seen that the decline in records for both the academic and research libraries groupings follows the same pattern as the WorldCat dataset. The pattern of increase/decrease annually is nearly the same for both of these library groupings. The only difference is in the number of records for the two groups, with the research libraries having the higher total and the academic libraries following the same pattern and keeping the same differential between totals for the time period.

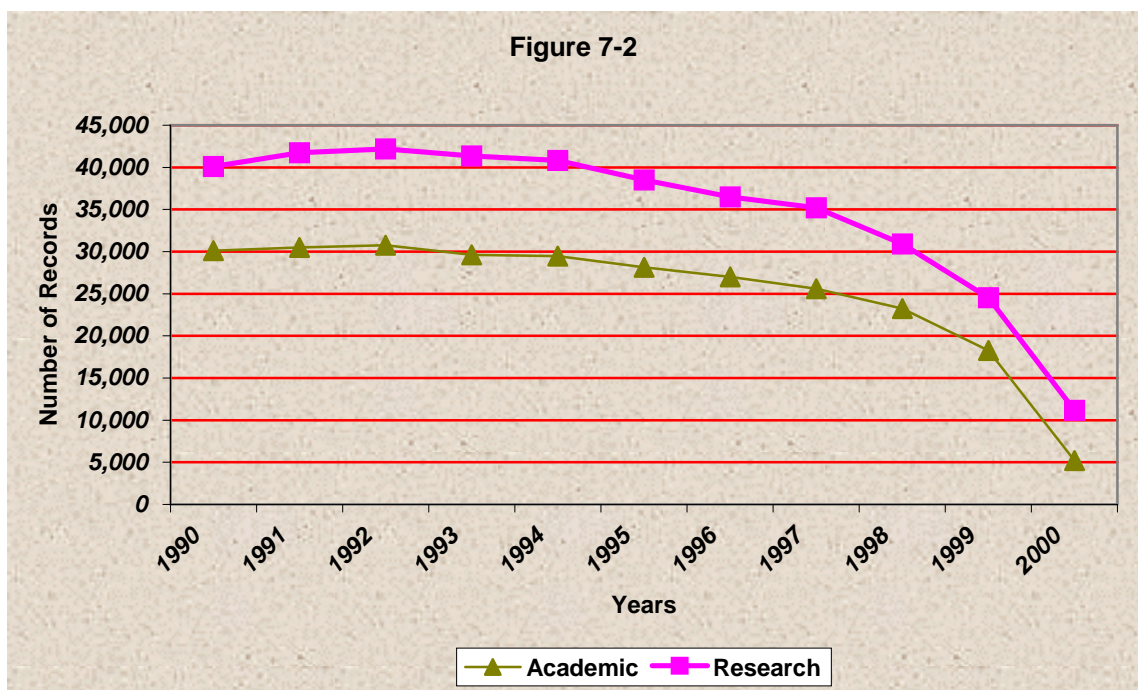
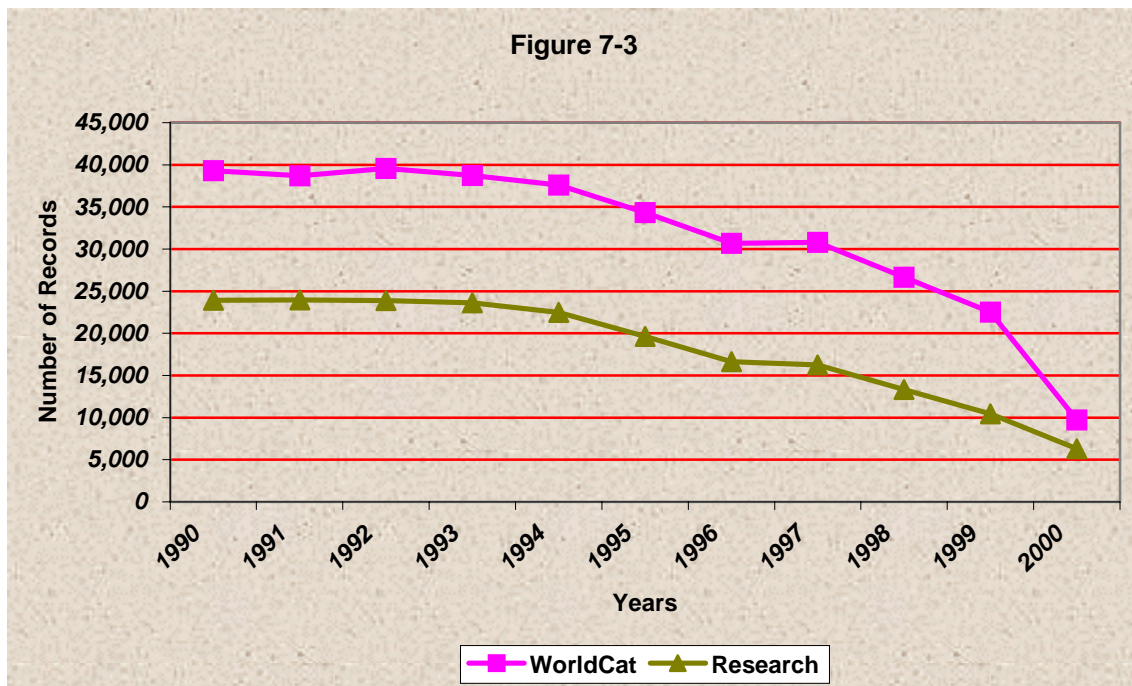


Table 7-3 and Figure 7-3 *WorldCat and Research Libraries Unique Records 1990-2000* show the pattern of decline in the number of unique records for both WorldCat and the research libraries for the 1990-2000 imprint years. For the majority of the years, the percentage decline in WorldCat is higher than the decline in the research libraries= records. Again, 1997 stands out in the unique records measure as an anomaly in the pattern for the decade. The number of unique records in WorldCat remains static from 1996-1997, but there is a slight decrease in the number of unique records in the research libraries. Both have considerable drops in the number of records from 1995 to 1996 and the decline continues in 1998. The other types of library groupings are not similarly analyzed here as the unique records analysis was performed only on WorldCat and the research libraries.



The series of tables 7-4a, 7-4b, and 7-4c and accompanying figures show the number of records from 1990-2000 imprint years for English language records and non-English language records. These are from the dataset of those records with call numbers used for the language analyses. In 7-4a and Figure 7-4a, *WorldCat English and Non-English Language Records 1990-2000*, it can be seen that the decline in the number of English language records and the total number of records begins in 1993, whereas the decline in non-English records commences in 1995. It is evident in table 7-4a that the decline is much steeper in the non-English language records, again a factor of the overall lower absolute numbers of those records.

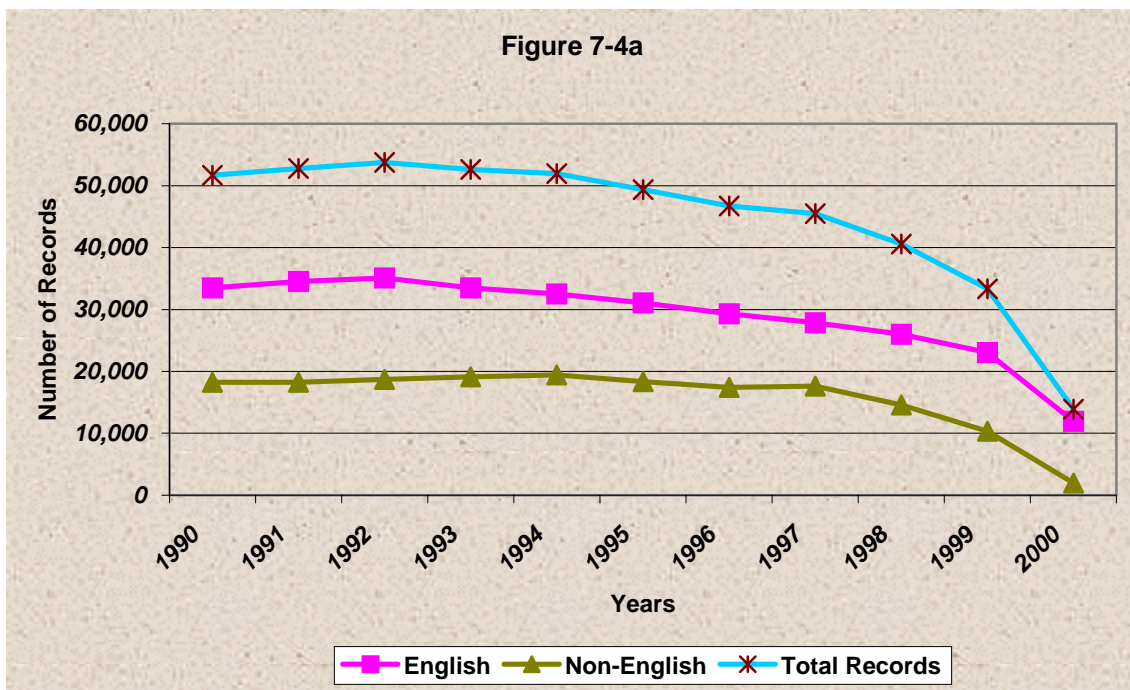


Table 7-4b and Figure 7-4b, *Research Libraries English and Non-English Language Records 1990-2000*, show the decline in English and non-English records for the research libraries. It is more obvious from figure 7-4b that the decline in non-English language records is much steeper than in English language records. For the years 1996 and 1997, the absolute numbers of records in the two language categories draw closer together. From 1995-1998, the number of English language records maintains a steady and modest decline around 5 percent. The number of non-English records plunges, dragging down the total number of records after 1997.

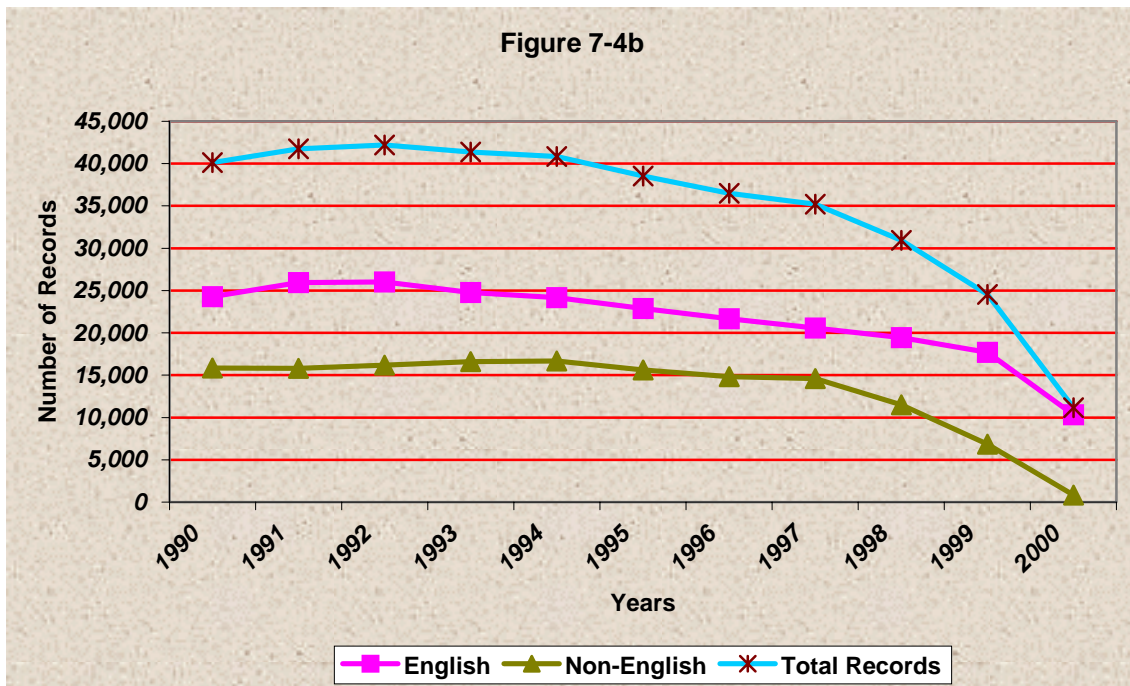
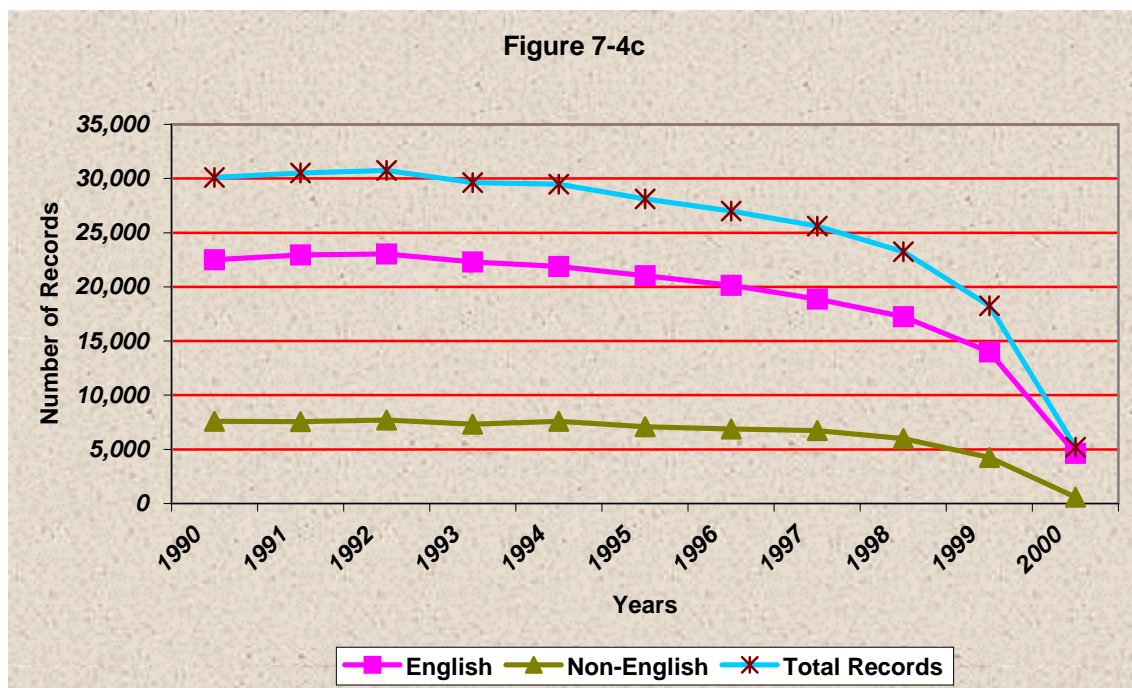


Table 7-4c and Figure 7-4c, *Academic Libraries English and Non-English Records 1990-2000* show the data for the same measures for the academic libraries. Figure 7-4c makes it obvious how low a number of non-English language records are owned by the academic libraries, and this grouping includes a considerable number of non-U.S. libraries. The trend line of non-English records is relatively flat until 1998. The international academic library membership of OCLC may be responsible for a relatively constant rate of acquisition and cataloging in foreign language records. Because of the predominance of records for English language publications in the academic libraries, the trend lines for the total and English language records have an identical pattern. The scale for Figure 7-4c is 10,000 records lower than in 7-4b. It was observed in chapter three that there is a differential of approximately 10,000 records between the research libraries and the academic libraries groupings in total records.



The last set of tables and figures depict the ISBN analysis from chapter seven. Table 7-5 and Figure 7-5 *WorldCat Records with ISBN Numbers*, shows the three ISBN analyses that were run for WorldCat. In the ISBN analyses, the downward trend seen in the other measures commences only a few years from the time of data extraction in 2001. The number of records with ISBN numbers and call numbers peak in 1996; the number of ISBN records lacking call numbers peak in 1997. The number of records with ISBN numbers but without call numbers is very low in comparison to those with call numbers which set the pattern for the total. The decline in these measures is much less severe than in the other measures. This stands to reason in that imprints with ISBN numbers are typically from mainstream publishers and large conglomerates internationally, and these would be titles which are added by libraries on a more current basis than more unique and esoteric titles which do not as readily come to attention and are not cataloged as close to the time of acquisition.

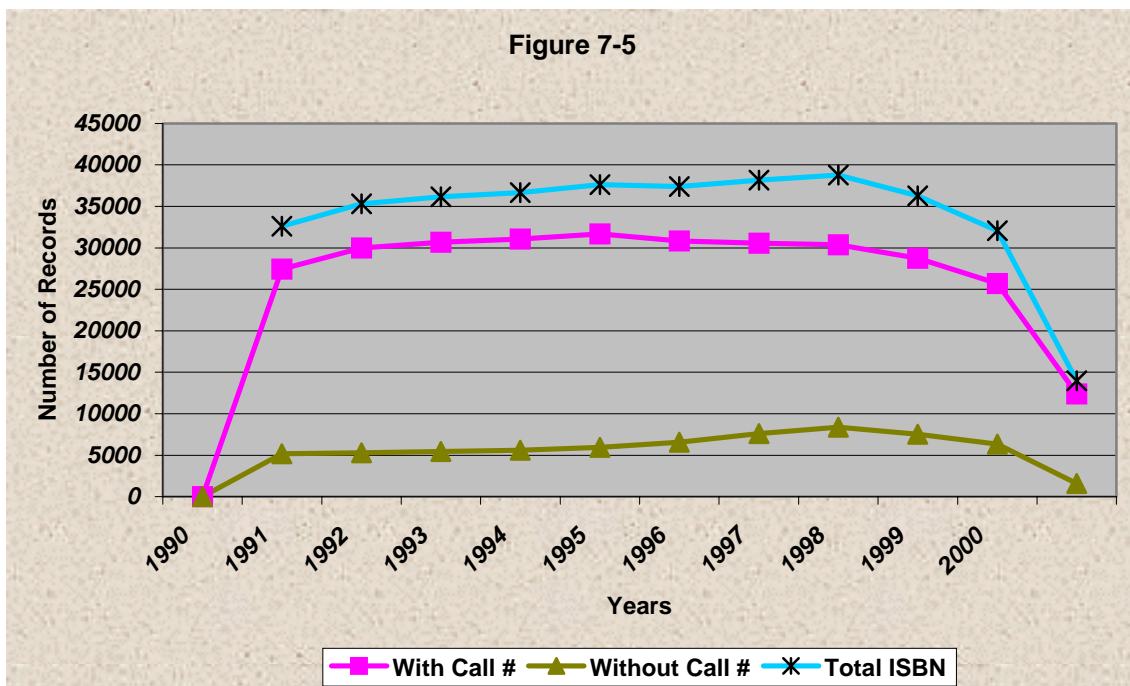
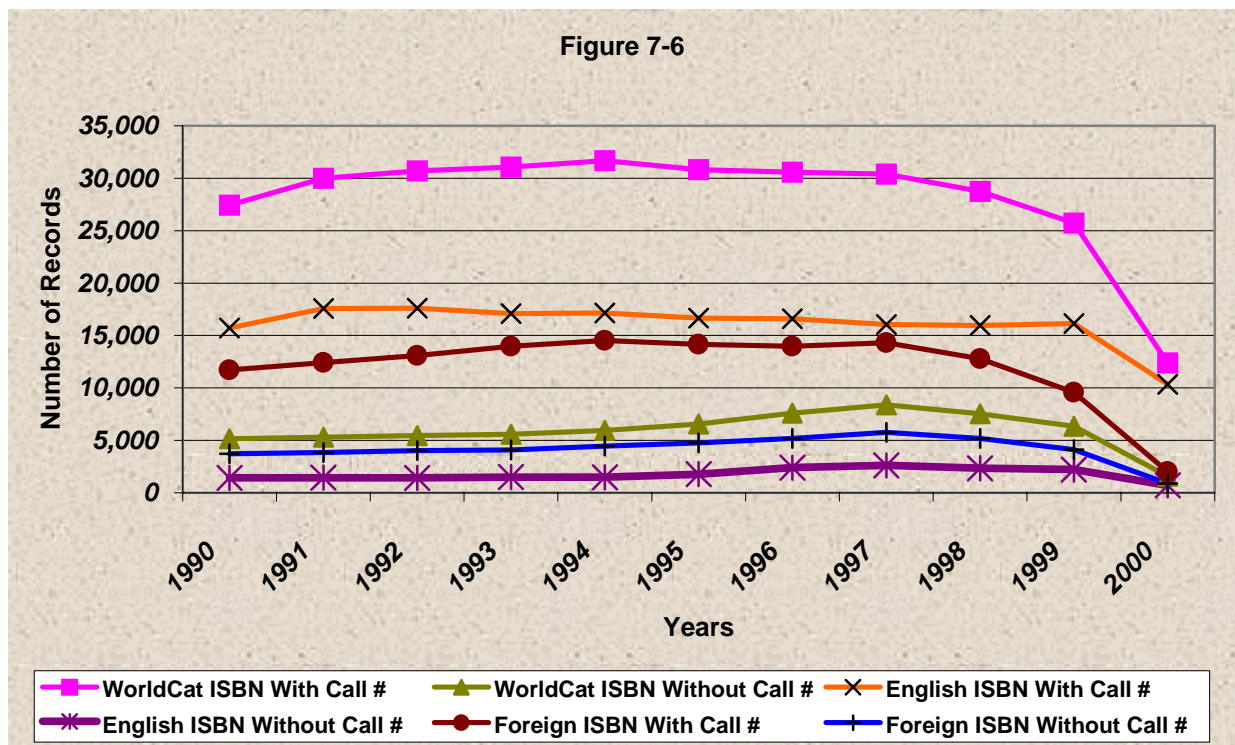


Table 7-6 and Figure 7-6 *WorldCat, English, and Foreign Language Records with ISBN Numbers*, show the ISBN analysis for WorldCat by both call number presence and language categories for the decade of the 1990s. The category with the lowest number of records with ISBN numbers is that of English language records without call numbers. As English language publications are those most likely to have ISBN numbers and also most likely to be owned by U.S. and libraries in other English speaking countries, it follows that most English language titles with ISBN numbers would most likely be cataloged and have call numbers.



The next to lowest category in the ISBN analysis is that of foreign (non-English) titles without call numbers. From the two lowest categories it can be seen that records without call numbers comprise the lowest number of records with ISBN numbers. This points to the type of materials which are not from mainstream publishers and tend to be local history and other titles of local interest, juvenile materials not classified, theses and dissertations, and reports of all kinds. The third line from the bottom in Figure 7-6 is the summation or total of the two bottom lines, those records in WorldCat with an ISBN number but not a call number.

The top three trend lines in Figure 7-6 are for records with both ISBN numbers and call numbers. The lowest of the three is foreign language records with call numbers, which are less numerous than English language records with call numbers. It can be seen that these two categories do not have exactly the same pattern for the imprint years shown. The English language records with call numbers begin to trend downward in 1992, but the decline is slight and the number of records in 1990 and 1998 are very close to the same number. The number of English language records with ISBN numbers and call numbers is actually higher in 1999 than in 1998. Whereas the foreign language records trend upward and remain in the same range until after 1997 but then show the same steep decline as other measures have shown. From Table 7-6 and Figure 7-6 it can be seen that English language records with both call numbers and ISBN numbers are added to the database in a more timely manner than the other categories of records analyzed. The decline is severe in foreign language records with call number and all records without call numbers. What this means is that mainstream publications from English speaking countries are the only category of materials added to the database in a timely manner. Those materials that are more unique and contribute to variety within the resources base are added in a slower manner over a longer period of time.

The implications of these findings and further interpretation are addressed in the next section.

Implications

The findings of this report have direct implications for resources sharing and coordinated cooperative collection development, both of which depend upon the availability of database bibliographic records. The profile of WorldCat revealed in the data analyses in this report provides a basis for discussions on the role of WorldCat in these activities.

OCLC WorldCat has facilitated resources sharing in a number of ways, first and foremost by developing a database of bibliographic records for shared cataloging with holdings information which forms the basis for resources sharing. The cataloging records are augmented by union lists of serials and newspapers with detailed holding information for volumes and issues. The interlibrary loan module of WorldCat revolutionized the procedural nature of resources sharing which went from a system of detective work and guessing to a certainty of which libraries hold a desired title.

The effect of WorldCat upon coordinated cooperative collection development has not been of the same magnitude as the interlibrary loan system upon resources sharing. Now it is more a matter of what capabilities of the database and existing products can be taken advantage of for cooperative collection development activities. ACAS (Automated Collection Assessment and Analysis Service) is a set of collection analysis and assessments tools developed by WLN to assist libraries and consortia in analyzing collections for resources sharing and CCD. The iCAS product from which the data analysis in this study was produced is one module of ACAS. ICAS has been used in consortia in the United States since its inception and it is now being utilized internationally. As an example, CURL, the Consortium of University Research Libraries in the British Isles, began a collection analysis project using the iCAS in June 2001.² The ACAS collection analysis products can assist in

²*OCLC Annual Report, 2000/2001*, p.15.

formulating cooperative collection development arrangements and in providing the data necessary for the implementation and functioning of such agreements. Both the ACAS products and the WorldCat database have the potential to augment CCCD activities and to play a larger role in the measurement of cooperative activity.

The extent to which the bibliographic records in WorldCat reflect the global universe of publication has a bearing on the support for resources sharing and CCCD. In considering the title of this report, *Global Collective Resources*, to what extent does WorldCat reflect the global universe of publication? While this question cannot be completely answered with the data in this study, it is easy to see that the profile of WorldCat reflects the collections of the libraries that contribute the records to the database. The farther back we go in time, it would seem, the more accurately WorldCat reflects the universe of publication as libraries have contributed records for retrospective materials. As we have seen from the previous section examining the annual decline in records in the last decade of the 20th century, certainly for that time frame the universe of publication is not yet represented in WorldCat.

What the findings for the 1990s do show is that as the international membership of OCLC continues to grow, the database will come ever closer to reflecting the universe of publication worldwide. The addition of many foreign language records in the 1990s are the result of the addition of bibliographic records from research and academic libraries internationally, which joined WorldCat.

Another aspect the data from the sample show is that the variety of resources in WorldCat is due, in large part, to the unique records for titles which are not from Amainstream® publishers. While the records for these titles may present some problems in analysis, i.e. the large number of records without call numbers that cannot be analyzed by subject, these records are present in the database and identify items, which may not be identified through any other venue, or at least any other electronic database. As with the study sample, and indeed as the study sample analysis has shown, the most unique titles appear to be added to the WorldCat

database much later and over an extended period of time as compared to current mainstream publications with ISBN numbers, both English and foreign language titles. Many of these unique titles are only discovered then acquired and cataloged over an extended period of time. In order to be most effective for acquisitions and resources sharing, bibliographic records for current, new publications need to be added to WorldCat within the time frame of publication. The lower number of records in the study sample at the time of data extraction shows what was available through the database at the time of data extraction. For purposes of any current activity, it does not matter that records will be added later and they do, indeed enrich the database. It is doubtful that most CCCD agreements will focus just upon the most common titles with ISBN numbers which are the predominant type of records appearing within the most current years of imprint records.

Agreements were made with international book vendors in the 1990s to contribute their bibliographic records to WorldCat. These book vendor records would not have been included in the study sample because there are no library holding symbols attached to those records. And at least some of them are brief cataloging and many are pre-publication cataloging. These records do provide bibliographic information for the most current publications available through those vendors, including many foreign language titles. Although these records are mostly for mainstream publications with ISBN numbers and do not assist in resources sharing, they are useful information for acquisitions and CCCD. And these records contribute to the universe of publication visible through WorldCat.

One of the steps taken in the OCLC effort to globalize the cooperative, was the establishment of a jointly owned organization, Pica B.V. to provide cataloging, interlibrary loan, local system and end-user services to libraries in the Netherlands, France and Germany. Pica B.V. is developing a European Central Catalog that will contain bibliographic and holdings information from many European academic libraries and other important libraries in Europe. What is not reflected in the sample

for this study is the large number of European libraries from Germany, France, and the Netherlands in Pica B.V., which are now considered to be members of OCLC, but have their records in the Pica B.V. database that is separate from WorldCat.³ At some point in the future these two databases will most probably be searchable as one and will truly reflect the universe of publication for North America and Western Europe.

WorldCat does represent an international universe of publications under bibliographic control with records for publications that are available for resources sharing and assistance in cooperative collection development activities. It seems certain that WorldCat will become more and more representative of the worldwide universe of publication as time goes on. It is doubtful, however, with a critical mass of over 45,000,000 records that the profile of WorldCat by imprint year and subject analysis for printed materials will change substantially for the foreseeable future. The snapshot in time, which is analyzed in this study, will form a picture of the database that will only gradually change, with the exception of the last decade of records.

These implications and questions can be explored in further research.

Further Research

The profile of WorldCat described in this report is the most detailed study to date of the monographic bibliographic records in WorldCat by multiple parameters of imprint date, subject, language, unique titles/overlap, and ISBN data. And yet, in some respects, the analysis is still very broad. The data analyses from the 10% random sample have not yet been mined out.[@] These suggestions are made for further research.

³ *OCLC Annual Report, 1999/2000*, p.30.

- \$ Subject analysis on the 10% systematic random sample thus far has been only on the iCAS broad 24 subject divisions. More detailed subject analysis should be conducted to look for more specific collecting patterns within the 24 subject divisions.**
- \$ Further analysis of the records without call numbers in the sample might assist in formulating cooperative collection development agreements.**
- \$ Further research is needed into using key word and subject headings to perform subject analysis instead of using only call numbers.**
- \$ Further exploration of the differences in profile between the set of U.S. libraries in the OCLC/CACD and the WorldCat sample which represents the entire OCLC international membership of contributing libraries may show the variety of resources reflected in WorldCat not owned by U.S. libraries.**
- \$ Comparisons could also be performed between the WorldCat analyses and those of large consortia using the iCAS product such as CURL in Great Britain, and several state consortia in the United States.**
- \$ An interesting study would be a comparison of the profiles of WorldCat and the Pica.B.V. European Central Catalog.**

Conclusion

The profile formed by the findings of this study captures the nature of the bibliographic contents of WorldCat at the end of the 20th century when the printed book was the predominant vehicle for the dissemination of knowledge and the preferred format for educational and recreational reading. The point of critical mass seems to have been reached in the first 30 years of WorldCat-s existence, in that from the beginnings with a few bibliographic records from Ohio libraries, it has

grown to over 45,000,000 million records. In June 2001, the total number of records for books was at 37,787,280 records, comprising 84% of records in the WorldCat database.⁴ Additions at this point will be for new publications and filling in with retrospective and foreign language records still not in the database. Undoubtedly, the profile will slowly change over time reflecting the broadening global membership of OCLC. There almost certainly will be an increase in records for foreign language materials. Libraries will continue to add records for retrospective materials and records for formats other than for printed monographs -- for web sites and electronic publications.

As the hegemony of the printed book wanes in the electronic environment, the 20th century will probably be regarded as the apex of print. As the WorldCat database continues to grow in retrospective records, it will increasingly reflect the profile of the universe of publication for the history of printed works. But the question remains as to the efficacy of the database in studying current library acquisitions patterns and as a basis for acquisitions and coordinated cooperative collection development. These concerns can only be addressed by the membership of OCLC.

⁴*OCLC Annual Report, 2000/2001*, p.18.