

The Proportion of NUC Pre-56 Titles Represented in the RLIN and OCLC Databases Compared: a Follow-up to the Beall/Kafadar Study

Christine DeZelar-Tiedman
Archives and Special Collections Catalog Librarian
University of Minnesota Libraries
160 Wilson Library
309 19th Ave S
Minneapolis, MN 55455

(612) 625-0381 (phone)
(612) 625-3428 (FAX)
dezel002@umn.edu

The Proportion of NUC Pre-56 Titles Represented in the RLIN and OCLC Databases Compared: a Follow-up to the Beall/Kafadar Study

Abstract

This article replicates a previous study that investigated the proportion of bibliographic records from the *National Union Catalog Pre-1956 Imprints* in the OCLC WorldCat database, and expands it to search a similarly-sized sample of records in the RLIN database as well. The author seeks to determine the impact that the merger of the RLIN and OCLC databases will have on the ability to locate catalog records for older materials, and whether there are still significant numbers of library materials for which there are no online bibliographic records. Entries for non-Roman language materials were not included in the study.

Introduction

In a previous article, Jeffrey Beall and Karen Kafadar investigated the proportion of bibliographic records from the *National Union Catalog Pre-1956 Imprints* (popularly called *Mansell*) represented in the OCLC Online Computer Library Center (OCLC) WorldCat database.¹ Based on a sample of 508 records from 26 different volumes of *Mansell*, Beall and Kafadar discovered that 27.8% of the titles from the sample were not found in OCLC WorldCat.

In May of 2006, OCLC and the Research Libraries Group (RLG) announced that they would combine operations, and that bibliographic records from RLIN, the RLG Union Catalog, would be integrated into WorldCat. Though RLIN is a smaller database than WorldCat, there is not complete duplication between the two bibliographic utilities. Therefore, it is possible that a proportion of the *Mansell* records not already in OCLC might be found in RLIN. If this is the case, then those records will become part of the OCLC database once the records from the two utilities have been merged. The purpose of the present study is threefold:

1. To confirm the validity of the Beall and Kafadar study by replicating the methodology and searching a similarly sized, but different, sample of titles in OCLC WorldCat.
2. To search the same set of titles in RLIN to determine the proportion of pre-1956 records not previously in WorldCat that will become more widely accessible once the integration of RLIN into WorldCat is complete.
3. To see whether the addition of the RLIN records into WorldCat has a significant impact on the percentage of records still unavailable in electronic form.

Literature Review

Other than the Beall and Kafadar study, the author could find only one other article comparing records found in *Mansell* with those in an online bibliographic database. In a 1984 article, Williams compared a sample of pre-1950 serial records in OCLC with records in the print resources *Mansell* and the *Union List of Serials (ULS)*.² Of his sample, 60 percent of the records in OCLC also had records in both print resources, 36 percent had records in either *Mansell* or *ULS*, and only four percent appeared in neither.³ He also found that the completeness of information provided for an individual title in the different resources varied: the print volumes tended to have fuller institutional holdings information, but many OCLC records contained more up-to-date data such as the International Standard Serial Number (ISSN).⁴

A number of studies have compared hit rates between OCLC and RLIN, as well as hit rates for now defunct bibliographic databases such as the Western Library Network (WLN). Most, however, date from the first decade of online cataloging, before most libraries had completed their retrospective conversion projects, and therefore would provide little in the way of reliable data to predict the numbers of records one would currently expect to find.

Several articles in the 1990s explored the availability of copy in the bibliographic utilities for foreign language materials. LeBlanc searched current French and Italian materials not found in RLIN, his institution's primary source of catalog records, in OCLC and in six large university library online catalogs.⁵ OCLC was shown to be a strong source for records lacking in RLIN, and a substantial majority of titles had a record in at least one of the catalogs.

Grover compared the cooperative cataloging of Latin-American books in RLIN and OCLC.⁶ He discovered that "both systems had cataloged almost the same number, although not the same books."⁷ In 1994, Erbolato-Ramsey and Grover followed up this study by comparing the amount of copy available in OCLC and RLIN for Spanish and Portuguese monographs.⁸ The researchers tracked hit rates over a period of months, and by the end of sixteen months, found records for 84 percent of the titles in RLIN, 91 percent in OCLC and RLIN combined, and nine percent in neither system.⁹ The article also explored the value of searching on a second system when copy is not found on the primary system. The average percentage of records found in OCLC and not in RLIN decreased over time, with an average of four percent over four years.¹⁰

In a 2003 study comparing hit rates among the Library of Congress catalog (LC), OCLC WorldCat, and RLIN, DeZelar-Tiedman, Genereux, and Hearn chose a sample of 433 titles newly ordered for the University of Minnesota Libraries.¹¹ Their research indicated that some level of copy (full or minimal level) was available for 51 percent of titles searched in LC, 74 percent in RLIN, and 81 percent in OCLC. Because there was not complete overlap between the titles found in RLIN and OCLC, the overall hit rate for all three resources was 91 percent.¹²

With the exception of the Williams and Beall/Kafadar articles, none of the studies cited above used random sampling to generate the set of titles to be searched. Instead, the samples were based on incoming materials encountered at the libraries at which the researchers worked. While this methodology is useful for predicting hit rates for the types of materials routinely cataloged at a particular library, the results are less transferable to different environments and different types of library materials. In addition, again excepting the Williams and Beall/Kafadar studies, the samples were comprised of newly acquired materials, and thus bear little relevance to the availability of records for older materials. Catalogers working on backlog projects, or cataloging gifts or special collections, often encounter older materials and face challenges in locating records for them.

Methodology

The integration of bibliographic records from the RLIN database into WorldCat commenced in February of 2007, with a target completion date of late summer 2007. As the Beall/Kafadar study limited itself to the WorldCat database, it is not known what impact the availability of the records formerly only in RLIN will have on the proportion of records from *Mansell* that are accessible in electronic form. The author wished to discover this impact by replicating and expanding on the methodology used in the previous study. The objectives of the present study are to test the following three hypotheses:

1. The Beall/Kafadar study is valid, and therefore the proportion of *Mansell* titles not found in WorldCat in the present study will not differ significantly from the original study.
2. A small but significant number of records will be found in RLIN that are not currently available in WorldCat.
3. A larger proportion of records than are available exclusively in RLIN will be available in neither WorldCat nor RLIN.

For logistical reasons, entries for non-Roman language materials were not included in the study. Non-Roman titles were excluded from the Beall/Kafadar study, and it was necessary for both studies to use the same selection criteria for *Mansell* records in order to accurately compare the results. In addition, titles in non-Roman scripts found in *Mansell* would have required transliteration before they could be searched in the online utilities. This would have considerably slowed the progress of the research.

As in the Beall/Kafadar study, the author used a random number generation program to select two *Mansell* volumes published in each of the years from 1969 to 1980, and one volume each from 1968 and 1981. For each volume selected, a list of thirty triples was randomly generated. The first number in each triple represented a page number for each volume from 1 to 694, the second was for the column number (1, 2, or 3) on a given page, and the third identified the record number in each column, from 1 to 7. If the record selected was for a non-Roman alphabet resource, a cross reference, or there was no such record (e.g. if a column had only five records), that triple was skipped and the next set of numbers in the list was used. Up to twenty records for each volume were searched. For

four volumes, the author was unable to find twenty valid records using the thirty triples selected. The total number of records searched was 502, as compared to 508 in the earlier study.

Between July and September of 2006, the author performed searches in both WorldCat and RLIN for each valid record identified in *Mansell*. The same criteria were used as in the Beall/Kafadar study when considering whether a match was found: the record had to match the *Mansell* entry exactly in title, format, date, and edition to be considered a match. Depending on the search results, the author coded each record as being found in “RLIN only”, “OCLC only”, “Both RLIN and OCLC” or “Neither RLIN nor OCLC”.

Results and Discussion

Table 1 compiles the results of the 502 searches. Slightly over half of the *Mansell* records searched (274, or 54.5 percent) were found in both RLIN and WorldCat. 69 records, or 13.7 percent of the sample, were found exclusively in WorldCat. 32 records, or 6.5 percent, were only in RLIN. More than 25 percent of the titles searched were not available in either bibliographic database.

(Insert Table 1)

To determine whether the number of hits in OCLC falls within the expected range, a standard deviation (σ) of 11.2 was calculated based on the full sample of 502 records. Since 72 percent of the *Mansell* records in the Beall/Kafadar study were found in OCLC, the result in the present study should fall within $2*\sigma$ on either side of 361.4, or 72 percent of 502.¹³ Therefore, in order to confirm the first hypothesis, the expected range of hits in OCLC is between 339.0 and 383.8. By adding the number of records found in both databases to those found only in WorldCat, the total number found in OCLC was 343, or 68 percent. This does fall within the expected range, and therefore, the validity of the earlier study is supported and the first hypothesis of the present study confirmed.

Though the number of records found only in RLIN was small, it nevertheless represents a notable percentage of records not presently available in WorldCat. With the integration of the two databases, users not formerly RLIN members will gain access to these records. It should also be said that libraries that were formerly members of RLIN but not OCLC will reap an even greater benefit once the databases are combined.

The most significant result, however, is that even with the combined database, 25 percent of the *Mansell* records searched are still unavailable in bibliographic utilities. These records generally represent old, rare, and obscure materials that are not widely held in libraries, and therefore may fly under the radar of library users and staff alike. However, they are more likely to be encountered by those who work with special collections or gift materials, and having access to other libraries' records for the same or similar material continues to be valuable. A library with a large uncataloged backlog of rare, special collections, or gift materials may want to think twice before de-accessioning *Mansell*, as tempting as the freeing up of shelf space might be.

One limitation of both the Beall/Kafadar and the current study is the exclusion of non-Roman language material. It is likely that materials in non-Roman scripts, particularly older material, are even more greatly underrepresented in bibliographic databases than those in Western European languages. This deserves further study. It might also be valuable to explore the quality and completeness of records for older materials in both WorldCat and *Mansell*. It has been demonstrated that 25 percent of *Mansell* entries are unlikely to be found in WorldCat, even with the inclusion of RLIN records. However, many *Mansell* records are minimal at best, only containing basic descriptive information. How useful are *Mansell* records to a cataloger when an electronic record cannot be located? Do the records in each resource provide an accurate description and an adequate number of access points?

Conclusion

Retrospective conversion of bibliographic records from card catalogs into online catalogs was a major effort of library catalog departments through the 1980s and into the 1990s. However, economies of scale caused many materials to be missed, or deliberately excluded, from these projects. Rare and special collections material, and materials in obscure languages and formats are more likely than mainstream monographs to remain in “hidden collections” in libraries, for obvious reasons. New areas of focus for libraries, including an emphasis on digital materials and alternative forms of access to library resources, make it unlikely that these backlogs will receive full cataloging on a large scale any time in the near future. Therefore, the ability to access information about these materials, even in “old-fashioned” book catalogs, remains crucial.

The current study demonstrates that a significant percentage of library materials, especially those that are old and rare, are not represented online in library catalogs. As libraries work toward tackling cataloging backlogs, it is necessary to paint an accurate picture of the completeness (or lack thereof) of retrospective conversion and the likelihood of finding copy for certain categories of materials. While the combination of the RLIN and OCLC databases will bring a slightly larger set of records to those libraries that previously subscribed to only one of the services, the number of records inaccessible electronically remains substantial.

Notes

1. Jeffrey Beall and Karen Kafadar, "The Proportion of NUC Pre-56 Titles Represented in OCLC WorldCat," *College & Research Libraries* 66, no. 5 (Sept. 2005): 431-35.
2. James W. Williams. "Pre-1950 Serials in OCLC: A Second Look at Database Records and a Comparison with *Union List of Serials* and *National Union Catalog, Pre-1956 Imprints*," *The Serials Librarian* 8, no. 4 (Summer 1984): 69-77.
3. Ibid., 72
4. Ibid., 76
5. James D. LeBlanc. "Towards Finding More Catalog Copy: The Possibility of Using OCLC and the Internet to Supplement RLIN Searching," *Cataloging & Classification Quarterly* 16, no. 1 (1993): 71-83.
6. Mark L. Grover. "Cooperative Cataloging of Latin-American Books: the Unfulfilled Promise," *Library Resources & Technical Services* 35, no. 4 (1991): 406-15.
7. Ibid., 409
8. Christiane Erbolato-Ramsey and Mark L. Grover. "Spanish and Portuguese Online Cataloging: Where Do You Start from Scratch?" *Cataloging & Classification Quarterly* 19, no. 1 (1994): 75-87.
9. Ibid., 80
10. Ibid., 82
11. Christine DeZelar-Tiedman, Cecilia Genereux, and Stephen Hearn. "Utilizing Z39.50 to Obtain Bibliographic Copy: A Cost-containment Study," *Library Resources & Technical Services* 50, no. 2 (Apr. 2006): 120-28.

12. Ibid., 125

13. Watson, G.S. "Hypothesis Testing," *Encyclopedia of Statistical Sciences*. Hoboken, NJ: Wiley & Sons, c2006.