

WorldCat and SkyRiver

A Comparison of Record Quantity and Fullness

Cathy Blackman, Erica Rae Moore, Michele Seikel, and Mandi Smith

In 2009, a new company, SkyRiver, began offering bibliographic utility services to libraries in direct competition to OCLC's WorldCat. This study examines the differences between the two databases in terms of hit rates, total number of records found for each title in the sample, number of non-English language records, and the presence and completeness of several elements in the most-held bibliographic record for each title. While this study discovered that the two databases had virtually the same hit rates and record fullness for the sample used—with encoding levels as the sole exception—the study results do indicate meaningful differences in the number of duplicate records and non-English-language records available in each database for recently published scholarly monographs.

The existence of SkyRiver means that libraries now have another choice of vendors from which to acquire bibliographic records and contribute original records. Therefore a comparison of the quantity and fullness of records available from OCLC's WorldCat and Innovative Interfaces's (III) SkyRiver can be helpful to libraries deciding which vendor would be best for their institution. This study compares the two databases in an attempt to determine whether there is a meaningful difference between them in terms of hit rates (percentages of records found in each database for each sample title), types of records (language of cataloging and format), and record fullness. Libraries can use this data in conjunction with other points of comparison, such as functionality, cost, and complementary services, when shopping for bibliographic services for cataloging. An understanding of the development of the two vendors and their products sets the stage for the comparisons made in this study.

Overview of OCLC and SkyRiver

In 1967, the presidents of Ohio's academic libraries established the Ohio College Library Center with the goal of using computer technology to share bibliographic records to help reduce cataloging costs. The shared database, now known as WorldCat, became a reality in 1971. Ohio's experiment was extremely successful and quickly grew into an international nonprofit membership organization, Online Computer Library Center (OCLC).¹

Cathy Blackman (cathyb@cameron.edu) is Assistant Professor/Librarian at Cameron University, **Erica Rae Moore** (emoore@cameron.edu) is Cataloging/Technical Services Associate at Cameron University, Lawton, Oklahoma, **Michele Seikel** (michele.seikel@okstate.edu) is Professor/Cataloger at Oklahoma State University, Stillwater, Oklahoma, **Mandi Smith** (masmith@cameron.edu) is Technical and Electronic Services Librarian at Cameron University.

Submitted June 18, 2013; returned to authors for revision September 25, 2013; revision submitted November 18, 2013; tentatively accepted for publication February 26, 2014 pending modest revision; revision submitted April 17, 2014; accepted for publication April 21, 2014.

The authors wish to note that each of them contributed an equal amount of work to this paper.

Between 1971 and 2009, OCLC increased the size of its WorldCat database to more than 200 million records through the original cataloging records contributed by its members and with the acquisition of the Research Libraries Group's Research Libraries Information Network (RLIN) database and the Washington Library Network's (WLN) database.² However, OCLC's exponential growth has not been without difficulty. OCLC has faced not only monopoly related allegations from its competitors, but has had to face many challenges as the company attempts to build its international base. Various languages, bibliographic formats, and cataloging rules have made OCLC's foray into the international market challenging and have resulted in problems for its end user clients. Despite those difficulties, OCLC now serves libraries in 170 nations.³

In 1978, Jerry Kline co-founded Innovative Interfaces Inc. (III), which created as its first product a hardware/software system that libraries could use to automate the transfer of bibliographic records from OCLC into local catalogs.⁴ Having observed the growth of OCLC for three decades, Kline decided to mount a challenge by launching SkyRiver in 2009, only three years after the merger of OCLC and its last competitor. In 2013, Kline sold his interests in SkyRiver and Innovative Interfaces, and the new owners merged the two companies. This merger brings the bibliographic database, SkyRiver, into the product line of the parent company, III, just as WorldCat is a product of OCLC.⁵

At the outset, SkyRiver's database consisted of a variety of public sources including the Library of Congress (LC), the British Library (BL), and the Cooperative Online Serials Program (CONSER) records.⁶ SkyRiver has grown its database by incorporating the bibliographic records of existing local catalogs (including records originating from member libraries using WorldCat) and by the addition of new, original records created by SkyRiver customers.⁷ Partnerships with vendors have also expanded SkyRiver's ability to provide a wider range of bibliographic records to its clients. For example, *Library Journal* reported that in 2012, SkyRiver began a partnership with the Donohue Group Incorporated (DGI), which provides catalog records for recorded books as well as publisher's cataloging-in-publication (CIP) records from small and independent presses.⁸ According to SkyRiver, its database had 43 million records as of July 2013.⁹

Development of a Study of OCLC and SkyRiver

One possible reason for the size disparity between WorldCat's nearly over 300 million records and SkyRiver's 43 million is that SkyRiver aims to provide one unique record per title without duplicates.¹⁰ The company believes this will "[save] catalogers time and [reduce] searching frustration."¹¹ However, because of OCLC's global reach, WorldCat may have multiple records for a title, including many

English-language records plus records from international libraries whose language of cataloging is not English.¹² The size disparity between the two databases forms the basis of the quantity-related aspects of this study.

The authors devised a study to determine whether there would be a meaningful difference between the hit rates for the two databases. During the period of this study, catalogers began to use Resource Description and Access (RDA) rules to create new records, but since those rules do not affect the analyzed elements and criteria used in this study, the authors do not believe that RDA implementation influenced the study results. Additionally, this study takes into consideration the different types (e.g., non-English, print, etc.) and quantities of records for each title found in each database from a sample of 368 scholarly monographs.

Libraries are not only interested in whether a record for a title is available, but also in the quality of the bibliographic record found. Record quality has often been an important factor when comparing competing databases. Despite this importance, the concept of record quality remains inherently subjective as evidenced by the varying definitions and standards reported by Bade.¹³ For that reason, the authors chose to focus on the inclusion and completeness of certain MARC21 fields in a bibliographic record as indicators of record fullness. Those elements include matching International Standard Book Number (ISBN), matching date in the 008 (leader field), encoding level, LC call number, physical description, the presence of LC subject heading(s), and contents and summary notes. Because the samples for the study are scholarly monograph titles taken from the new purchases of two American university libraries, the inclusion of the LC call number (050/090 fields) was considered a more cogent indicator of record fullness for the sample than the Dewey classification number (082/092 fields). This was because the majority of academic libraries in the US use the LC classification system.¹⁴ LC subject headings were chosen as an analyzable element because they are carefully constructed and monitored by LC and Subject Authority Cooperative Program (SACO) members. The two university libraries from which the samples were drawn are not located in the same state as those of the researchers, nor do they have any direct ties.

Literature Review

There is certainly a historical tradition of comparing WorldCat's quantity and quality of bibliographic records with those of other databases. Describing the results of an exhaustive 1993 study of OCLC and WLN records, Ross concluded that "the hit rate for new monographic titles differed only by 1.4% between OCLC and WLN even though their databases vary substantially in size, OCLC with 24.8 million

bibliographic records and WLN with 7.8 million records.¹⁵

Hillman and Sugnet published a study of the OCLC and RLIN databases. Their 1983 findings indicated that OCLC's database would be more likely to produce results for older materials.

Probably the most difficult factor to analyze is the difference in coverage and size of the database. For some older material and state documents especially, the hit rate on RLIN is much poorer than on OCLC. . . . Searching on OCLC, the cataloger may come up with an older, retrocon record needing extensive revision. Searched on RLIN, one is more likely to find no record at all, which means that the cataloger must do that title originally.¹⁶

Writing in 1989, Intner could not find a statistical difference between WorldCat and RLIN, despite the two utilities' dissimilar philosophies of quality control and the commonly held belief "that OCLC is big and dirty, while RLIN is small and clean."¹⁷ While RLIN's original focus was on cataloging quality, OCLC initially focused on increasing the size of its database. Intner measured quality by measuring the accuracy of punctuation, capitalization, spelling, etc., in various MARC fields.¹⁸ Ross chose different indicators of record quality for her 1993 study. Her quality measurements included the encoding level and the inclusion of LC or Dewey call numbers.¹⁹

Instead of focusing on quality (that elusive term), recent research is focusing on how frequently various MARC fields are used. In 2006, Moen et al. published a large-scale study of millions of bibliographic records in OCLC's WorldCat. For this study, each occurrence of a field was counted, even if repeatable fields had multiple occurrences for a single record. For 2004, the most recent year studied, six MARC fields were found to have the most occurrences in monographic bibliographic records: 650 (subject heading), 245 (title statement), 008 (fixed fields), 500 (general note), 100 (personal name main entry), and 700 (personal name added entry), respectively. It should be noted that the study excluded system-generated fields, such as 001 (control number), 040 (cataloging source), and 029 (other system control number). While occurrences for MARC fields 260 (publication distribution, etc.), and 300 (physical description) were reported for previous years, they were not listed as commonly occurring fields for 2004.²⁰

Ongoing experimental OCLC research is also looking at the frequency of MARC fields; however, instead of focusing on the number of occurrences of various MARC fields, OCLC is studying what percentage of records in their database utilize a particular MARC field. Consequently, the number of occurrences of repeated fields is irrelevant to this research. All monographic bibliographic records analyzed

include an 040 and 245. Other fields that were used more than half of the time were 260 (93.65 percent), 300 (89.33 percent), and 100 (63.77 percent).²¹

In preparation for this study, the authors could not find any direct comparisons focusing on record quality or fullness between OCLC and its most recent competitor, SkyRiver. This holds true for both formal, peer-reviewed literature and less formal reports publicly available, including committee minutes from various library related groups. While not a comparison, a 2009 post to the OCLC-CAT discussion list outlines some statistics on encoding levels and LC-created records for WorldCat. At that time, nearly two-thirds (64 percent) of the records in WorldCat were considered minimum-level records, records that are cataloged as "less-than-core" (encoding level 2, 3, 4, 5, 7, K, or M). Additionally, less than one-tenth (8.6 percent) of the cataloging in WorldCat originated at LC.²² Beyond the promotional material from SkyRiver about its initial database, the authors did not discover any third-party statistics concerning any quality- or fullness-related indicators for the SkyRiver database.

Some of the less formal reports included quantity-related information concerning hit rates between OCLC and SkyRiver. Michigan State University (MSU) reported that the hit rate for approval plan books decreased only slightly with SkyRiver: 95–98 percent for OCLC compared to 93–95 percent for SkyRiver. Michigan State was one of SkyRiver's earliest large university library clients, and this favorable report, published in the *Association for Library Collections and Technical Services (ALCTS) Newsletter Online (ANO)*, may have encouraged other libraries to consider SkyRiver seriously.²³ In a 2010 report, Janes reported that the hit rate for her sample of new scholarly monographs at the Mabie Law Library in the University of California-Davis (UC-Davis) was 100 percent for OCLC and 98 percent for SkyRiver, a statistically insignificant difference.²⁴

Conversely, the committee minutes from two consortia do not show SkyRiver's hit rate in such a favorable light. During the March 2011 administrators' meeting of System-Wide Automated Network (SWAN), a Chicago-area consortium of eighty libraries, one participant mentioned a 2:1 ratio for original cataloging, which would result in increased costs for original cataloging activity if the group switched to SkyRiver.²⁵ The statistics reported to the ILS committee of the South Central Library System in Wisconsin in August 2012 show a hit rate of 90 percent for OCLC, but only 50–60 percent for SkyRiver. The report mentions that SkyRiver staff can supply records for 25–30 percent of items not found within forty-eight hours on request, but a committee member expressed concern about workflow while waiting for the supplied records.²⁶ Neither the Chicago-area consortium nor the Wisconsin reports mention the sample or method used for their comparisons.

Research Methods

To study whether there were meaningful differences in either hit rates or record fullness between WorldCat and SkyRiver, the authors chose to analyze a sample of titles. The sample for this study was provided by two academic libraries who had previously indicated a willingness to provide data, the University of North Carolina-Charlotte (UNCC) and the University of Arkansas-Little Rock (UALR). At the time of the study, UNCC had a full-time equivalent (FTE) of more than 20,000 students and, per the American Library Survey (ALS), was listed as Carnegie class Masters I. UALR had an FTE of approximately 8,000 students and, per the ALS, was listed as Carnegie class Doctoral/Research-Intensive. Each university library emailed the sample titles, 13-digit ISBNs, author and editor names, publication dates, and edition numbers of their recent print monograph purchases—368 titles in total—to the authors. Both libraries sent the information to the researchers using their Baker & Taylor YBP Library Services (YBP) order carts, so this could be described as a convenience sampling. The sample from UNCC consisted of the materials purchased from August 6 to September 20, 2012, a total of 244 titles, and the sample from UALR was for materials purchased from October 24, 2012 to February 28, 2013, a total of 124 titles. The sample information was incorporated into a Microsoft Excel spreadsheet, which was used to track and organize both the information provided for each book and the data found from the searches in both databases. The sample included monographs published in both English and Spanish. Publication dates for the sample ranged from 1977 to 2013; 89.9 percent were published from 2009 to 2013. There was no duplication of titles between the two samples.

Determining Hit Rate

To determine the hit rate in each database, an ISBN search in WorldCat and SkyRiver was conducted for each title in the sample. If the ISBN search failed to yield results, a title search was conducted. All searching and recording of data for record results was the same for WorldCat and SkyRiver, and took place during the same seven-day period, to lessen the likelihood that records could be added or modified in each database. To the authors' knowledge, there is no way to determine whether any of the SkyRiver records found and analyzed originated in WorldCat because SkyRiver records do not utilize an 035 field (system control number) containing an OCLC record number. However, SkyRiver contains "the Library of Congress MARC files and CONSER files," according to its webpage, as does WorldCat.²⁷

The total number of records resulting from either the ISBN or title search in both databases was recorded. The records found were categorized by type of record, and

the number of records of each type was recorded. Types of records found included non-English language records, English-language records, English-language print records, and e-book records in any language. The most widely held English-language print record found with a matching date was analyzed for fullness. If there were no records with a matching date, the most widely held English-language print record was analyzed. If a title search was necessary, the record's ISBN would not match the provided ISBN. The choice to count these records as hits was made even though some libraries may opt to create a new bibliographic record per local practice. Additionally, OCLC's Bibliographic Formats and Standards states that the absence, presence, or difference in ISBN does not justify the creation of a new record.²⁸

Many MARC elements were analyzed for the record which had the most holdings in OCLC and SkyRiver for each title (see tables 3, 4, and 6). For the 040 (cataloging source) field, the authors recorded the subfield a, which lists the code for the institution which created the record, and subfield b, which lists the language of transcription of the record. The authors considered subfield a to compare the composition of contributors of bibliographic records for the sample studied. The completeness level of the LC call number was also noted. LC call numbers were considered complete (with LC classification, Cutter, and publication year/volume number in a series), partial (missing one of the three elements), or absent. Call numbers for monographs published before the early 1980s, when adding the publication year to the call number became common practice, were considered partials, since the current practice would require altering the call number in most instances during the cataloging process. Additionally, the physical description of the resource (the 300 field) in each analyzed record was examined to determine whether it was complete (pagination and dimensions were present), or partial (if either of those were missing). The number of LC subject headings present in the record were also counted and recorded. Finally, the authors noted whether 505 (table of contents) and 520 (summary) fields were present in each record because they are two notes which are thought to be useful to patrons. Both notes are frequently found on full-level bibliographic records created within the past few years.

Results

This study focused on a comparison of both the quantity and fullness of records found in the two databases. Quantity was further broken down into the hit rate and the counts of the various types of records. There was little difference found between the two databases' overall hit rates (see table 1). Of the 368 titles searched, 363 (98.64 percent) were found in WorldCat and 362 (98.37 percent) in SkyRiver.

Table 1. Hit Rate Results for WorldCat and SkyRiver

Type of Search	WorldCat		SkyRiver	
	No.	%	No.	%
ISBN search	356	96.74	352	95.65
ISBN search (matching date)	350	95.11	341	92.66
ISBN or Title search	363	98.64	362	98.37
ISBN or Title search (matching date)	356	96.74	347	94.29

Table 3. MARC 21 Fields Used as an Indicator of Fullness

Field	WorldCat		SkyRiver	
	No.	%	No.	%
Matching ISBN	356	96.74	352	95.65
Matching Date	354	96.20	347	94.29
Full LC Call Number (050)	354	96.20	350	95.11
Complete Physical Description (300)	359	97.55	354	96.20
LC Subject Headings (6xxs)	357	97.10	354	96.20

More noticeable differences were discovered when focusing on the various types of records (see table 2). In WorldCat, of the 368 titles searched, 296 (80.43 percent) had records whose language of cataloging is not English, while SkyRiver had only 1 (0.27 percent) record of that type. Additionally, because SkyRiver states that “sophisticated matching algorithms minimize duplication,”²⁹ the percentages of searches for each database that resulted in only one or two total records, one or two English-language records, and one English-language print record were noted. Overall, a sizeable difference between the databases was found, with 70 (19.02 percent) of 368 items searched in WorldCat resulting in only one or two total records, contrasted with 320 (86.98 percent) of 368 titles searched in SkyRiver. The difference decreased after removing non-English language records and nonprint records. After these records were removed from consideration, of the 368 searched items, 160 (43.48 percent) of the titles searched resulted in a single English-language print record in WorldCat, as opposed to 304 (82.61 percent) in SkyRiver.

Table 3 shows the percentage of records found in WorldCat and SkyRiver with elements whose completeness indicate fullness. The two databases had no substantial differences when the authors compared these elements. Considering the elements individually, in WorldCat, 356 of the 368 titles (96.74 percent) had matching ISBNs and 352 of the 356 titles (95.65 percent) did in SkyRiver. When comparing the most-held record in terms of matching date, there was a slightly larger discrepancy with WorldCat having 354 titles (96.20 percent) and SkyRiver having 347 (94.29 percent).

Table 2. Hit rates for Various Types or Categories of Records

Records	WorldCat		SkyRiver	
	No.	%	No.	%
Non-English language	296	80.43	1	0.27
e-books with print ISBNs included	235	63.86	191	51.90
1 or 2 total records per item	70	19.02	320	86.96
1 or 2 English-language records per item	167	45.38	319	86.68
1 English-language print record per item	160	43.48	304	82.61

Table 4. Original Cataloging Source

Field	WorldCat		SkyRiver	
	No.	%	No.	%
Library of Congress (DLC)	234	66.03	240	65.22
Baker and Taylor (BTCTA)	35	9.51	31	8.42

The breakdown of complete call numbers for the two databases are WorldCat at 354 (96.20 percent) and SkyRiver at 350 (95.11 percent). Both databases had more records with complete physical descriptions, WorldCat with 359 (97.55 percent) and SkyRiver with 354 (96.20 percent). The smallest difference can be found when comparing the number of most-held records with at least one LC subject heading. Out of the 368 titles searched, WorldCat had 357 titles (97.10 percent) with at least one LC subject heading and SkyRiver had 354 titles (96.20 percent). It is significant to note that both WorldCat and SkyRiver scored approximately 95 percent or higher for each of the analyzed record elements.

A comparison of the most-held records for each title showed that 84.51 percent of the titles had the same cataloging source as indicated by the MARC field 040, subfield a (see table 4). Further study indicated that the highest percentage of records analyzed were created by LC. Of the 368 titles searched, 243 (66.03 percent) of the most-held records in WorldCat and 240 (65.22 percent) in SkyRiver were initially created by LC. The second most common cataloging source was the vendor Baker and Taylor. Of the 368 titles searched, 35 (9.51 percent) of the most-held records in WorldCat and 31 (8.42 percent) in SkyRiver were created by Baker and Taylor.

Differences were discovered when comparing record encoding levels (see table 5). SkyRiver has integrated WorldCat’s encoding level terminology. The encoding level dropdown box used in the SkyRiver platform explicitly states “OCLC” and offers OCLC definitions for the various encoding levels. When searching WorldCat, 217 (58.97 percent) of the 368 most-held records analyzed had a blank encoding level, while 255 (69.11 percent) records had a

Table 5. Record Encoding Level

Field	WorldCat		SkyRiver	
	No.	%	No.	%
blank	217	58.97	255	69.11
1	0	0.00	0	0.00
2	0	0.00	0	0.00
3	3	0.82	0	0.00
4	61	16.58	26	7.07
5	0	0.00	0	0.00
6	0	0.00	0	0.00
7	4	1.09	7	1.90
8	7	1.90	13	3.53
u	0	0.00	0	0.00
z	0	0.00	0	0.00
i	59	16.03	50	13.59
k	1	0.27	1	0.27
l	0	0.00	0	0.00
m	10	2.72	8	2.17
e	0	0.00	0	0.00
j	0	0.00	0	0.00

blank encoding level in SkyRiver. As previously mentioned, approximately two-thirds of the WorldCat and the SkyRiver records analyzed may have had the same LC origin. Additionally, of the 368 records analyzed, 61 (16.58 percent) had a “4” encoding level in WorldCat and 26 (7.07 percent) in SkyRiver. The differences are eliminated if the blank and 4-level records are added together (75.55 percent for WorldCat and 76.18 percent for SkyRiver).

Of the 368 titles searched in WorldCat, 260 (70.65 percent) had records that included a table of contents and 137 (37.23 percent) had records that included a summary note (see table 6). In SkyRiver, 273 (74.18 percent) of the 368 titles searched had records that included a table of contents note and 161 (43.75 percent) included a summary note.

While each institution has different priorities when choosing acceptable records, most institutions consider more than a single field when making that choice. Therefore, examples of how the MARC fields analyzed can be combined to help with the decision-making process are provided (see table 7). In WorldCat, 217 (58.96 percent) of the 368 titles searched had records that included a blank encoding level, had a full LC call number, a complete physical description, and at least one LC subject heading. In SkyRiver, 246 (66.85 percent) of 368 titles searched resulted in records that had those same characteristics. When adding the condition of having only one English-language print record among the search results, however, the differences

Table 6. Presence of TOCs and Summaries

Field	WorldCat		SkyRiver	
	No.	%	No.	%
Table of Contents (505)	260	70.65	273	74.18
Summary (520)	137	37.23	161	43.75

between the two databases became more apparent with WorldCat having 146 (39.67 percent) and SkyRiver having 276 (75.00 percent) that fall into that category (see table 7).

Discussion

The goal of this study was to determine whether there are meaningful differences between WorldCat and SkyRiver in terms of hit rate, fullness, and types of records. Overall, there were no noticeable differences in hit rate. The databases for both companies had a greater than 98 percent hit rate; OCLC had only a single title (0.27 percent) more than SkyRiver. This study supports the assertion made by Janes and MSU that the differences in search results are insignificant, despite the size disparity between the two databases.³⁰ According to the results of this study, WorldCat’s larger database (over 300 million records) did not result in a noticeably better hit rate than SkyRiver’s smaller database (over 40 million records) for the sample. This study’s results did run counter to the hit-rate results documented in the minutes of the Chicago area consortium’s administrative meeting or the South Central Library System’s committee meeting. The authors discovered a much higher hit rate (98.64 percent) than the 2:1 ratio or 50–60 percent offered in the minutes.³¹ The reason for this may be that at the time of those studies, SkyRiver was still in the early stages of its evolution; its database has grown rapidly in the past two years.

It was not possible to compare most of the fullness-related results (ISBN, dates, LC call number, physical description, and LC subject headings) of this study with previous studies due to the lack of recent studies available for either database. Overall, there was very little difference between the fullness for the most-held records for the sample in WorldCat and SkyRiver. Nearly half of all records (44.29 percent) were exactly the same for every recorded element, down to the number of LC subjects. As previously noted, approximately 95 percent or more of the records analyzed from both databases had matching ISBNs, matching dates, complete LC call numbers, complete physical descriptions, and the inclusion of LC subject headings. There were no differences for this collection of fullness-related elements. Since approximately three-fourths of the analyzed records in both WorldCat and SkyRiver originated from LC or Baker and Taylor, it is understandable that there would be many

Table 7. Records with Combined Multiple MARC21 Fields

Example 1				Example 2				Example 3			
Any no. of records				Any no. of records				1 Eng.-lang. print record			
"Blank" Encoding level				Any Encoding Level				Any Encoding level			
Full LC Call Number				Full LC Call Number				Full LC Call Number			
Complete Physical Desc.				Complete Physical Desc.				Complete Physical Desc.			
>1 LC Subject Heading				>1 LC Subject Heading				>1 LC Subject Heading			
WorldCat		SkyRiver		WorldCat		SkyRiver		WorldCat		SkyRiver	
Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
217	58.96	246	66.85	351	95.38	334	90.76	146	39.67	276	75.00

records that are essentially identical, regardless in which database they were found.

Because this study focused on a subset of records for each title, the percentage of records created by LC in WorldCat varied greatly from the 8.6 percent stated by the Director of WorldCat Quality Management in his 2009 posting to OCLC-CAT.³² Instead of considering all records in WorldCat as he did, this study analyzed the most-held record for each title in the sample. With this subset, the percentage of LC-created records in WorldCat rose to 66.03 percent, a negligible difference when compared with SkyRiver's results of 65.22 percent. It was not possible to compare this study's SkyRiver results with that of any previous studies.

As was the case for the statistics for LC-created records, this study's encoding level results did not match those previously reported in the OCLC post. While the post reported that 64 percent of all WorldCat records were minimum level records, only 21.48 percent of the analyzed records in the sample from WorldCat were minimum level. Again, the authors attribute the difference to the fact that only the most-held records found for each title were analyzed. Another possible explanation could be the nature of the sample used. It was not possible to compare SkyRiver's encoding results from this study with that of any previous studies.

While this study discovered that the two databases had virtually the same hit rates and record fullness for the sample used—with encoding levels as the sole exception—dramatic differences were discovered when various types or counts of records were compared. Although many libraries in the US, Great Britain, Australia, or any country that follows the Anglo-American Cataloging Rules (AACR2) and/or RDA may be able to utilize SkyRiver for their bibliographic needs, libraries that follow different cataloging rules or need records in a language other than English would likely be better served by WorldCat. OCLC's global focus is evident with over four-fifths (80.43 percent) of all titles searched resulting in at least one record transcribed in a non-English language. However, with OCLC serving 485 languages and

dialects, this study's results of an average of 3.1 non-English records per title suggests that it is possible that not every member library would always find a useable record for their particular needs.³³ Based on the results of this study, WorldCat's inclusion of non-English language records is currently much higher than that for SkyRiver. Out of the 368 titles searched in SkyRiver, only one (0.27 percent) resulted in a non-English language record.

In addition to non-English language records, the authors documented the number of print records, the number of e-book records, and the total number of records found for each searched title. Because only print ISBNs were searched, the documented figures do not represent an accurate depiction of the e-book record composition for either database; instead, the results may indicate the percentages of records for each database where the print ISBNs were included on the e-book records as suggested by the Program for Cooperative Cataloging's (PCC) "Provider-Neutral E-Monograph MARC Record Guide."³⁴ WorldCat had a higher percentage (63.86 percent) of print ISBNs included on e-book records than SkyRiver (51.90 percent).

This documentation of the number and types of records occurred because, as previously mentioned, the authors wanted to test SkyRiver's public statement that "sophisticated matching algorithms minimize duplication and sub-standard records, saving catalogers time and reducing searching frustration."³⁵ However, as many e-book records included print ISBNs, a decision was made to gather statistics for the number of searched titles with one or two records total because the authors' assumption was that SkyRiver should have one record for the print resource and one record for the electronic resource. Again, there was a dramatic difference between the two databases when comparing the percentage of titles that had a total of one or two records for each title searched. Nearly nine-tenths (86.98 percent) of all titles searched in SkyRiver resulted in one or two records. This result was more than four times larger than that found in WorldCat, which had less than one-fifth (19.02 percent) of the searched

titles resulting in only one or two records. Part of the disparity can be accounted for by the fact that WorldCat contains records whose language of cataloging is not English. When non-English records were removed from the comparison, the gap narrowed. After removal, nearly half (45.38 percent) of all titles searched in WorldCat resulted in one or two total English-language records, while SkyRiver's percentage remained approximately the same at 86.68 percent.

Another area of focus was the number of searched titles that resulted in a single English-language print record per title, after removing all non-English and nonprint records from consideration. These figures corresponded consistently with the previous results. Of the 368 titles searched, 43.48 percent in WorldCat resulted in a single English-language print record as compared to 82.61 percent in SkyRiver.

While the number of resulting records in WorldCat ranged from a minimum of 0 and a maximum of 42 records per title, the average number of records resulting from each searched title was 6.56 total records. The searched titles in SkyRiver resulted in a minimum of 0 records and a maximum of 5 records, with an average of 1.81 total records per searched title. Differences in the number of results per searched title can have a tremendous effect on the decision-making process. Some libraries may prefer having more records available from which to choose. Other libraries may prefer having one distinct record per title, or at least fewer records to evaluate during the selection process.

Because the results of this study may factor into libraries' decision-making processes when considering bibliographic services, it is important that any limitations and issues with the research method are clearly outlined. As previously mentioned, unlike Intner's method, the authors did not verify the accuracy of analyzed elements in the records; only their inclusion was measured. This is particularly relevant when discussing encoding levels because that was the single record element with noteworthy differences between the databases.

As previously stated, all searching was done in the same one-week period. Although all searching and record analysis occurred in the same one-week period after the full sample was compiled, many of the titles had been published and available for distribution several months before the searching. All the searched titles in the sample had a gap of at least three months between the publication date and the searching date. It is entirely possible that the results would have been different if the searching had taken place within the same week, or within a few weeks, of each title's publication date. One database may have a faster turn-around time for the inclusion of records that might affect hit rates. Further research is needed to determine whether the two databases differ in how long it takes to include new records and whether that difference significantly affects the hit rate comparison.

The final research-related issues are connected with the sample itself. The sample is in no way representative of library acquisitions at large. Without knowing the exact

scope of the project, the authors chose to limit the sample to print monograph titles to contain any potential issues that might come up with nontraditional formats. E-books, DVDs, streaming videos, audiobooks, music CDs, cartographic materials, and other types of resources were not part of this study. Widening the sample to include more formats or more non-English language materials might have affected the study's results. The majority of the print monograph titles in the sample were scholarly books in English, published within the past four years. Each book was chosen by two specific academic libraries. Public libraries, special libraries, and even other types of academic libraries may have very different acquisitions needs—even in terms of print monographs.

The limited sample necessitates further research with different or larger samples to gain a better understanding of how WorldCat and SkyRiver compare. Further research can focus on factors other than hit rates, types and counts of records, and record fullness. For example, further research will need to be conducted to study the effect of the adoption of RDA. While there are many factors that need to be considered to obtain a more complete picture of the two databases, it is highly recommended that future research focus on functionality, cost, and complementary services offered by each company. When the functionality of WorldCat and SkyRiver were compared by the catalogers at UC-Davis, they reported that SkyRiver was less complicated to learn and more efficient when used with Innovative's Millennium integrated library system (ILS).³⁶ Further research could be done to see if these efficiencies hold true for other libraries, especially those using a different ILS. If OCLC implements the recommendation of its Global Advisory Group on Credits and Incentives to transition its current Financial Credit Program into a subscription pricing model, new cost comparison studies will need to be conducted.³⁷ Complementary services, especially holdings-related services, can be deal-breakers when comparing the two companies and a more holistic comparison should include such services.

Conclusion

There is no meaningful difference between the percentage of records found for each title in WorldCat and SkyRiver for this study's sample. Record fullness was also very similar in each database, possibly because for both databases approximately three-fourths of the most-held records were created by LC or Baker & Taylor. Because of the virtually identical hit rate and record fullness, the results of this study suggest that it may be possible to eliminate these factors from the decision-making process when choosing a vendor. In terms of this study, it may be more prudent to focus on the more pronounced differences between the two databases: the total number of records found per search and the number of records whose language of cataloging is not English. The

figures show that WorldCat is currently much more global in scope than SkyRiver, containing, for many of the titles searched, non-English language records. The results also support the conclusion that SkyRiver is thus far adhering to its implied intention of limiting duplicate records, as approximately nine-tenths of all titles (87 percent) had only one or two records. However, this study is a snapshot that examines the state of each company's database in 2012–13. Given that III, SkyRiver's parent company, has a large international customer base, SkyRiver's database may acquire more non-English records in the future. Studies featuring bibliographic records for other types of materials would be of interest in further determining difference in both quality and quantity between the two companies and their databases.

References

1. "OCLC's presidents," OCLC, accessed May 3, 2013, www.oclc.org/en-US/about/leadership/presidents.html.
2. Judy Janes, "SkyRiver or OCLC?" *Spectrum Online*, November 21, 2011, accessed March 1, 2013, www.aallnet.org/main-menu/Publications/spectrum/Spectrum-Online/skyriver.html.
3. Doris Small Heifer and Helen Heinrich, "OCLC: Is Its Future Up in the Clouds?" *Searcher* 20 no. 2 (2012): 22–23.
4. Brian Kenney, "Being Innovative," *Library Journal* 129, no. 14 (2004): 38–39, accessed May 3, 2013, www.library-journal.com/lj/ljinprintcurrentissue/872626-403/being_innovative.html.csp.
5. Meredith Schwartz and Bob Warburton, "III Drops OCLC Suit, Will Absorb SkyRiver," *Library Journal* 138, no. 6 (2013): 12, accessed May 3, 2013, www.infodocket.com/2013/03/04/innovative-interfaces-integrates-all-skyriver-services-and-withdraws-antitrust-lawsuit-against-oclc.
6. Joshua Barton and Lucas Mak, "SkyRiver at Michigan State University Libraries: A Brief Overview," *ALCTS Newsletter Online* 21, no. 2 (2010), accessed April 25, 2013, www.ala.org/alcts/ano/v21/n2/feat/system.
7. "Frequently Asked Questions," SkyRiver, accessed May 7, 2013, theskyriver.com/faqs.
8. David Rapp, "SkyRiver, Donohue Group Announce Partnership," *Library Journal* 137, no. 2 (2012): 18.
9. "SkyRiver," Innovative Interfaces, accessed April 14, 2014, www.iii.com/products/skyriver.shmtl.
10. "A Global Library Resource," OCLC, accessed May 7, 2013, www.oclc.org/en-US/worldcat/catalog.html; Janes, "SkyRiver or OCLC?"
11. "SkyRiver."
12. Heifer and Heinrich, "OCLC," 22–23.
13. David Bade, "The Perfect Bibliographic Record: Platonic Ideal, Rhetorical Strategy or Nonsense?" *Cataloging & Classification Quarterly* 46, no. 1 (2008): 109–33.
14. Jay Shorten, Michele Seikel, and Janet Ahrberg, "Why Do You Still Use Dewey?" *Library Resources & Technical Services* 49, no. 2 (April 2005): 123–36.
15. Rosemary E. Ross, "A Comparison of OCLC and WLN Hit Rates for Monographs and an Analysis of the Types of Records Retrieved," *Information Technology & Libraries* 12, no.3 (1993): 359–60.
16. Diane Hillman and Christopher Sugnet, "Comparison of OCLC and RLIN: A Question of Quality," *Cataloging & Classification Quarterly* 4, no.1 (1983): 70.
17. Sheila Intner, "Much Ado about Nothing: OCLC and RLIN Cataloging Quality," *Library Journal* 114, no. 2 (1989), 38.
18. *Ibid.*, 39–40.
19. Ross, "Comparison of OCLC and WLN," 355–59.
20. William E. Moen et al., "Catalogers' Use of MARC Content Designation over Time: An Analysis of MARC Records from 1972 to 2004," 8, in *MARC Content Designation Utilization: Inquiry and Analysis*, 2007, accessed September 30, 2013, www.mcdu.unt.edu/wp-content/CatalogersUseOverTime_Final_30Dec2007.pdf.
21. "MARC Usage in WorldCat," OCLC, accessed March 14, 2014, experimental.worldcat.org/marcusage.
22. Glenn Patton, "Posting to OCLC-CAT February 26, 2009," accessed May 7, 2013, <http://listserv.oclc.org/archives/oclc-cat.html>.
23. Barton and Mak, "SkyRiver at Michigan State University Libraries."
24. Janes, "SkyRiver or OCLC?"
25. "SWAN Administrators' Quarterly Meeting Minutes, March 3, 2011," accessed April 25, 2013, www.mls.lib.il.us/swan/archive/2011_3-3_SWAN_Quarterly_Notes.pdf.
26. South Central Library System Wisconsin, "ILS Committee Meeting Minutes, August 1, 2012," accessed April 25, 2013, www.scls.info/committees/ic/minutes/2012-08-01.pdf.
27. "Skyriver."
28. "When to Input a New Record," OCLC, accessed March 14, 2014, www.oclc.org/bibformats/en/input.html#CHDJFHA.
29. "SkyRiver."
30. Janes, "SkyRiver or OCLC?"; Barton and Mak, "SkyRiver at Michigan State University Libraries."
31. "SWAN minutes"; "South Central minutes."
32. Patton, "Posting to OCLC-CAT."
33. "A Global Library Resource," OCLC, accessed March 10, 2014, oclc.org/worldcat/catalog.en.html.
34. Becky Culbertson, Yael Mandelstam, and George Prager, "Provider-Neutral E-Monograph MARC Record Guide," accessed March 10, 2014, www.loc.gov/aba/pcc/bibco/documents/PN-Guide.pdf
35. "SkyRiver."
36. Janes, "SkyRiver or OCLC?"
37. OCLC Global Advisory Group on Credits and Incentives, "Final Report," accessed May 5, 2013, www.oclc.org/content/dam/oclc/councils/global/global-advisory-group-on-credits-and-incentives.pdf.