PAUL METZ AND JOHN ESPLEY

The Availability of Cataloging Copy in the OCLC Data Base

A sixteen-week longitudinal study was conducted to determine the effectiveness of OCLC as a source of cataloging data and to optimize the timing of searches for cataloging copy for various categories of materials. The findings indicated a high rate of success and, further, suggested that for many types of materials a holding pattern might be unnecessary. A mid-sized research library should be able to clear about half of its monographic receipts immediately, if it is willing to accept CIP copy. For materials not searched immediately, or for subsequent searches of materials not cataloged at once, the data may be used to determine the best timing and frequency of searches.

ANY LIBRARY that relies on an on-line bibliographic utility as its primary source of cataloging copy confronts a number of critical decisions that determine how effectively and efficiently the on-line data base can serve its needs. For example, a blanket decision to accept, to reject, or to inspect and modify cataloging copy from particular sources represents an important choice between the goal of quality and the goals of speed and economy. An especially critical series of decisions must be made addressing the questions of when it is most profitable to search the data base for contributed copy, how often and at what intervals to repeat the search when copy is not found, and at what point to abandon the search in favor of original cataloging. Decisions of this nature represent a balancing of several goals, in that the library seeks simultaneously to minimize the extent of original cataloging, to process materials as quickly as possible, and to minimize the number of searches required to find copy. The library will also generally seek to safeguard the integrity of its authority structure, often by maximizing its use of Library of Congress copy.

This study presents empirical data that might provide a basis for informed decisions about cataloging searches of OCLC, the largest and most heavily used cataloging data base. Other studies have evaluated the effectiveness of OCLC as a resource for ILL and preacquisitions verification and for cataloging data.[1-4] Meyer and Panetta, in their comparison of OCLC and B/NA as cataloging data bases, touch briefly on how the probability that copy for a new title will be found on OCLC increases with time.[5] But even the most comprehensive and authoritative study, Hewitt's *OCLC: Impact and Use*, while pointing to the need for "an evaluation of the relationship between original find rates, holding patterns, and final find rates," could not specify these relationships. Hewitt did point to a reduced turnaround time for cataloging under OCLC, mainly due to an escape from the inefficiencies of local card production but partly due to the speedier arrival of cataloging copy in useful form. He also made the significant point that the characteristics of the materials being acquired would be an important determinant of find rates and of the effects of holding patterns.[6]

Paul Metz is acting user services librarian and John Espley is automation supervisor, Cataloging Department, at the Carol M. Newman Library, Virginia Polytechnic Institute and State University, Blacksburg.

Methodology

The study was conducted at the Carol M. Newman Library of Virginia Tech (Virginia Polytechnic Institute and State University) in Blacksburg, Virginia. Newman is a medium-sized ARL library that adds about fifty thousand monographic titles per year. Newly acquired titles represent a broad range of subjects and come from a wide variety of sources. Since variations among the ways books come to the library are crucial in determining the relationships of interest, the findings will be reported in terms of the sources of receipt. Reporting in this fashion should make it possible for other libraries to adjust the findings to their own collections' patterns and thereby to generalize about their own situations.

For each of three consecutive weeks beginning in March 1979, approximately 140 newly unpacked monographic receipts were selected for the study. Serials were excluded. Selection was not strictly random, but was guided to achieve a rough match between the distribution of sampled books and the distribution of the library's annual receipts in terms of country of origin and means of purchase. As table 1 shows, American imprints, British imprints, and imprints from other nations were sampled in an approximate ratio of 4:2:1. Blanket orders accounted for half the sample, while standing orders (which are like blanket orders, but are specific to a publisher and not a dealer) and firm-order books accounted for about one-quarter of the distribution apiece.

It should be noted that the sample of firm-order materials was confined to monographs with either 1978 or 1979 dates of imprint. This decision was based on the assumption that for older materials, cataloging data would most often be available at the beginning of the test period and that if it were not, it would be unlikely to arrive during the period. One result of this decision was to focus the study quite specifically on the use of OCLC as a source of cataloging data for *current* imprints. In order to keep the distinction between firm orders and other materials as clear as possible, the study included as firm orders only those materials that wold fall outside the scope of all of the library's blanket and standing orders, either because of their subject matter or because their publishers were not covered by any of the vendors.

Trained OCLC searchers looked for copy for each item, using all reasonable access points to find cataloging copy, The results of each search were coded for one of the five categories: full Library of Congress copy, LC Cataloging in Publication (CIP) copy, "good" copy, "other" copy, and no copy found. A code of "good" indicated that copy had been contributed by a library on a list, compiled by Virginia Tech's cataloging professionals, of twenty libraries whose contributed cataloging has been of noticeably superior quality for some time and is considered less apt to need close review and revision. "Other" refers to copy from OCLC members other than the Library of Congress and "good" libraries. If multiple cataloging copy was found for a given imprint, the best data available at the time was coded, with priorities assigned in the order listed above. Only copy for the exact piece in hand was considered; in the relatively rare cases in which copy for a different edition was found but not copy for the piece in hand, the search was coded as "no copy."

Each title in the sample was searched during the week of its receipt, one week later, the next week, and then every alternate week until the sixteenth week. Searching ended only with the sixteenth week or with the arrival of full LC copy, whichever came first. After the test period had ended, the coded sheets were compiled to identify the

TABLE 1

Distribution of Sampled Monographs

	American	British	Other	Totals (Source)
Blanket order	120	57	29	206
Standing order	59	14	17	90
Firm orders	58	30	12	100
Totals (nation)	237	101	58	Grand total = 396

arrival dates for the first copy found and for the best copy ultimately found. Cumulative statistics were also kept for incidents in which copy was "upgraded," with copy being supplanted by other copy higher in priority.

General Findings

Before considering the arrival of cataloging copy and the effects of various library policies, it might be useful to make some general observations about the frequency with which a library like Virginia Tech's can expect to find useful OCLC copy for various categories of materials. The data showed that OCLC is a highly productive tool for the distribution of cataloging copy. Some copy for the full piece in hand was found within sixteen weeks for 87.1 percent of the books. Full Library of Congress copy was available for 59.3 percent of the sample. These results are displayed by category of materials in tables 2 and 3. As table 2 shows, copy is almost invariably present for American imprints and for firm-order materials. Copy is least likely to be found for British and other blanket orders; thirty-four of the fifty-one cases without copy, or two-thirds, came from these two categories.

The distribution of full LC copy (table 3) shows the same general pattern as the distribution of any found copy, except that the gaps between the success rates for American versus British and other imprints and for firm-order materials versus the other two sources widen. While full LC copy is available within sixteen weeks for about three-fourths of American imprints, it is available for only about one-third of the rest. And while full LC copy is obtained for 81 percent of the firm orders, it is found for only about half of blanket and standing orders. Whereas there is more overall copy for standing orders than for blanket orders (table 2), full LC copy is more frequently found for blanket orders; this difference is due to the very low incidence of LC copy for foreign standing orders.

The difference between the overall rate of 87.1 percent and the 59.3 incidence of full LC copy is of course accounted for by those cases where the best available copy came from "good" or "other" libraries, or represented CIP data that had not been upgraded. Table 4 shows the distribution of the best copy that had been found within sixteen weeks across the five categories. The table seems to suggest two conclusions for an OCLC member library. The first is that the availability of member-contributed (non-LC) copy, which for many members is a prime motivation for joining a network, is substantial: nearly 22 percent of materials would have no copy at all except for the contributions of members other than the Library of Congress. A second, more tentative conclusion is that the maintenance of a "good" list is more trouble than it is worth. Only 8 percent of best copy came from "good" libraries. Whether such a list is worth keeping depends on how much less review a library gives to cataloging copy from highly regarded members, and on how difficult it is to train searchers to recognize

TABLE 2

Percentage of Materials for Which Copy Found within Sixteen Weeks, by Category

	American	British	Other	Totals (Source)
Blanket order	98.3	68.4	44.8	82.5
Standing order	93.2	78.6	76.5	87.8
Firm orders	98.3	93.3	91.7	96.0
Totals (nation)	97.0	77.2	63.8	Grand total = 87.1

TABLE 3

Percentage of Full LC Copy, by Category of Materials

	American	British	Other	Totals (Source)
Blanket order	76.7	29.8	10.3	54.4
Standing order	66.1	7.1	11.8	46.7
Firm order	91.4	66.7	66.7	81.0
Totals (nation)	77.6	37.6	22.4	Grand total = 59.3

TABLE 4

DISTRIBUTION OF BEST COPY FOUND (SIXTEEN WEEKS)

	Full LC	CIP	Good	Other	(Member Subtotal)	None
Number	235	24	28	58	(86)	51
Percentage	59.3	6.1	7.1	14.6	(21.7)	12.9
Percentage of materials with copy	68.1	7.0	8.1	16.8	(24.9)	NA

the symbols of all approved libraries and to give their copy special treatment. The advice of Hogan in *OCLC: A National Library Network* supports the view that the categorical distinction between "good" and "bad" libraries is not worth making.[7]

HOLDING PATTERNS AND THE TIMING OF COPY AVAILABILITY

As noted, the key purpose behind this study was to provide information useful in determining holding patterns for the various categories of materials, so that a balance could be achieved between minimizing the number of searches for copy and making materials available as quickly as possible. For this purpose, the emphasis must not be on what type of copy is available, but rather on when it appears.

Taken together, tables 5 and 6 show that while some copy is available for two-thirds of materials as soon as they arrive, the most desirable copy, full LC, is immediately available only about 18 percent of the time. In fact, the only category for which full LC copy is immediately available more than half the time is American imprints ordered on a title-by-title basis. For both LC full copy and for copy in general, rates of immediate availability are far better for American imprints and for firm orders than for other materials.

If a library considers CIP copy to be nearly as good as LC full copy (in other words, if it considers the effort of supplying missing data preferable to extended waiting), rates of immediate availability are greatly improved, especially for American imprints. Table 7 shows the rates of immediate availability for any LC copy, whether full or CIP.

The data shown so far suggest that not all materials need to be put into a holding pattern. Copy is immediately available for a significant proportion of materials in some categories, such as firm orders. American imprints would also be such a category, if a library were to decide to accept CIP copy when available. Such a decision would have significant consequences, since CIP constitutes such a large percentage of the immediately available cataloging copy. Only 31.3 percent of the exact LC copy that was available at the end of the test period had been there from the beginning, while 44.7 percent represented upgrades of CIP copy that was extant at week one. The decision to accept CIP copy makes an immediate

TABLE 5

IMMEDIATE AVAILABILITY OF COPY, BY CATEGORY OF MATERIALS (PERCENTAGE)

	American	British	Other	Totals (Source)
Blanket order	89.2	29.8	3.4	60.7
Standing order	69.5	35.7	52.9	61.1
Firm order	93.1	76.7	58.3	84.0
Totals (nation)	85.2	44.6	29.3	Grand total = 66.7

TABLE 6

IMMEDIATE AVAILABILITY OF FULL LC COPY, BY CATEGORY OF MATERIALS (PERCENTAGE)

	American	British	Other	Totals (Source)
Blanket order	5.8	0.0	0.0	3.4
Standing order	13.6	0.0	5.9	10.0
Firm orders	69.0	43.3	33.3	57.0
Totals (nation)	23.2	12.9	8.6	Grand total = 18.4

TABLE 7

IMMEDIATE AVAILABILITY OF FULL LC COPY OR CIP, BY CATEGORY OF MATERIALS (PERCENTAGE)

	American	British	Other	Totals (Source)
Blanket order	80.8	3.5	0.0	48.1
Standing orders	49.2	7.1	17.6	36.7
Firm orders	84.5	46.7	33.3	67.0
Totals (nation)	73.8	16.8	12.1	Grand total = 50.3

search for copy for many materials much more attractive and may help to reduce in-process time significantly. The data showed that waiting for CIP copy to be upgraded can introduce a significant delay. For the 105 books for which CIP copy was ultimately superseded by full LC copy, the latter was typically not available until the sixth or eighth week. Moreover, there were twenty-four additional cases where CIP was still the best available copy after the entire sixteen weeks of the study had expired.

The Virginia Tech library has accepted the conclusions of this study and has instituted a policy of immediate searching for copy for all monographs obtained on firm order or through American blanket or standing orders. As expected, this change has resulted in a reduction of about one-half in the proportion of monographic titles going into a holding pattern. Public service librarians have expressed strong approval of the new policy. As a necessary part of the new plan, searchers have been trained and authorized to upgrade CIP records by supplying collation and other omitted data. This has represented a modest addition to their workload, but an efficient reduction in the load of work previously performed by library assistants.

It should be noted that in deciding to use CIP data, as upgraded by its own clerical staff, the library has made a judgment that the demands of efficiency and prompt user availability justify some possible sacrifice in cataloging data. Differences between CIP and final LC cataloging often involve more than simply the collation portion of the record. Dowell has pointed out that about two-thirds of CIP copy is ultimately changed by LC, that the mean number of changes per CIP title is about 1.2, and, most important, that about one CIP title in four will generate subsequent differences in final LC cataloging that could be called "significant." Significant changes include differences in main entry, title, series, subjects or other added entries, ISBN, or call number. Many, but by no means all, of the differences that fall into these categories could be expected to affect user access, according to Dowell.[8] According to a recent survey of libraries participating in OCLC, the majority of libraries have decided to delegate CIP upgrading to nonprofessional staff.[9]

In order that individual libraries may draw their own inferences from the data and not be limited to the conclusions drawn here, the most salient data have been laid out in tables 8 and 9. In table 8, the times at which various categories of materials had any copy available are laid out in four-week intervals, beginning with the date of receipt. Summary statistics are given for each purchase source and point of origin, as well as for all materials taken together. The data can be used as the basis for determining holding patterns, though where the number of cases is small (for example, firm orders from "other" countries) the findings cannot be precise. Table 9 is analogous to table 8, but is restricted to arrival patterns for Library of Congress copy (full or CIP).

The data do appear to support a few final generalizations. The very small increase with time in the proportion of firm orders having copy helps to underscore the suggestion that these should be searched immediately and further suggests that if copy is not found, original cataloging might be called for. The significant growth in the find rate for British and other materials demonstrates that for these materials a holding pattern pays definite dividends. It is really in the categories of other blanket and standing orders that member copy is most useful, as a comparison of the data shown here with other data indicates that in these cases member copy constitutes an actual majority (63.2 percent) of the best copy available

TABLE 8

PERCENTAGE OF MATERIALS HAVING ANY COPY, BY CATEGORY AND OVER TIME

	Immediate	4 Wks.	8 Wks.	12 Wks.	16 Wks.	*n*
American BLO	89.2	95.8	96.7	97.5	98.3	120
British BLO	29.8	45.6	49.1	61.4	68.4	57
Other BLO	3.4	20.7	24.1	37.9	44.8	29
American SO	69.5	81.4	81.4	83.1	93.2	59
British SO	35.7	64.3	71.4	78.6	78.6	14
Other SO	52.9	64.7	64.7	64.7	76.5	17
American firm	93.1	94.8	98.3	98.3	98.3	58
British firm	76.7	80.0	86.7	90.0	93.3	30
Other firm	58.3	91.7	91.7	91.7	91.7	12
American total	85.2	92.0	93.2	94.1	97.0	237
British total	44.6	58.4	63.4	72.3	77.2	101
Other total	29.3	48.3	50.0	56.9	63.8	58
BLO total	60.7	71.3	73.3	79.1	82.5	206
SO total	61.1	75.6	76.7	78.9	87.8	90
Firm total	84.0	90.0	94.0	95.0	96.0	100
Grand total	66.7	77.0	79.3	83.1	87.1	396

TABLE 9

PERCENTAGE OF MATERIALS HAVING FULL LC OR CIP COPY, BY CATEGORY OF MATERIALS AND OVER TIME

	Immediate	4 Wks.	8 Wks.	12 Wks.	16 Wks.	*n*
American BLO	80.8	82.5	85.8	87.5	89.2	120
British BLO	3.5	14.0	15.8	28.1	35.1	57
Other BLO	0.0	0.0	0.0	6.9	10.3	29
American SO	49.2	54.2	55.9	61.0	72.9	59
British SO	7.1	7.1	7.1	14.3	14.3	14
Other SO	17.6	17.6	17.6	17.6	17.6	17
American firm	84.5	84.5	90.0	90.0	91.4	58
British firm	46.7	46.7	46.7	50.0	66.7	30
Other firm	33.3	66.7	66.7	66.7	66.7	12
American total	73.8	75.9	79.3	81.4	84.0	237
British total	16.8	22.8	23.8	32.7	41.6	101
Other total	12.1	19.0	19.0	22.4	24.1	58
BLO total	48.1	51.9	54.4	59.7	62.1	206
SO total	36.7	40.0	41.1	45.6	53.3	90
Firm total	67.0	71.0	74.0	75.0	81.0	100
Grand total	50.3	54.0	56.8	60.9	65.4	396

within sixteen weeks. Finally, with respect to exactly what holding pattern might be best, the data indicate that each additional four weeks of waiting pays rewards but suggest that the greatest incremental benefit comes in the first four weeks. For all categories of materials the growth in the find rate after the first four weeks is so gradual that it would be difficult to justify a researching interval of less than eight or twelve weeks. Of course, local variations in policy or in collection patterns may lead to different conclusions for other libraries. For example, a library whose jobbers were slower to deliver materials than Virginia Tech's could expect to discover that the find rate for first searches would be higher, and vice versa.

CONCLUSIONS

The most general conclusion to which this study points is that OCLC provides its member libraries access to an impressive wealth of cataloging data. For a library like Virginia Tech's, copy is available within sixteen weeks for the great majority (87 percent) of materials, while full LC copy is available for a high percentage (59 percent). Some 22 percent of the best cataloging records available for monographs comes from

members other than LC. This may be taken as one index of the value of network participation (stated otherwise, a library that does not use these records has little reason to use a utility for cataloging).

With respect to the arrival times of copy, the study shows that it is apparently in a library's best interest to search all firm-order materials immediately. Full LC copy will usually be there, and even for recent monographs there is only a fairly small likelihood that first copy or improved copy will appear during the course of any reasonable holding pattern.

It is not so clear that other materials should be searched immediately. There is only a small chance that useful copy will be immediately present for some materials, though this depends on the criteria of acceptance. A critical decision point is whether to accept CIP cataloging when it is immediately available rather than waiting for full LC copy. A library that decides to accept CIP data will probably find that an immediate search for all American monographic receipts is justified. Together with the firm orders cleared by immediate searching, these materials should bring the rate of immediate clearance up to the neighborhood of 50 percent.

It is harder to draw definitive conclusions about the optimum holding pattern for other materials or for American imprints and firm orders that are not found at first. The data, however, indicate that the most productive period for any holding pattern is the first month or so and that thereafter the hit rate will grow steadily but slowly. No doubt an asymptotic upper limit is approached at some point, but this apparently does not happen until materials have been in the holding area for quite some time.

References

1. Marion T. Reid, "Effectiveness of the OCLC Data Base for Acquisitions Verification," *Journal of Academic Librarianship* 2, no.6:303, 326.
2. Joe A. Hewitt, *OCLC: Impact and Use* (Columbus: The Ohio State University Libraries, Office of Educational Services, 1977).
3. Christian M. Boissonas, "Quality of OCLC Bibliographic Records: The Cornell Law Library Experience," *Law Library Journal* 72:80–85 (Winter 1979).
4. Cynthia C. Ryan, "A Study of Errors Found in Non-MARC Cataloging in a Machine-Assisted System," *Journal of Library Automation* 11:125–32 (June 1978).
5. R. W. Meyer and Rebecca Panetta, "Two Shared Cataloging Data Bases: A Comparison," *College & Research Libraries* 38:19–24 (Jan. 1977).
6. Hewitt, *OCLC*, p.68.
7. Allan D. Hogan, "Acceptance of Cataloging Contributed by OCLC Members," in Anne Marie Allison and Ann Allan, ed., *OCLC: A National Library Network* (Short Hills, N.J.: Enslow, 1979), p.133.
8. Arlene T. Dowell, "Discrepancies in CIP: How Serious Is the Problem?" *Library Journal* 104:2281–87 (Nov. 1, 1979).
9. Sally Braden, John D. Hall, and Helen H. Britton, "Utilization of Personnel and Bibliographic Resources for Cataloging by OCLC Participating Libraries," *Library Resources & Technical Services* 24:135–54 (Spring 1980).