Cost Savings to Canadian University and Large Urban Public Libraries from Their Use of National Library of Canada MARC Records

Jamshid Beheshti, Andrew Large, and Pat Riva

The authors present a study to determine the savings incurred by Canadian university and large urban public libraries as a result of using Canadiana printed monograph cataloging records generated by the National Library of Canada (NLC) rather than cataloging these items themselves. The study employed three methodologies: questionnaires were sent to 90 Canadian university and college libraries and to 30 member libraries of the Council of Administrators of Large Urban Public Libraries (CALUPL); follow-up telephone interviews were held with 18 university and 12 public libraries; and a sample of 100 bibliographic records for Canadiana printed documents was selected by the NLC from its catalog and then compared with records in a sample of 20 university and 10 public library OPACs to determine the extent to which NLC records form the basis for copy cataloging by other libraries. The saving per library through using NLC records as the basis for copy cataloging rather than originally cataloging items was $16,400 per annum for university libraries and $7,800 for large urban public libraries. An extrapolation to all university and large public libraries suggests an annual saving of $1,476,000 for all Canadian university libraries, and $249,000 for all Canadian large urban public libraries. Many libraries make use of NLC name or series authority data, and use NLC copy in their acquisitions processes or for other bibliographic purposes. The monetary benefits accruing to the libraries from these services and activities have not been quantified.

Cataloging with copy has become a ubiquitous process in all but the very smallest of North American libraries. Much has already been written about the rapid expansion of shared cataloging since LC card sets first appeared. Instead of undertaking original cataloging of all acquired materials, libraries now make every effort to obtain catalog records that have already been created elsewhere. These records are then integrated into the library's own database, probably with various degrees of modification to meet the specific needs of the library. One such source of cataloging records in many countries, including Canada, is the national library. Full-level source records are particularly sought as they constitute authoritative records created according to national standards,

Jamshid Beheshti is Director, Graduate School of Library and Information Studies, McGill University, Montreal, Quebec, Canada. **Andrew Large** is CN-Pratt-Grinstad Professor of Information Studies, Graduate School of Library and Information Studies, McGill University. **Pat Riva** is Cataloguing Librarian/Database Specialist, Library Technical Services, McGill University.

The authors would like to acknowledge the librarians who commented so helpfully upon the pilot questionnaire, who took time to complete the questionnaire with such care and thought, and who despite their hectic work schedules volunteered to be interviewed, whether or not they were finally included in our interview sample. Marc Richard of McGill University Libraries and France Bouthillier of McGill's Graduate School of Library and Information Studies generously gave their time and expertise to translate the English-language questionnaire and accompanying letter into French. Liz McKeen of the NLC was a source of encouragement, support, and help at various stages of the project. Finally, Amy MacLean and Valerie Nesset, our research assistants, as is often the case, did the bulk of the day-to-day work, and we are greatly indebted to them.

Manuscript submitted September 25, 2002; accepted November 26, 2002.

and carry full authority work for all access points. The National Library of Canada (NLC) generates approximately 30,000 catalog records annually, and these are made available to other libraries directly or indirectly through a variety of services. A major motivation for copy rather than original cataloging is the expectation of cost savings that the library will reap. The library generating the original record, of course, does not make any such saving from this transaction, although it may in turn obtain other libraries' cataloging records on a cooperative and mutually beneficial basis.

The NLC is a major source of cataloging records, especially for Canadiana—documents published in Canada, by Canadian authors, or about Canada. Many libraries make use of its catalog records. But the NLC itself must expend considerable resources in producing these records. In late 2001 the NLC contracted the Graduate School of Library and Information Studies, McGill University, to determine the dollar value of savings incurred by two types of Canadian libraries—university and large urban public—as a result of using Canadiana printed monographic and federal government cataloging records generated by the NLC rather than cataloging these items themselves. Other types of libraries and materials were excluded, for the time being at any rate, to make the study manageable and realizable given the time and financial resources available. The study was conducted between January and March 2002.

Previous Studies

A number of studies have been undertaken in various countries to identify cataloging costs in general, and more specifically to compare copy cataloging with original cataloging costs. A recurrent theme in these studies, however, is the difficulty of establishing a satisfactory methodology that meets the four criteria established by Orr (1973) for measuring library services: reliability (identical results will be generated from identical situations); validity (appropriate for the situation); precision (capable of taking into account all relevant internal and external factors that might modify the results); and feasibility (can be undertaken with the kind of human and other resources available in a library). Lancaster (1973) points out that cost analyses of library technical processing generally suffer from two limitations: it is not clear exactly how data were derived, and there are no generally accepted standards for what should be measured and how costs should be derived and presented.

In response to the need for standardization to facilitate inter-institution comparisons, and the importance of basing local management decisions on sound cost analyses, the ALCTS Technical Services Costs Committee (1991) created a comprehensive checklist and formula to "help the technical processing manager determine the unit cost for any acquisitions or cataloging function." (49) However, many published cost studies predate the ALCTS guidelines, and as Morris et al. (2000) have more recently stated: "the literature on cost studies for technical services operations is extensive . . . but for the most part it is fragmentary, limited in scope, and short on detail." (70)

Deriez and Giappiconi (1994) provide an interesting discussion of methodological problems and possible solutions. Cost calculations reported in the literature may be based on a calculation of direct costs only, or include certain elements of indirect costs and overhead. Although a comparison between two workflows at the same institution can be made by looking only at direct costs, as indirect costs and overhead likely would be the same for each workflow, comparisons between institutions generally need to look at indirect as well as direct costs. The difference in results between differing methodologies is illustrated by LC's adoption of a "full costing" methodology that is in compliance with the U.S. Federal Accounting Standards Advisory Board's "Management Cost Accounting Concepts and Standards" as decreed in July 1995 and beginning in fiscal year (FY) 1996. As described in *LC Cataloging Newsline* (1996), full costing includes direct labor costs, personal fringe benefits, and indirect costs such as salaries of office personnel, equipment, and facilities. The FY 1995 cost per record (including decimal classification) had originally been calculated as $48.34, but was recalculated as $93.19 using full costing. In FY 1996, the cost per record for full original cataloging was $107.52, and for copy cataloging, $45.15. Few studies use as comprehensive a full costing method. The Iowa State University longitudinal study is one of the most comprehensive, but focuses on personnel time and cost; costs for equipment and facilities are not included. In Morris (1992), the per-item cataloging costs (for 1989/90) of $9.02 for copy, and $32.99 for original, are calculated with such "overhead" items as administrative tasks, staff participation in nontechnical services tasks, and vacations, holidays, and sick leave; but without other unavoidable associated tasks such as "training, procedure and policy documentation, revision, or separate authority work activity."

Many studies have been confined to an investigation of a single library's operations, mostly academic libraries, and frequently ARL members. U.S. libraries are represented in Leung (1987) (University of California, Riverside), El-Sherbini (1995), Rider and Hamilton (1996) (both Ohio State University), Morris (1992), and Morris et al. (2000) (Iowa State University). These deal either specifically with monographs or blend costs for all types of materials. As Osmus and Morris (1992) and Morris et al. (2000) point out, however, serials cataloging is far more expensive than monograph cataloging, taking one-third more time per title.

Blended costs mask this because the proportion of cataloging that is serials cataloging is generally low. In the Iowa State study, Morris et al. (2000) report copy cataloging costs per title that were appreciably lower than original cataloging costs (in 1989–90, $9.02 and $32.99 respectively, and in 1997–98, $12.22 and $88.24 respectively). The copy cataloging is broken down into 60% DLC/CIP and 40% member records (in 1997–98).

Rider and Hamilton (1996) at Ohio State University examined the cost savings when using the OCLC PromptCat service as a distribution vehicle for cataloging copy. PromptCat was able to supply a record for all 200 books in the sample, using 65% CIP, 25% full DLC, 8% member records, and 2% UKMARC.

Actual cost figures are highly sensitive to technology and procedural factors. An example is the study by Jenda (1992), carried out from 1985 to 1987 at the University of Botswana which was using a card catalog at that time, to estimate costs incurred when cataloging a title using Library of Congress card sets compared to the cost of original cataloging (the former was 40% less than the latter). The cost comparison included both staff time and the cost of materials and services.

Several studies have attempted to gather data more widely. Kantor (1986) took a detailed look in 1984 at the costs of choosing, ordering, and cataloging monographs in 8 U.S. academic and major research libraries. He noted significant differences in the average performance of the libraries, but overall found that original cataloging was more than three times as expensive as copy cataloging. McCain and Shorten (2002) conducted a survey of ARL libraries, based on FY 1998/99. Statistics gathered were extensive, but only 27 (including 2 Canadian) of 111 libraries (24%) were able to respond. As the focus was on defining "best practices," taking both efficiency and effectiveness into account, per-item cost figures for copy and original cataloging were not reported. In assessing the benefits of BIBCO for LC, Wiggins (2000) calculates that LC was able to use 5,585 records created by BIBCO libraries between October 1995 and September 2000. Taking into account the LC cost per record for full original cataloging ($138 in 2000), and also including the tasks that LC still needed to perform to add the records, Wiggins estimated a saving of $577,377 for LC. This figure suggests that the aggregate savings for many libraries due to copy cataloging could be substantial.

The Western Australian Group of University Librarians (WAGUL), with four members, undertook in 1996 a Collaborative Cataloging Project to look into cataloging operations (Wade and Williamson 1998). Specifically, information was sought on original and copy cataloging costs. Average overall cataloging costs ranged between the four libraries from $23.11 to $37.06, but when weightings were introduced to reflect variations between different kinds of cataloging (original, clone, difficult copy, copy, and additional copies) the new range was between $14.25 and $21.90. Among the member libraries full original cataloging accounted for only 8% to 12% of all titles.

A few studies of Canadian academic libraries have focused upon or touched upon cataloging costs: Oldfield (1987) (University of Waterloo), Carter (1997) (University of Alberta), and Partington and Talbot (1997) (University of Manitoba). None specifically indicates the proportion of NLC records among the sources of copy. Oldfield (1987) reports unit costs for 1984–85 in four categories: copy with MARC records, $6.12; copy (i.e., "manual copy"), $13.27; original, $21.70; and abbreviated, $3.29. In volume, the manual copy represented less than 5% of the copy cataloging total, but was still significant enough to track separately. In later studies, manual copy has disappeared as a category; all copy is assumed to be derived from MARC records.

Methods

Unlike such detailed case studies of specific libraries, the study reported in this article sought an aggregate response to the question of cumulative savings due to the use of NLC MARC records as a source of copy for cataloging. This aggregate figure is arrived at without comparing details of procedures, practices, policies, and technology available at each participating library. In addition, by applying the same methodology to two quite distinct types of libraries—academic and large urban public—some observations relating to the similarities and differences between the two groups can be made. Of existing studies, only Deriez and Giappiconi (1994) considered the case of public libraries, which they felt to have quite different collection profiles and cataloging priorities than university libraries. In particular, Deriez notes that academic libraries are co-contributors to shared cataloging with their national bibliographic agencies to a greater extent than public libraries, a factor that increases their costs.

Three methodological approaches were used to determine cost savings. These approaches were deemed to be the most effective and efficient methods of collecting information given the financial and time constraints confronting the researchers.

- Questionnaires were e-mailed to Canadian university and college libraries (henceforth called simply "university libraries") identified in the *Directory of Canadian Universities,* and to member libraries of Canada's Council of Administrators of Large Urban Public Libraries (CALUPL).
- Follow-up telephone interviews were conducted with a sample of respondents to the questionnaire.

- A sample of NLC records was matched against the holdings in a sample of university and large urban public library OPACs to determine the percentage of NLC records in these OPACs and the proportion of the records that have been copy cataloged using the NLC records.

Questionnaire

The questionnaire was intended to determine how many cataloging records are received annually by the target libraries, what proportion are copy cataloged rather than originally cataloged, and the sources of the former. It asked questions about the typical cost incurred in copy cataloging a record compared with undertaking original cataloging. It also offered an opportunity for librarians to comment upon NLC's cataloging service.

The questionnaire contained 14 closed questions (two questions included both an *a* and a *b* part), one open question, and one invitation to add any comments whatsoever about NLC cataloging policies and procedures. A draft English-language version of the questionnaire was piloted in 1 university and 1 public library. A copy also was sent to the NLC for feedback. The questionnaire was then modified in light of the pilot and NLC comments, and was translated into French. A second translator checked the French translation. An explanatory letter to accompany the questionnaire was also developed and translated into French.

Bearing in mind that for their own management purposes most libraries would not be tracking the detailed statistics that would enable an easy and precise answer to the research question, respondents were encouraged to supply either "actual" or "estimated" figures for many questions, qualifying them as such. Additionally, statistics were requested for the most recent completed fiscal year, as the exact months covered were not relevant to the result. No specific formula for calculating per-item costs was prescribed, but respondents were expected to use the same method for calculating both copy and original cataloging costs. These features were intended to allow libraries to participate with a minimum of recalculation of their in-house statistics and were certainly factors leading to a higher response rate than seen in other similar studies (for example, Bedford (1989) sent a survey instrument to 26 large academic research libraries but received full data sets only from 4 of them).

A list of 92 Canadian universities was obtained from the *Directory of Canadian Universities.* In 4 cases the libraries were affiliated with other libraries listed, and the questionnaire was therefore e-mailed to 88 university libraries. A list of the 32 members of CALUPL was also obtained from CALUPL itself, and the questionnaire was e-mailed to all of them with the exception of 2 libraries from which the message bounced back due to e-mail address problems. The e-mail therefore was received by 118 libraries (see appendix A for the English-language version). Two reminder e-mails were sent to all nonrespondents, and a copy of the questionnaire was attached to the second e-mail in case the original had gone astray.

Telephone Interviews

At the end of the questionnaire, respondents were asked to indicate whether they were willing to participate in a short follow-up telephone interview. Forty questionnaire respondents agreed to be interviewed. From these, 30 (75%) were selected for interview (18 university libraries and 12 public libraries). Selection was based upon the desire to represent technical services departments in different regions of the country, of varying sizes, both independent institutions and consortium members, and supporting French, English, and bilingual catalogs. The interviews were conducted with the technical services librarian of each library, 4 in French and 26 in English. Interviews lasted between 20 and 45 minutes. The primary purpose of the interviews was to verify and authenticate answers provided in the questionnaires, especially relating to cataloging costs. In addition, a preliminary analysis of the returned questionnaires identified a number of general questions to ask relating to procedures, record sources, and NLC services. All interviews were conducted by the same member of the research team to ensure consistency. The interviewer took detailed notes, but the interviews were not taped.

Record Matching

The record-matching process was intended to provide more reliable quantitative data than the questionnaires, but using only a small sample of NLC records and a subset of university and large urban public libraries. The intention was to extrapolate the results from the sample population to the entire population. Its objective was to determine what percentage of Canadiana titles cataloged by NLC are to be found in the sample libraries, and what percentage of these are cataloged using copy from NLC records.

NLC was asked to select a small sample of records for Canadiana printed monograph titles cataloged by it in 1999. In discussion with the NLC it was agreed to obtain two samples, one of federal government documents and the other of commercially published fiction and nonfiction monographs; all other types of publications were excluded. The initial NLC sample comprised 105 discrete records: 35 for government documents and 70 for nongovernment documents. Where bilingual publications were cataloged in both English and French (10 records representing 5 discrete government

documents), both records were included in the sample. The "E" and "F" suffixes in the 016 field denote these in the MARC records. For the purposes of matching the records against the libraries' OPACs, however, only one version of the bilingual record pair was used. The final record collection to be used in the matching process therefore comprised 100 records: 30 government and 70 nongovernment. The records were supplied by the NLC to the project team in full MARC format as well as in ASCII format.

The size of the library sample was largely determined pragmatically by the time available to conduct the record-matching procedure. An average of three to five minutes was estimated to compare each of the 100 sample records against the OPAC in each of the sample libraries. It was decided that a sample of 30 libraries (university and public) would be manageable, given the time and resources available to the research team.

Each of the public library Web sites was examined to determine if it offered a Web-accessible OPAC. Only 11 provided publicly accessible MARC records, and 2 shared the same catalog. This left therefore 10 public library catalogs providing accessible MARC records, and it was decided to include all of them in the sample. The university library OPACs were examined to eliminate those libraries that did not publicly display records in MARC format. This reduced the number of eligible institutions to 71. When several shared a common OPAC (e.g., NOVANET serves university libraries within Nova Scotia), one catalog was chosen to represent the consortium, as catalog practice and software constraints would be similar for all institutions belonging to the same OPAC network. This reduced by 30 the number of eligible libraries. A random sample of 20 then was selected from the remaining 41 university libraries, to give the total of 30 OPACs for record sampling that had been determined as manageable (10 public and 20 university).

Each sample MARC record supplied by NLC was matched against the MARC records in the chosen university and public library OPACs using 3 distinct access points. An initial search was carried out on the ISBN. If a match was not found, the title and personal name fields were then searched. In the case of government documents, the NLC record number was searched if the OPAC offered this search key, but this rarely was the case. When the record was found in the OPAC, the MARC format was examined to see if the library's record was derived from the NLC record in the sample. In MARC21 field 040 (Cataloging Source), either subfields *a* (Original cataloging agency) and *c* (Transcribing agency), or *d* (Modifying agency) were expected to contain the MARC21 code for NLC (CaOONL) or the OCLC participant code (NLC). However, lack of field 040 does not conclusively indicate that the record was produced by original cataloging, as some redistribution vehicles for NLC records do not retain field 040. In all cases, MARC21 field 016 (National Bibliographic Agency Control Number) subfield *a* (or MARC21 field 015 in the case of OCLC records) had to match with the 016 from the sample record to be counted as an exact match. For bilingual records, target libraries were expected to hold only one record of each pair for a full match. A true match was counted only if the library held the exact item represented by the target record (for example, the precise edition).

Data Analysis

This article presents the data collected only in so far as it relates to the question of cost savings. Both the questionnaires and telephone interviews collected librarians' opinions about the NLC's cataloging service (in general, very positive) together with some suggestions for enhancements. These opinions, however, have not been included here.

Questionnaires

Of the 118 questionnaires mailed, 69 (58% response rate) were returned, 48 from university libraries (55% response rate) and 21 from public libraries (70% response rate). While all returned questionnaires were valid and could be analyzed, some respondents did not answer all the questions. Answers from the 14 closed questions were entered into SPSS. The answers to the open questions, along with any general comments added at the end of the questionnaire, were assigned to subject topics by two members of the research team.

Table 1 shows the quantitative data collected from the questionnaires. The diversity of the libraries in terms of their size and, to a lesser extent, their mandates results in a wide range of data. The collections of the public libraries in the sample are relatively large, with an average of more than 333,000 (median 295,000) printed monographic titles. The university libraries are much more varied in collection size, ranging from 15,000 to more than 2,000,000 printed monographic titles, with an average of 572,000 but a median of 240,000.

The number of titles cataloged in the last fiscal year (in almost all cases 2000–01) on average was close to 14,000 across all the libraries, but there were marked differences between the university and the public libraries. The mean for the former is around 11,000 whereas for the latter it is almost 20,000. The gap between those libraries undertaking a lot of cataloging and those undertaking little is great: the number of titles cataloged in the last fiscal year ranged from just under 200 to 45,000.

Copy cataloging is a common practice among the libraries, with a mean of 85% of all cataloging being copy. The mean and median figures for copy cataloging are very similar (85% and 90% respectively), and there is little difference between the university and the public libraries in this respect. However, NLC-derived copy only constitutes a small part of the cataloging, although there are differences depending upon whether the copy relates to monographs or government documents. The mean for all copy cataloging derived directly from NLC records for monographs is 14%, of which 3% is for federal government documents. The use of NLC copy for monographs is little different between the university and public libraries, but public libraries are less inclined than university libraries to use NLC copy for federal government documents. As may be expected, the proportion of Canadiana publications that are not copy cataloged by the libraries is very low, at around 5%, with the university libraries handling more original than the public libraries. A large majority of libraries (87%) stated that these data represented a "typical" fiscal year.

Determining their costs of copy and original cataloging proved to be more problematic than assembling catalog production figures for the responding libraries. Only 4 reported the "actual" costs of copy cataloging, ranging from less than $1 per record to $50 per record (these and all subsequent figures refer to Canadian dollars). For the purpose of statistical analysis, these "actual" costs were combined with the estimated costs submitted by the other libraries. The average cost per document is slightly higher than $13 dollars for copy cataloging and almost $31 for original cataloging (with medians of $10 and $26.50). However, large differences are reported by the two categories of library. University libraries give an average cost of almost $17 for copy whereas public libraries only report just over $8 per record; in the case of original cataloging, again the discrepancies are large, with almost $38 for universities but less than $19 for public libraries.

Many libraries use NLC name or series authorities—47% indicated "occasional" use, while 35% are "frequent" users. Only 18% "never" use this particular service by NLC.

Libraries on average wait about 3.5 months for a record to become available before undertaking original cataloging. The length of time, however, differs significantly between university and public libraries ($t=3.23$, $df=56$, $p=0.002$). While university libraries on average may wait 4.8 months (median of 3 months), public libraries are willing to wait for only 1.5 months (median of zero).

Analysis of variance shows that cataloging costs for both the copy and original are dependent on the type of library ($F=5.154$, $df\ 1,36$, $p=0.029$; $F=6.457$, $df\ 1,30$, $p=0.016$). Since in many cases the data are highly skewed, the median may be a more accurate measure of the central tendencies than the arithmetic mean. Figures 1 and 2 show the median costs of copy and original cataloging for university and public libraries. The graphs illustrate the considerable differences between university libraries and public libraries in both costs. Further analysis based on nonparametric statistics and medians also confirms these results. Both the Kruskall-Wallis test and the Median test show significant differences between university and public libraries in terms of the cost of cataloging ($p<0.05$). They strongly suggest that in calculating cataloging costs, university and public libraries should be treated separately.

As figures 3 and 4 indicate, the cost of original cataloging for university libraries ranges from $2 to $100 with a mean of $38, while for public libraries the range is from $2 to $35 with a mean of $19. Q-Q probability plots were used to determine whether the distribution matched a normal distribution. As a result, in the case of the university libraries one outlier was eliminated and the cost figures were recalculated. The average cost of original cataloging for university libraries is $37.6 (median of $30) with a 95% confidence interval of $25 to $44. The figure for public libraries remains the same at $19 (median of $15.50), with a 95% confidence interval of $11 to $26.

Figures 5 and 6 show the estimated costs of copy cataloging for university and public libraries respectively. The cost for university libraries ranges from $1 to $50, whereas for public libraries it is from $1 to $30. Q-Q probability plots were used to eliminate the outliers. The mean cost of copy cataloging based on the modified data for university libraries is $16.9 (median of $13) with a 95% confidence interval from $9 to $18, and for public libraries $8.3 (median of $6) with a 95% confidence interval from $4 to $11.

Differences in currencies, fluctuating exchange rates, labor costs, and inflation prevent any direct comparison of the absolute costs of cataloging reported here with those provided in the previous studies. The ratio of copy to original cataloging, however, indicates that the results of this study are comparable to those reported in the literature. The cost of copy cataloging is 41% of the cost of original cataloging for university libraries and 37% for public libraries.

Cost Savings for Libraries

The potential cost saving for the libraries is calculated as follows (figures are rounded to the nearest one-hundredth dollar): total number of published print monographic titles in collection cataloged in last fiscal year, multiplied by percentage of published print monographic titles in collection that were copy cataloged using NLC bibliographic records in last fiscal year, multiplied by cost per bibliographic record for full original cataloging of published print monographs, minus cost per bibliographic record for copy cataloging of published print monographs.

Table 1. Questionnaire Data Analysis Overview

Questions	All Libraries			University			Public	
	Mean	Median	N	Mean	Median	N	Mean	Median
Monographic title holdings	501,272	269,422	45	572,033	240,000	19	333,679	295,000
Titles cataloged (last fiscal year)	13,948	10,853	45	11,269	6,773	20	19,976	16,751
Total titles copy cataloged	85%	90%	44	84%	87%	21	88%	90%
Total titles copy cataloged using NLC	14%	10%	36	14%	11%	16	14%	10%
Gov. titles copy cataloged using NLC	3%	1%	33	4%	2%	15	1%	1%
Canadiana titles cataloged originally	6%	4%	28	7%	5%	15	3%	2%
Cost per record—copy cataloging	$13	$10	23	$17	$15	19	$8	$6
Cost per record—original cataloging	$31	$26.50	21	$38	$30	12	$19	$15.50
Months will wait for copy to be found	3.5	3	40	4.8	3	19	1.5	0

N=number of libraries answering the relevant question

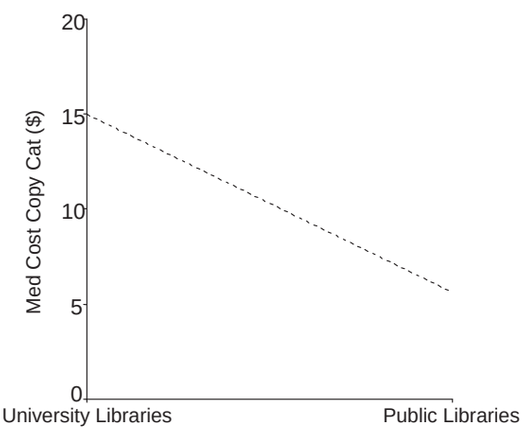

Figure 1. Median Cost (CAN $) of Copy Cataloging for University Libraries and Public Libraries

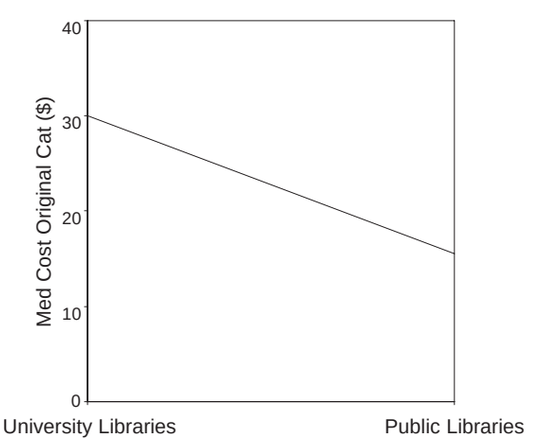

Figure 2. Median Cost (CAN $) of Original Cataloging for University Libraries and Public Libraries

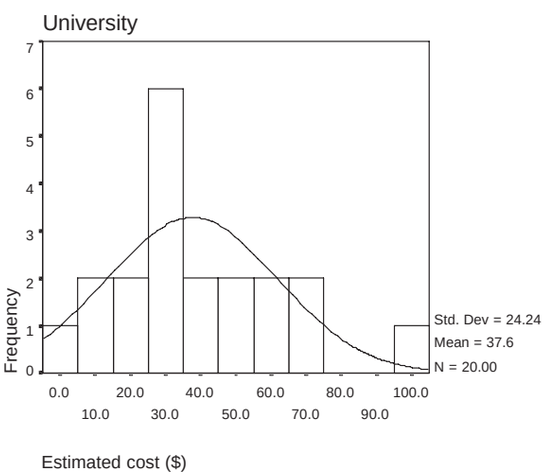

Figure 3. Cost of Original Cataloging for University Libraries

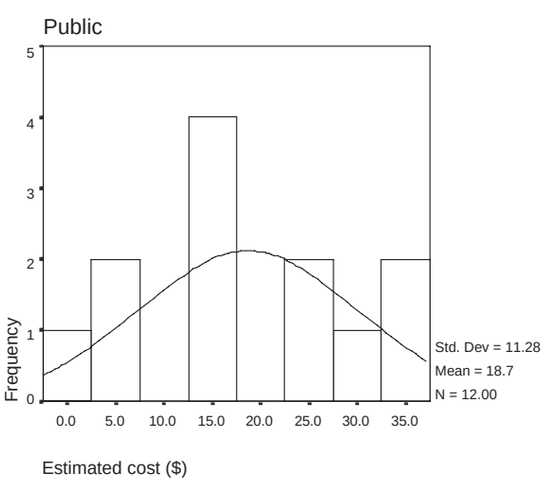

Figure 4. Cost of Original Cataloging for Public Libraries

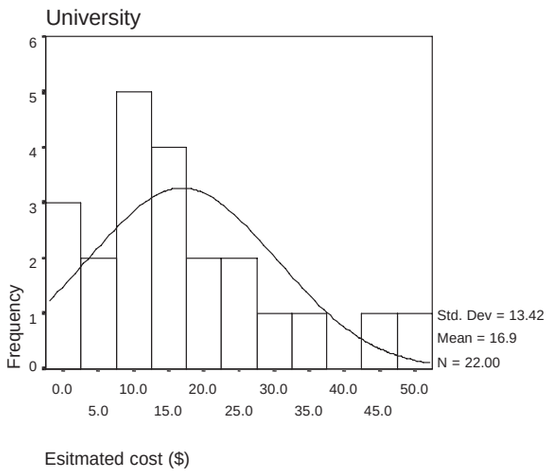

Figure 5. Cost of Copy Cataloging for University Libraries

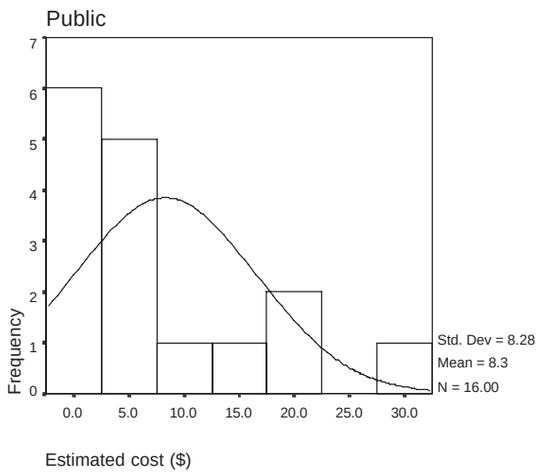

Figure 6. Cost of Copy Cataloging for Public Libraries

The total cost savings for the 20 (22% of total population) university libraries that responded to *all* the questions pertaining to the formula is $572,800. The mean cost saving for a university library is $28,600 (median of $20,700) with a 95% confidence interval of $9,700 to $47,600. As figure 7 shows, however, several outliers are skewing the data, which were then removed with Q-Q probability plots. The modified mean cost saving for an academic library based on 18 responses is then $16,400 (median of $16,500), with a 95% confidence interval of $8,700 to $24,200.

If we extrapolate from these figures, the mean total cost savings for university libraries in Canada as a result of using NLC MARC records is $1,476,000. A 95% confidence interval around the mean ranges from $783,000 to $2,178,000.

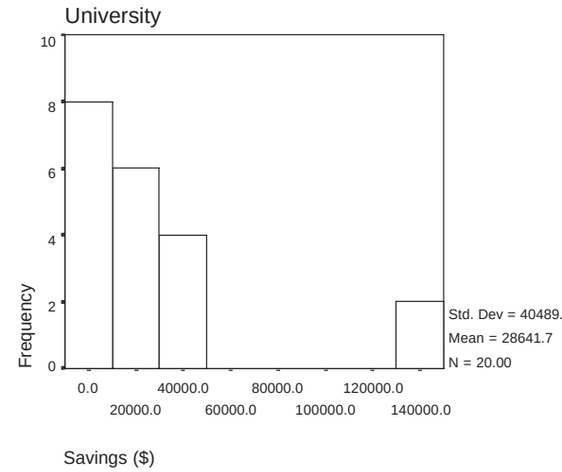

Figure 7. Cost Savings by University Libraries

The cost savings for the 9 (28% of total population) public libraries that responded to all the questions pertaining to the formula is $137,600. The average annual saving for a public library is $15,300 (median of $6,100), with a confidence interval of zero to $33,400. Figure 8 shows that one response is an outlier. Recalculating the mean without this outlier results in an average cost saving for public libraries of $7,800 (median of $5,400). The 95% confidence interval around the mean is from $1,200 to $14,500. The wide range of the confidence interval reflects the small sample size and the diversity of the responses.

If we extrapolate from these figures, the mean total cost savings for large urban public libraries in Canada as a result of using NLC MARC records is $249,600, with a 95% confidence interval of $38,400 to $464,000.

Bibliographic Sources

Figure 9 shows the bibliographic sources used by all the libraries for cataloging Canadiana (n = 68). The most frequently cited *single* source is Amicus Online (75%). Web OPACs including Z39.50 servers are used by 76% of libraries, followed by printed CIPs (56%). These data demonstrate that the majority of libraries opt for those bibliographic sources that are free of charge and readily available. The most cited commercial sources are AG Canada (40%) and OCLC Online (35%). Amicus, alone or in combination with one or more other sources, is by far the most used bibliographic resource for cataloging Canadiana; more than 33% of libraries indicated that they rely on it.

As more than half of respondents had reported using CIP printed in books as a source of cataloging, the nature of that use was clarified during the interviews. Libraries were divided as to whether a record created by transcribing CIP should be counted as copy cataloging or original cataloging, but all reported that printed CIP is used much less frequently than formerly, as the MARC records generated by the CIP program normally now are rapidly distributed.

The frequency of use of bibliographic sources is relatively evenly distributed between the university and public libraries, with the exception of three sources: OCLC online is used significantly more by university libraries at 47% versus 10% in public libraries; Canadiana CD-ROM (19% versus 4%) and BiblioFile (33% versus 13%) are used more by public than by university libraries.

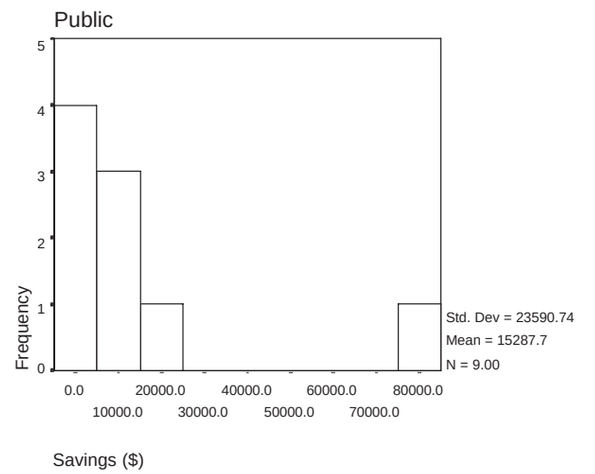

Figure 8. Cost Savings by Public Libraries

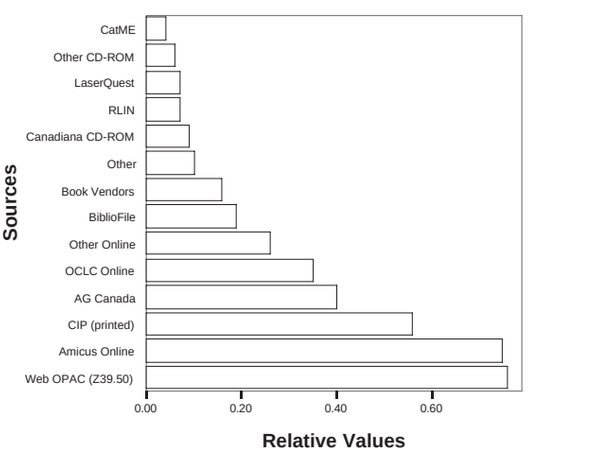

Figure 9. Bibliographic Sources for Cataloging Canadiana

Discussion of Questionnaire Analysis

The high overall response and reply rate indicates that libraries are sufficiently interested to participate in this sort of survey. However, as the follow-up telephone interviews clearly revealed, respondents were not necessarily able to provide answers to all the questions requiring actual or even estimated statistics. For this reason, the number of respondents to any one question differs; table 1 (above) shows that the total number of university and public libraries responding to the specific questions analyzed in that table ranged from 33 to 65, even though 69 libraries did return questionnaires. In practice it proved difficult to unambiguously frame questions without converting each question into an essay—which would have greatly reduced the chances of anyone returning a completed questionnaire! Some survey questions, even after pre-testing, remained open to varying interpretations by respondents. One example relates to outsourcing. Some libraries that outsource did not know whether they should report on all additions to their catalog or only their in-house cataloging. Furthermore, the original source of records supplied by the outsourcers were generally not known by the client libraries.

In some cases libraries responded to questions, particularly those dealing with the costs associated with either copy or original cataloging, with a range of figures. At the data entry stage the decision was made to enter the high figure of each range rather than to use the low figure or an average figure. This inevitably has affected the mean and median figures as calculated above. The reasons why some libraries felt the need to report ranges for cost figures has not been explored, but possible factors include:

- The question asked for all types of copy, but the library has separate figures for "good" source copy (the low end), and "partial" copy (the high end), and so reported both rather than an average.
- Copy may be divided into workflow streams by difficulty and these are handled by different levels of staff, with different salary scales (libraries specifically cited non-Roman alphabet material).
- For original cataloging, again, the difficulty of the material being cataloged may result in different levels of staff handling some categories, and these staff may have different salaries.

We found a correlation between the cost of original cataloging and the type of library (university or public). It is not surprising to find that university library costs are higher,

because of the nature of the materials being collected and cataloged by those libraries. We asked for costs for the library's overall cataloging operation, not just in regard to Canadiana, or English- and French-language monographs (our reason for this was an acceptance that libraries simply would be unable to allocate costs either for copy or original cataloging only to Canadiana in their collections). However, material requiring original cataloging in university libraries may well overrepresent other languages, rare material, and formats other than regular print monographs (serials, music, maps, electronic resources, etc.), all of which generally take longer to catalog, and may require more specialized staff. Even material handled by copy cataloging may include a significant proportion of more difficult documents and of poorer copy for non-North American imprints.

Attributing staff time specifically to cataloging also is complex whenever staff spends more than a negligible amount of time on any noncataloging duties. Copy catalogers may be involved in acquisitions, processing, database maintenance, or other tasks. The definition of which technical services tasks are an intrinsic part of the cataloging process is open to interpretation. The duties of original catalogers can also be very varied, including significant amounts of authority work, or training, or policy setting. The allocation of costs for sources of copy when those same databases are also used for ILL and reference and acquisitions is equally complex. All these factors mean that costs in practice are unlikely to be directly comparable between institutions.

In using the mean cost for all types of copy cataloging in our calculation, we are not making any allowance for the fact that some forms of copy cataloging are significantly less costly than other forms. The staff time and knowledge needed to complete the processing varies a great deal for different forms of copy cataloging. Specifically, using full source records found in the first source searched requires the least time and the least checking, while resorting to an incomplete record (particularly one that lacks elements of subject analysis) after a number of searches costs the most. Using full-level NLC source bibliographic records involves less expensive workflow, while alternative sources of copy would tend to be more expensive.

Timeliness in the availability of records also produces a saving for libraries, but this figure is difficult to calculate. If a full record that can be used for copy cataloging is found in the first source checked, then the library saves the staff time and per-search costs that would be incurred in searching through multiple sources for the record. As we elicited only a global average for copy-cataloging costs, we cannot put a figure on the cost savings resulting from the early availability of records for use in copy cataloging.

For libraries using copy in their acquisitions processes, there is an additional financial saving as the availability of records reduces the cost of inputting bibliographic information for ordering purposes. In addition, libraries are realizing savings through other uses of bibliographic records and through their use of authority records, which could not be quantified by the data collected in this study.

As a result of these considerations, the cost savings estimate as calculated from the questionnaire data may represent only a portion of the impact that the use of NLC bibliographic and authority records has on the respondents' total savings.

Record Matching

Since libraries were unsure about the exact amount of copy cataloging for which NLC MARC records were used, the record-matching procedure was intended to provide an alternative means to estimate use of NLC-derived cataloging records and the resulting cost savings to libraries. Figure 10 shows the distribution of NLC records in the 30 OPACs examined. Of the total 100 NLC records, the mean percentage hits per library is 4%, with a range for all libraries in the population between 2% and 6% in 19 out of 20 instances. If these results are extrapolated to the population, then we may conclude that the mean number of NLC hits per library is 1,200, with a 95% confidence interval of 600 to 1,800 (this wide range is a consequence of the relatively small sample size and the wide quantitative range of responses from the sample).

The average number of hits for NLC monographic records in the university libraries (7%) is significantly higher than in the public libraries (2%). Similarly, the average number of hits for NLC government publication records in university libraries (11%) is higher than in public libraries (2%).

Figures 11 and 12 show the mean number of NLC sample records found in the 30 OPACs. Of the randomly chosen sample of 70 monographic records, on average only 2 NLC records are found per library. An equal number (2 per library) is found with MARC records derived from other sources. In the public libraries, however, the number of hits for non-NLC-derived records is 2.2 per library versus 0.5 for NLC-derived records (figure 11). For government publication records, the average number of hits for academic libraries is 3.8 for non-NLC records and 3.4 for NLC copy records per library. For public libraries the average number of hits for government records per library are 1.3 for non-NLC and 0.7 for NLC-derived records (figure 12).

The mean number of documents per university library for which the copy cataloging were derived from NLC was reported in the questionnaires to be 1,203 with a 95% confidence interval of 721 to 1,684. The record-matching data show that on average 5% of the records (95% confidence interval of 4% to 7%) in university OPACs are derived from

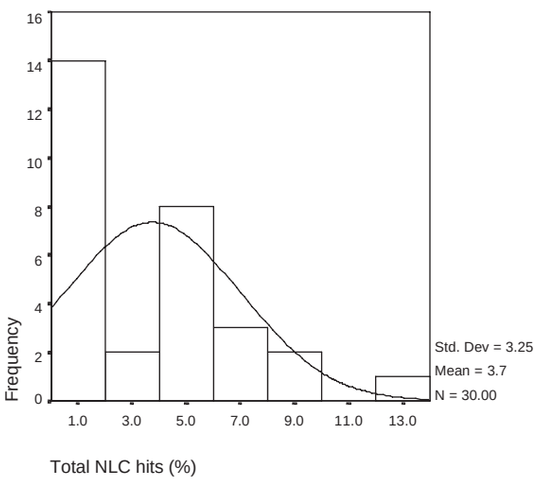

Figure 10. Percentage of NLC Sample Hits in 30 OPACs

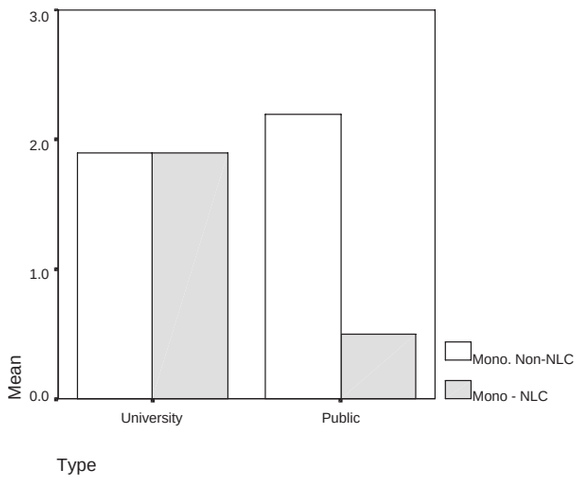

Figure 11. Mean Number of Hits in 30 OPACs-Monographs

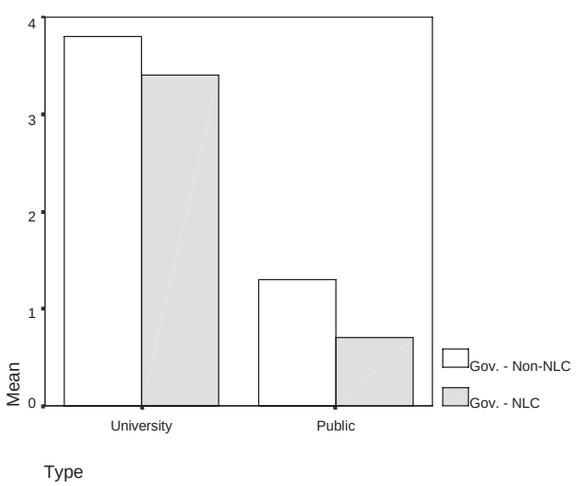

Figure 12. Mean Number of Hits in 30 OPACs-Government Publications

the NLC. Converting the average from percentage to an absolute number using the approximate figure of 30,000 documents cataloged by NLC per year yields 1,500 records. Therefore, the data provided by university libraries in their questionnaires approximately matches the result calculated from the record matching. The mean number of documents per public library with NLC-derived cataloging records was reported in the questionnaires to be 2,109 with a 95% confidence interval of 1,223 to 2,995. Record-matching data, however, indicate that only 1% of the records (95% confidence interval of 0.3% to 2%) in public libraries are derived from NLC. When this figure is converted to absolute numbers, the mean number of records derived from NLC by public libraries is only 300. Public libraries' responses regarding their usage of NLC-derived records, therefore, may be overestimates.

The number of hits per document is measured by examining the data for 70 monographic and 30 government document records when searching the 30 OPACs. For monographs, only 28 documents out of 70 (40%) could be located in these OPACs, with an average hits per document of 3.5, with a 95% confidence interval of 2 to 5 hits per document. Of these records, 18 were derived directly from NLC. Although these records were not distributed evenly among all the OPACs, the average number of derived NLC records per document is 2 with a range of 1 to 3 in 95% of cases. Proportionally, the average number of derived NLC records per document out of the total retrieved is fairly high at 61%, with a range of 45% to 78% in 19 out of 20 instances.

In the case of government documents, 20 out of the 30 records were located in the OPACs. The average hits per document is 8 with a range of 4 to 12 in 95% of cases. Of the retrieved records, 13 were derived from NLC. While these records were not distributed evenly among all the libraries, the average number of derived NLC records per document is 6 with a range of 2 to 9 in 95% of cases. Proportionally, the average number of derived NLC records per document out of the total retrieved is relatively high at 49% with a range of 36% to 61% in 19 out of 20 cases.

The record-matching procedure was designed to avoid any involvement from the library whose OPAC was being examined. In this way it cost the libraries in the sample nothing, which is an attractive feature of this study. It is also a weakness, because the determination of the source of the record had to be made by someone who (in all but one library) was not familiar with the integrated library system

used by that library or with the institutional history relating to cataloging policies. In the surveys and interviews, libraries pointed out characteristics of their databases that would complicate the determination of source; an insider would know how to compensate reliably, while an outsider may only be able to make an educated guess. One example is the 040 fields. At least two of the OPACs use software that seems to move the original 040 into a (locally defined) 046, and insert a new 040. The researcher can spot this visually, but will not know if the software does this for all sources of records or only some sources.

Another effect of lack of local involvement is that we were unable to exclude from the population of candidate libraries those whose collection or cataloging policies result in a lack of interest in copy cataloging for significant segments of the record sample. Actual library collections are not random, but are selected in response to the mission of the library, and the subject areas and formats of material collected vary a great deal. None of the libraries whose collections were examined have as their mission to exhaustively collect Canadiana.

Questionnaire and interview responses showed that some libraries are controlling either paperback fiction or government documents only with brief records. This form of cataloging control can be very inexpensive, but it does not provide a comparable level of access to the material as would full or core cataloging. Responses that grouped such minimal level records with other original records make original cataloging appear much cheaper in comparison with copy cataloging than really is the case. Several libraries held some of the target government documents, none of whose records were NLC copy, but these turned out to be brief control records and not full original cataloging nor derived cataloging from other sources.

Conclusions

The Canadian university and large urban public libraries reported in the questionnaire survey that about 10% of their cataloging is derived from NLC MARC records. The university libraries' responses match closely the results of the record-matching methodology used in the project. On average approximately 1,200 records are derived from NLC per year by this category of libraries. Matching a sample of Canadiana records to the public libraries' collections, however, suggests that the data reported by the libraries in the questionnaires may be overestimated. While the result from the record matching indicates that public libraries on average only use about 300 NLC records per year, their responses to the questionnaire show an average of more than 2,000 records. In both cases, the average number of hits per record derived from NLC is high, indicating that certain Canadiana publications are popular among all libraries.

The average annual cost saving for a university library when using NLC MARC records for derived cataloging for Canadiana monographs and federal government documents is $16,400, while the average saving for a large urban public library is $7,800. In general, the reported data show that large urban public libraries spend significantly less on cataloging than academic libraries while acquiring a smaller proportion of Canadiana, particularly government documents, for their collections. If we use the data provided by libraries to extrapolate the range of cost savings for all academic and large urban public libraries, we may conclude that NLC is saving the libraries approximately $1,725,600 (with a range of $821,400 to $2,642,000) per year.

Libraries rely heavily on two additional services provided by NLC. The major single source of the derived MARC records is Amicus Online, and 82% of libraries in the study reported using NLC name and series authorities. The savings outlined above do not take account of the financial benefits accruing to Canadian libraries from these services.

Works Cited

ALCTS Technical Services Costs Committee. 1991. Guide to cost analysis of acquisitions and cataloging in libraries. *ALCTS Newsletter* 2, no. 5: 49–52.

Bedford, Denise A.D. 1989. Technical services costs in large academic research libraries: A preliminary report on the findings of the Samuel Lazerow Fellowship Project. *Technical Services Quarterly* 6, nos. 3–4: 29–48.

Carter, Kathy. 1997. Outsourced cataloging and physical processing at the University of Alberta Library. In *Outsourcing library technical services operations: Practices in academic, public, and special libraries,* Karen A. Wilson and Marylou Colver eds. Chicago: American Library Association, 3–14.

Deriez, Rene, and Thierry Giappiconi. 1994. Analyser et comparer les coûts de catalogage. *Bulletin des Bibliothèques de France* 39, no. 6: 28–35.

El-Sherbini, Magda. 1995. Contract cataloging: A pilot project for outsourcing Slavic books. *Cataloging and Classification Quarterly* 20, no. 3: 57–73.

Jenda, Claudine Arnold. 1992. Time and workflow study of the cataloging process used to evaluate Library of Congress cardsets as a cataloging support service. *Library Resources and Technical Services* 36, no. 4: 426–40.

Kantor, Paul B. 1986. Three studies of the economics of academic libraries. *Advances in Library Administration and Organization* 5: 221–86.

Lancaster, F.W. 1973. *The measurement and evaluation of library services*. Washington, D.C.: Information Resources.

Leung, Shirley W. 1987. Study of the cataloging costs at the University of California, Riverside. *Technical Services Quarterly* 5, no. 1: 57–65.

Library of Congress. 1996. New cataloging cost methodology. LC Cataloging Newsline 4, no. 13. Accessed Sept. 23, 2002, www.loc.gov/catdir/lccn0413.html.

McCain, Cheryl, and Jay Shorten. 2002. Cataloging efficiency and effectiveness. *Library Resources and Technical Services* 46, no. 1: 23–31.

Morris, Dilys E. 1992. Staff time and costs for cataloging. *Library Resources and Technical Services* 36, no. 1: 79–95.

Morris, D. E., et al. 2000. Cataloging staff costs revisited. *Library Resources and Technical Services* 44: 70–83.

Oldfield, William R. 1987. Cataloguing and catalogue maintenance: Functional cost allocation system. *Technical Services Quarterly* 5, no. 2: 55–65.

Orr, R.H. 1973. Measuring the goodness of library services: A general framework for considering quantitative measures. *Journal of Documentation* 29, no. 3: 315–32.

Osmus, Lori L., and Dilys E. Morris. 1992. Serials cataloging time and costs: Results of an ongoing study at Iowa State University. *Serials Librarian* 22, nos. 1–2: 235–48.

Partington, Lynne, and George Talbot. 1997. Outsourcing cataloguing and physical processing: A Canadian experience at the University of Manitoba Libraries. In *Outsourcing library technical services operations: Practices in academic, public, and special Libraries*, Karen A. Wilson and Marylou Colver eds. Chicago: American Library Association, 88–99.

Rider, Mary M., and Marsha Hamilton. 1996. PromptCat issues for acquisitions: Quality review, cost analysis and workflow implications. *Library Acquisitions: Practice and Theory* 20, no. 1: 9–21.

Technical Services Cost Studies in ARL Libraries (SPEC Kit 125). 1986. Washington, D.C.: Association of Research Libraries.

Wade, Rona, and Vicki Williamson. 1998. Cataloguing costed and restructured at Curtin University of Technology. *Australian Academic and Research Libraries* 29, no. 4: 177–89.

Wiggins, Beacher J. 2000. Library of Congress benefits from BIBCO. *RLG Focus* 47. Accessed Sept. 9, 2002, www.rlg.org/r-focus/:47library.html.

Appendix A: Questionnaire (English-language version)

The Use of NLC Source MARC Records in Canadian Libraries

The objective of this questionnaire is to collect information about the extent to which Canadian university and large urban public libraries make use of MARC catalog records generated by the National Library of Canada.

The questionnaire is being sent to the libraries of all institutions listed in the *Directory of Canadian Universities,* and to all members of the Council of Administrators of Large Urban Public Libraries (CALUPL). The analyzed and aggregated data will be incorporated in a final report to be submitted to the National Library of Canada, and may also form the basis of published articles or conference papers. The responses will be treated with full confidentiality. Individual libraries will not be identified in any reports, conference papers, or publications.

This research is being undertaken under contract by the Graduate School of Library and Information Studies, McGill University, on behalf of the National Library of Canada.

The questions below relate to published print monographic titles only: fiction and nonfiction; adult and children's; commercial and government publications. Please exclude all electronic and other nonprint materials.

As each institution is receiving one questionnaire only, please answer these questions for your entire library system. If you are unable to do this, please specify the branch library or other part of your library system for which you are responding:

_____

1. Name of your library:
   _____
2. Total number of published print monographic titles in your collection: _____
3. Total number of published print monographic titles in your collection, cataloged in the last fiscal year for which you have data (excluding RECON):
   (a) Number _____
   (b) Year _____
4. Percentage (%) of published print monographic titles in your collection, copy cataloged (whether using either partial or full copy, from all sources) in the last fiscal year for which you have data (excluding RECON). (Please give estimate if complete data unavailable.) _____
5. (a) Percentage (%) of published print monographic titles in your collection, copy cataloged (whether using either partial or full copy) using NLC bibliographic records in the last fiscal year for which you have data (excluding RECON).

(Please give estimate if complete data unavailable.)_____

(b) Of these, what percentage are Canadian federal government documents? (Please give estimate if complete data unavailable.)_____

6. Percentage (%) of Canadiana published print monographic titles for which you undertook full, original cataloging (excluding RECON) in the last fiscal year for which you have data. (Please give estimate if complete data unavailable.) _____

   For the purposes of this question, the term "Canadiana" refers to publications that meet any one of the following four criteria: publications from a Canadian publisher, by a Canadian author, on a Canadian topic, or any Canadian (federal and provincial) government publications.
7. Was the fiscal year used in your previous answers a typical one for your cataloging activities? If not, please elaborate.
8. Average cost per bibliographic record for copy cataloging (whether using either partial or full copy) of published print monographs. Please incorporate all direct costs: personnel, subscriptions, etc.
   (Please give estimate if complete data unavailable.)
   actual _____ OR estimate _____
9. Average cost per bibliographic record for full original cataloging of published print monographs. Please incorporate all direct costs: personnel, subscriptions, etc. (Please give estimate if complete data unavailable.)
   actual _____ OR estimate _____
10. Please list any categories of published print material
    (e.g., fiction, government documents) for which you almost never use copy cataloging.
11. Please list any categories of published print materials for which you almost always use copy cataloging.
12. (a) Which sources does your library actually use to find bibliographic records for cataloging purposes? Only answer for published Canadiana print monographic titles. For the purposes of this question, the term "Canadiana" refers to publications that meet any one of the following four criteria: publications from a Canadian publisher, by a Canadian author, on a Canadian topic, or any Canadian (federal and provincial) government publications.
    Indicate as many as necessary
    ____ Amicus Online
    ____ AG-Canada
    ____ OCLC Online
    ____ RLIN
    ____ Other online sources (please specify)
    ____ Canadiana CD-ROM
    ____ CatME
    ____ LaserQuest
    ____ BiblioFile
    ____ Other CD-ROM sources (please specify)
    ____ Web OPACs (including Z39.50)
    ____ Book vendors
    ____ CIP as printed in books
    ____ Other (please specify)
    ____ Not applicable

    (b) Which of these is your major source for bibliographic records for Canadiana?
13. On average (excluding high priority, rushed, or urgent items), how many months will you wait for a bibliographic record of a published print monograph title to become available before cataloging the title originally? _____
14. Do you make use of NLC name or series authorities?
    Never_____ Occasionally_____
    Frequently_____
15. How might the NLC's bibliographic records be of more use to you for cataloging purposes? Please answer in as much detail as you wish.

Please add any other information relating to your cataloging practices for Canadiana material or about NLC cataloging practices that you think relevant.

We should very much like to hold a short follow-up telephone interview with you in March. If you are willing to participate, please give your name and telephone number.

Thank you for your help.