Book Vendor Records in the OCLC Database

Boon or Bane?

Laura D. Shedenhelm and Bartley A. Burk

This case study is based on a 1998 sample of recently acquired Spanish-language firm-ordered materials, all of which had vendor records in the OCLC database. Vendor records were compared to final fully cataloged records to study differences in the basic bibliographic description fields (1xx, 245, 300, 4xx, 5xx). Identified were the types of errors found in the records and the duplication rate with records already in the database (full LC and member records, partial member records, and other vendors). Both areas are problematic for cataloging units. Secondary research objectives included tracking titles for usable copy cataloging and analyzing the cost impact for typical cataloging workflow. The researchers conclude that the records, though sometimes problematic, are useful. Suggestions are given for areas of improvement.

In recent years, there has been an exceptionally close scrutiny over public funds. Libraries are included in this realm and are among the most cost-conscious entities in the public sector. This undoubtedly reflects our innately conservative approach to fulfill our responsibility to acquire, maintain, and preserve the intellectual and cultural patrimony of our parent institutions and to serve our clients. While we have watched the steady increase of prices for the materials we purchase, we have sought innovative ways to hold down costs for acquiring and providing access to our collections. Book vendors are also seeking ways to lower their overhead costs, provide needed services to their clients, acquire new clients, and maintain a profit margin which will allow them to stay in business. There is a symbiotic relationship between libraries and book vendors that is necessary for both groups to succeed in a fast-paced and competitive world. This study looks at part of the symbiosis that appears to be deleterious to one part of that relationship: the impact on cataloging processing units within libraries of vendor-produced records in the national bibliographic databases.

Vendor records are very brief bibliographic records originally designed to advertise an item for sale by the vendor. They are based on the files the vendors use to create their own sales catalogs and include minimal information: the author and title of the work, publication information, and extent of the item. Often, they also will have notes about the work, "edition" information, and some subject analysis. When this information is accurate, it is very useful for bibliographers who must decide if they want to purchase the item for their collections. However, a problem arises when the data presented do not accurately represent the bibliographic item they are meant to describe. The descriptions in vendor catalogs are often so brief it takes real skill to decide if the item would be a unique addition to a collection. For example, "edition" entries in these catalogs often merely indicate new printings, not new versions of the works. The form of

Laura D. Shedenhelm is Special Projects Cataloger, University of Georgia, Main Library, Athens, and **Bartley A. Burk** is Social Sciences Cataloger, University of Notre Dame, Indiana.

the author's name, and sometimes even the choice of author entry, rarely matches a library's entry in its catalog. Once the item is purchased for a collection, catalogers must describe it in terms of library cataloging rules more detailed than those adequate to advertise an item. Catalogers must distinguish the basic bibliographic unit from all other renditions of the intellectual content presented, and they must make it fit intelligently into the larger bibliographic universe. What happens, then, when these minimal records are added to the national databases that traditionally provide accurate and full cataloging that generally meet industrywide agreed-upon standards? Is there an impact on the work performed in libraries' cataloging units?

Our premise is that there is an impact. We took a small sample of vendor records as a basis for a study on their impact on a typical cataloging workflow. We looked at:

1. How much manipulation of the data in vendor-produced records must be done in order to bring them into alignment with library standards for cataloging. We did this by comparing the original versions of the vendor records to the title pages of the works they represented. We asked: Are there errors in vendor records that make them difficult to utilize as nationally acceptable standardized library cataloging records? If the answer is yes, is there a pattern to the errors?
2. How well vendor records fit into the greater bibliographic universe. We searched the OCLC database for any other records that would also represent the items in the sample. We asked: Do vendor records duplicate any records in the national bibliographic databases? If so, at what rate?

These two areas, editing of records and determining appropriate records to use for cataloging, constitute the majority of work performed in cataloging units. Anything that negatively impacts these two areas would have an unfavorable effect on cataloging production. Timely and cost-effective cataloging is a primary goal for any cataloging unit. Because of this we realized that there were two other questions (which we will call 3 and 4) that our study needed to address:

3. How soon does full-level copy cataloging appear in the national bibliographic databases for items represented by vendor records? We tracked the sample items through an eleven-month cycle to see when any source of copy cataloging would appear. We performed a follow-up search fourteen months later.
4. Do vendor records provide a cost-efficient method for giving access to our collections? We looked at costs for a typical cataloging routine to see if there

were any savings as a result of vendor records being in the national databases.

Spanish-language catalogers were among the first to feel the impact of vendor records. During the last few years, though, other non-English-language catalogers have begun to see similar records appear for their materials. Most European vendors now routinely send tapes to be loaded into the bibliographic utilities. While this study looks specifically at records from Spanish-language vendors, the basic issues are applicable to all of these records. This was verified through colleagues working with German-language material at University of Notre Dame and French-language materials at University of Georgia. We note here that these issues apply most directly to OCLC, which uses the single "master record" method for displaying bibliographic cataloging and holdings. Our colleagues using the RLIN utility with its "clustering" technique have fewer dilemmas with these records.

History of Spanish-language Vendor Records in Bibliographic Utilities

The history of vendor records in the bibliographic utilities goes back to the early 1990s (Peet 1999). During 1990 and 1991, representatives of Casalini Libri approached the Library of Congress with the idea of using electronically formatted vendor records to streamline the acquisition process. Unfortunately, the idea was "before its time," since the online system at LC was not configured to be able to take advantage of this technological enhancement. LC decided not to pursue this idea.

Various factors converged during the next few years forcing the Library of Congress to take a more aggressive approach to implementing computer-based processing. There was a federal hiring freeze and a vacancy rate of library technicians at LC that reached nearly 40.00%. In May 1993, at the Guadalajara meeting of SALALM (Seminar on the Acquisition of Latin American Library Materials), members of LC's Hispanic Acquisitions approached Puvill Libros with this same idea of using electronically formatted vendor records in order to address the growing backlog of materials at the library that needed to move through the acquisition process. By 1993, LC had an overall backlog of 8,500 recent Spanish imprints (two- to three-year-old publications) that needed minimal-level records in their database. After several months of testing, LC received the first set of MARC-formatted records from Puvill in early July 1994. Always open to pursuing good, apparently cost-saving ideas, Puvill approached OCLC in 1995, suggesting the sharing of these records with a wider audience. Many other vendors have followed this same path, leading us to the current situation.

Literature Review

A quick search of library literature will show that minimal-level records in the utilities is not a new idea. Our initial literature search indicated that there had been nothing written about vendor-produced bibliographic records. However, this has recently changed.

Vendor-produced Bibliographic Records

Beall "subjectively examines the impact the addition of vendor record to the bibliographic utilities, chiefly OCLC, has had on cataloging" (Beall 2000, 230). He analyzed vendor records for content and quality, then looked at the impact on cataloging and patron access. His study is based on opinion and calls for three areas of quantitative analysis. Our study validates much of his commentary, and addresses in part his suggestion for research about the quantity of vendor records that were enhanced after one to two years in the utility.

Minimal-level Records and Duplication

Two of the best histories about minimal-level records are by Patton (1991) and Stamm (1996). Everyone is aware of CIP records and those of us who have been around for a while will remember the early days of mass tape-loading projects that added many minimal acquisition-level records to OCLC. In our drive to get the most out of our processing dollars, we seem to be wedded to the idea that "something is better than nothing."

Yet, as professionals, we have questioned that marriage often through the years. Fox and Preece (1991) and Preece and Fox (1992) cover issues most closely related to vendor records. They discuss three reasons to upgrade minimal-level records: (1) to provide increased patron service through complete cataloging; (2) to maintain the library's commitment to quality cataloging; and (3) to enhance record quality for the online catalog (Fox and Preece 1991). As with vendor records, the minimal-level records in their study had no subject headings and there was no authority control for access points. They note that accuracy in transcription and coding of the title and extent fields is an area of highest concern: "[to] support key word title access, accurate transcription of the entire 245 field is critical" (Fox and Preece 1991, 29). They conclude that very experienced catalogers must do a detailed review of the entire record. They go on to recommend ensuring that the minimal-level record is not a duplicate in the OCLC database. "Duplicate" means one or more record(s) representing one bibliographic item that do not exactly match, resulting in a duplication of the OCLC master record, which will not automatically combine with the master record when OCLC's merge program is run. In their subsequent article (Preece and Fox 1992), Preece

and Fox review the theory that minimal-level records in the database help librarians predict when full-level records will appear. They use a similar technique to our study, checking for record upgrades over a six-month period. They found that minimal-level records duplicated full-cataloging member records at a rate of 20.00%. They also indicate that catalogers discovered upgrading minimal-level records took as much time to catalog as items without copy, and required additional work to verify if there were duplicate records in the OCLC database. When duplicates were found, OCLC was notified, adding more time to the process of what should be inexpensive, simple copy cataloging. They cite a 1990 survey of Association of Research Libraries as supporting their observations. Similar discussions on upgrading minimal-level records can be found in case studies by Ferguson (1991) and Handman (1991).

Horny (1991) and Intner (1994) both look at the promise of minimal-level cataloging versus the end product. Horny points out that if the materials are not processed through the system, they are not available to the patron. Minimal-level cataloging promises to move items through more quickly. However, she notes that the information most easily left out of records is the data that is most easily and quickly entered into them. The most time-consuming aspect of cataloging is authority control where "savings are difficult to achieve" (Horny 1991, 10). Intner concludes that the items that would most benefit from minimal-level cataloging are those most easily identified and available rather than the most esoteric (Intner 1994).

Duplication and Errors

Beyond Fox and Preece, much has been written about the difficulties arising in all areas of library work due to duplication of records in the database. Johnson and Josel (1981) discussed the types of errors that cause duplicates and the resulting costs of duplication. Wanninger (1982) looked at the impact of duplication on search results and difficulties related to multiple searching strategies. O'Neill, Rogers, and Oskins (1993) looked at typical errors found in records that result in duplication in the OCLC database. The characteristics include "(1) typographical errors, (2) erroneous tags and subfield codes, (3) omitted information, and (4) inconsistencies between the variable and fixed fields" (O'Neill, Rogers, and Oskins 1993, 61). Our findings closely parallel theirs. They assert that it is transcription in a form that is similar but not exactly the same that leads to most duplicate records, and point to batch loading as a major source of duplicate records. The most significant characteristics identified for duplication occurred in the date, author, or publisher areas of the record. In particular, one-third of the duplicated records contained author entries that did not match. It should not be assumed, however, that only records

that are duplicated show these characteristics. A study conducted by Romero (1994) that looked at typical errors made by both entry-level and experienced catalogers indicated that there were areas of difficulty in cataloging in general. She found that most errors occurred in choice of heading and description, but that typographical errors were minimal. In the area of description, the publication field was the most problematic. Our findings corroborate this. Also, the results of our study will demonstrate that all of the problems seen with earlier versions of batch-loading minimal-level records are intrinsic in loading of vendor tapes into the utilities.

History of the Project

From the first appearance of vendor records in the bibliographic utilities, there has been intense, often heated, discussion about these records during the SALALM annual meetings. While it was noted that acquisitions departments, including the Library of Congress' Hispanic Acquisitions Section, find these records useful, there was a diversity of opinion among catalogers at SALALM about their overall utility. The SALALM Cataloging and Bibliographic Technology Subcommittee has members from a variety of public and private, large and small institutions that subscribe to either RLIN or OCLC, or both. As such, the subcommittee is representative of the greater library community. Variations in processing workflow and attitudes about patron access at different libraries influence whether these records are more or less troublesome. This is especially true when coupled with a library's decision to catalog the items locally or nationally. SALALM members that catalog locally and only attach their holdings to a national utility through tapeloads have fewer complaints about vendor records than do members who catalog directly onto the utilities.

The seemingly endless discussion about these records prompted some of us at SALALM to do a study of the vendor records. Our hypothesis was that, while possibly useful at the acquisition stage, these records posed several problems for catalogers. Chief among these problems is that converting these records to nationally acceptable standardized library cataloging records would be time-consuming. We felt that it was easier and more cost-effective to create an original cataloging record. We believed that comparing the unenhanced vendor record to the actual item would illustrate this. We knew that the vendor records would always lack specific information we require in our fully cataloged records: classification (in this case Library of Congress Classification, or LCC), bibliographic and content notes in English, and subject access through Library of Congress Subject Headings (LCSH). All of this must be added to the record before it can pass out of processing as fully usable cataloging, which means that the library can never accept the record just as it originally appears from the vendor. Additionally, name headings would have to be verified, as it was unlikely that the vendors would create the nationally compliant access points. We also wanted to see how these records fit into the overall picture of bibliographic copy available to us in our daily work. Did these records represent any duplication of records already in the bibliographic utilities? If the answer is yes, then how much duplication is created? Finally, we realized that the deciding factor about the utility of these records would be based on whether they were cost-efficient or not. At what point could these records be handled by the least costly staff? A précis of the methodology and preliminary data used here was presented at the meeting of the Cataloging and Bibliographic Subcommittee during the 1999 annual meeting of SALALM held in Nashville, Tennessee.

Method

The purpose of this small study is to see if the complaints about the quality of vendor records (which were raised during the SALALM meetings), the rate of duplication in the OCLC database, and the resultant increased cataloging workload and cost are justified. To do this, we compared the vendor record to the item, seeking any errors in transcription or choice of data transcribed; we searched the OCLC database for any duplicate records; we tracked the records for potential use in copy cataloging; and we analyzed the costs.

Our sample was gathered from the workflow at the University of Notre Dame Hesburgh Library, which uses vendor records in its acquisition process, but is not an OCLC enhance library. All of the titles would have been searched in OCLC at the point of order. If a record existed at that time, it would have been downloaded into the local system to generate the order. Otherwise, a record would have been created on the local system. The items were always searched by the International Standard Book Number (ISBN). This is a simple and inexpensive search that should give good results provided the ISBN appears in the record. The assumption was that this search takes approximately five minutes per item.

When the items were received, the local database was searched. Items with full- or nearly full-cataloging records were appropriately distributed to the copy-cataloging unit. Items with strictly acquisition-level records were researched by ISBN in the OCLC database, then distributed accordingly. Normally, titles lacking full cataloging are put into a searching cycle for a maximum of 18 months. About every three months the title is searched for an acceptable record, and sent through processing or put back into the search cycle until it completes 18 months in the cycle. If there is no

full cataloging at the end of that period, the item is sent to the professional staff for local processing.

We collected a set of 64 titles from all items routed to the social sciences cataloger at Notre Dame from two shipments from Puvill Libros received in May and June of 1998. The sample was of titles that had only vendor records available when received. The original versions of the vendor records associated with the sample titles were printed. We kept these as examples of how they appeared in OCLC, prior to any upgrade or merging with other records.

In early July 1998, the title pages and versos of each sample title were photocopied and matched with the original vendor record. Using this process, we sometimes missed series information that would have appeared on covers, series title pages, or spines. Therefore we will not include any discussion of the series fields in our analysis. The title page and verso information was then compared to the original versions of the OCLC records, noting any discrepancies, such as typographical errors, omissions of data required by the cataloging code, miscoding of fields, etc. The 300 field (pagination, etc.), in general, was omitted from this review as the work was being done without the book. The exception was for multivolume sets. The issues related to these are discussed under "duplication" below.

During late July and early August of 1998, the OCLC database was searched for all records that might match the title pages. A combination of techniques was used in this process. We searched by ISBN, all types of derived searches, and scan title in order to ensure that we found any record that could be used in the cataloging process. This process was repeated in October of 1998 and April of 1999. Due to the appearance of the Beall (2000) article, another search was done in July of 2000 for any title that had not received full cataloging by April of 1999. Each time the records were searched, detailed notes were made about the records found. These notes include the number and type of duplicate records found (e.g., other records from LC, a member library, or a vendor), the number of holding libraries for each record, differences in the records versus the title pages, and differences in the records that might cause them not to merge using a duplicate-detection algorithm.

Finally, in order to try to analyze costs, we kept track of when each item would have been able to move through the copy-cataloging process that requires no additions or changes, the least expensive processing. Then, we looked at the overall costs associated with the processing of these titles.

Analyzing Collected Data

Our data involves four areas: (1) errors in the vendor records that lead to increased editing time, and that may be the cause of duplication in the bibliographic databases; (2) the amount of this duplication; (3) timeliness of copy cataloging for items represented by vendor records; and (4) processing costs for cataloging items using vendor-generated records.

Errors

As noted in the literature review, O'Neill, Rogers, and Oskins (1993) and Romero (1994) found characteristic types of errors records in the database. These errors are consistent with our findings for vendor records (see tables 1 and 2). While we saw some problems with accurate transcription, we believe the typing errors were typical of any randomly chosen set of records from the database. There was also the occasional odd decision for 245 filing indicators. These two problem areas would result in difficulty locating the record in the database and could contribute to duplication, but overall were minor.

Romero found that both experienced and beginning catalogers made the most errors (41.60% and 49.06%, respectively) in (1) choice and form of headings and (2) description of the items (Romero 1994). We also found that the most difficulties were in these two areas (see tables 1 and 2). The 260 field was particularly problematic due to inaccurate presentation of the publisher (33 of 49 errors, or 67.35% of description errors). From the information in table 1, one can conclude that the error rate in the description in vendor records is significantly higher (50.00%) than in records produced by people with minimal professional training. Additionally, one-fifth of the records (20.31%) had errors in choice or form of headings, either main or added entries.

Accurately editing fields that are inconsistently entered is time-consuming and difficult. Having to spend extra time meticulously correcting a record automatically raises the cost of using it. Romero concluded that these were errors that could only be addressed through improved training. We

Table 1. Types of Errors in Vendor Records

Typographical		Choice/ Form of Headings		Bibliographic Description	
# of records	% of records	# of records	% of records	# of records	% of records
7	10.94	13	20.31	49	76.56

Table 2. Multiple Error Types per Vendor Record

Typing and Heading		Typing and Description		Heading and Description		Typing, Heading, and Description	
# of records	% of records	# of records	% of records	# of records	% of records	# of records	% of records
1	1.56	4	6.25	7	10.94	1	1.56

absolutely agree with this assessment: these problems could be overcome through training programs developed specifically for vendors.

With regard to form of headings, Russell Marr, Library of Congress, Hispanic Acquisitions Section, helped us research the issue of vendors' access to the LC name authority file. According to LC personnel, this file is not used by the vendors that LC contracts for original acquisition tapes (Marr 1998). They would need to contract with OCLC or RLIN for access to it. The use of authorized versions of personal and corporate names is the heart of appropriate access and collocation in our catalogs. Anyone who has ever attempted to look for information in periodical and newspaper indexes about Muammar Qaddafi can attest to the problem inherent with multiple ways of entering a name! While most of the personal names in the sample could be relatively easily identified and corrected, corporate names, especially conference names, were rarely correct. These were usually tagged as 100 fields instead of the appropriate 110 or 111 designation. Again, the correct tagging of the fields could be easily improved with minimal training, but significantly more in-depth instruction would be necessary for confident construction of corporate names to fit into the bibliographic setting that our libraries use. The lack of authorized headings in the vendor records has the potential to be very costly over time. When one considers that the correction to name fields must be done by every library that uses the record, until such time as a library locks and corrects the record, one begins to see the extent of the associated costs. Unfortunately, without daily searching of each record, we would have no way of knowing at what exact point any particular record was upgraded, or how many libraries had used it prior to the enhancement. Therefore, we cannot state unequivocally how much libraries throughout the world are expending toward these corrections.

Duplication

Duplication of records in the OCLC database is problematic for efficient cataloging. We found that vendor records exacerbated this problem. Our initial search, using the variety of techniques described above, yielded duplicate records for 17 of the titles, 12 of which had usable cataloging copy. This means that 26.56% of our titles were duplicated in the database. Also, 18.75% of our sample had usable copy in the database without the distraction of the vendor records. (See table 3.) The second and third search cycles increased the duplication rate to 31.25% and 37.5%, respectively.

During the second search, one of the original records was upgraded by the Library of Congress (LC), one was upgraded by a member library, but also duplicated by a LC record, and two were duplicated (one by LC and one by a member record). This raised the overall rate of available copy cata-

loging to 25%. By the third search, the duplication rate had risen to 37% of the total. Of the 24 titles duplicated, 8 duplicates had been merged. Also, 23 had either been upgraded or the new duplicate represented full cataloging. The other title had been upgraded but lacked LC classification.

We believe that the two most common categories of errors noted above, description errors and incorrect form or choice of headings, are the roots of the duplication found in our sample. This is due to the very broad de-duplication algorithm used by OCLC that results in a lack of match when the program is run. Jay Weitz, Consulting Database Specialist for OCLC, sent us information about the Duplicate Detection and Resolution (DDR) software that was developed about a decade ago. He stated that the "DDR runs through the WorldCat database on an irregular basis (currently, roughly every six months)" (Weitz 2000a). He indicated that the algorithm uses fourteen descriptive elements on which to form a match for merging records. Also, there are about ten conditions that prevent merges. He noted that more than one million duplicate records for the books format have been merged since DDR first ran in June 1991. In his subsequent message, he listed the elements included in the algorithm: "cataloging library; LCCN [Library of Congress Control Number]; ISBN; government document classification number; media; author; title; statement of responsibility; edition statement; place of publication and publisher; publication date; number of pages or volumes; size; and series statement" (Weitz 2000b). He went on to say "we tend to err on the side of adding or leaving duplicates rather than merging away unique records whenever there is uncertainty" (Weitz 2000b). Given the need to match on publication data and choice and form of author, the areas with the most errors in the vendor records, it is not surprising that only eight of the duplicate records found during our study had been merged by the time of our third search of the database. It is also not surprising that so many duplicates are created at the initial tape loads. It would be interesting to know, of the eight merged records, how many were merged automatically and how many were reported by diligent libraries trying to clean the database.

Whenever there is duplication in the database, the cost of locating appropriate copy cataloging automatically increases. Instead of being able to do a simple numeric search of an ISBN, which should result in one record in OCLC, the

Table 3. Duplicates

Date	# of Duplicate Records	% of Total	# of Records with Full Copy	% of Total
Aug. 1998	17	26.6	12	18.8
Oct. 1998	20	31.3	16	25.0
Apr. 1999	24	37.5	23	35.9

library must pay more experienced, and therefore more expensive, personnel to distinguish among several records, and decide which is the best choice. Beall (2000) also shows that libraries, which had previously been able to utilize OCLC's Cataloging Micro Enhancer (CatME) software to great advantage in finding suitable copy cataloging, are now stymied by the influx of minimal-level records that the CatME software cannot distinguish from full-level cataloging. The automated search results in multiple records being downloaded, and the consequent increase in cost. The issue of duplication for vendor-produced records is particularly significant for Spanish-language cataloging as there are multiple vendors giving records to OCLC. Our study showed several cases when there were multiple vendor records for a single title.

The issue of multiple vendor records for a single title was particularly evident with multi-volume sets. In our sample, we found separate records for individual volumes of multivolume sets. Standard cataloging practice dictates that these volumes would be described on one bibliographic record in most cases. The exception, of course, would be a monographic set that warranted analysis of each individual volume (a rare case). We had two cases of multivolume/multirecords for items that were fully cataloged incomplete monographic sets. OCLC's reliance on machine-driven merge algorithms to identify and merge duplicate records will not work in these cases. We must rely on people to report these for manual merging, thereby increasing the cost of having them in the database (Johnson and Josel 1981).

Tracking Copy

When we looked at the time frame for appearance of copy cataloging for the titles in our sample, we were pleasantly surprised. There was a steady increase in the amount of copy available through either upgrades of the vendor record or duplication of these records by full-level LC or member-generated records (see table 4). Slightly more than 20.00% of the titles had usable copy at the time of the first search period during the summer of 1998. Twelve of the 13, however, came from duplicate records. In only one case was the record an upgrade of the vendor record, and it was done by Notre Dame. By October of 1998, almost half (46.87%) had usable copy. When the last search was done in April of 1999, 11 months after the sample records were gathered, 81% had usable copy. For our original study period, only 12 remained without complete cataloging. Of those, 2 only needed a classification number to complete the cataloging.

Beall posed three research questions, one of which focused on percentage of vendor records enhanced at the end of a one-year period, and a two-year period (Beall 2000). Since we had been looking at availability of full cataloging for titles represented by vendor records, we did searches for the

12 titles that did not have full-level cataloging at the end of our initial project. We hoped our findings would shed light on this issue. Our findings do not directly answer Beall's question since we looked for *any* available copy, not just enhancements to the vendor records. However, our findings show that full-level cataloging is available for items represented by vendor-produced records within his timeframe. In table 4, we show that 61 of the 64 sample titles had full cataloging at the end of two years. This accounted for 95.31% of our sample. We agree that further research needs to be done in this area. In particular, more study needs to be done based on the criticism that less original cataloging is being done because of the shift of the workload to upgrade these nonstandard minimal-level records (Beall 2000).

Another issue raised by Beall concerns uploading improved versions of the vendor records for national consumption (Beall 2000). An earlier study by Sercan (1994) found a marked decrease in the amount of full cataloging for Spanish-language materials available between 1983 and 1992 (18.00% versus 4.00% for LC copy, and 13.00% versus 5.00% for RLIN member copy). Erbolato-Ramsey and Grover's (1994) findings "would seem to indicate that most libraries accept and input less than full AACR2r [Anglo-American Cataloging Rules, 2d ed., rev.] level records online and do not update them [nationally] at a later time" (83). Beall points to the lack of financial incentive as a likely reason for the decrease in upgraded copy (Beall 2000). Our findings do not seem to support this theory. However, our sample is extremely small given the overall output of Spanish-language publications. A larger study would be useful.

Costs

The costs described here are based on the 1998 INCOLSA network prices and the labor costs at Notre Dame for that year. The following were the labor costs: support staff for transcription = $0.33 per item (figured at $15.00 per hour and 5 minutes per item); professional completion of records lacking only LCC = $3.75 per item; professional completion for other records = $21.86 per item. Network costs were: numeric or derived search = $0.34; update = $0.43; export (to local system) = $0.12. Network credits are: nonenhance

Table 4. Tracking Copy

Date	# with full copy	% with full copy	# lacking copy	% lacking copy
Aug. 1998	13	20.31	51	79.69
Oct. 1998	30	46.87	34	53.13
Apr. 1999	52	81.25	12	18.75
July 2000	61	95.31	3	4.69

upgrade = $2.10; enrichment (call numbers, subject headings, etc.) = $0.52; nonenhance original addition = $3.83. Beyond these specific costs, one must remember that there are additional hidden costs for sorting through multiple records as identified in the section about duplication above.

Table 5 shows the per-item processing costs during the original one-year period for titles in our sample using the cycle of searching scenario earlier. Overall, this information is fairly irrelevant since the items would have been processed regardless of the cost. However, these costs show the same disincentive to upgrade the record that Beall found (Beall 2000): an enrichment credit of $0.52 to add the classification number to a record does not compensate for the processing costs of $6.87. Similarly, a processing cost of $24.98 for an item needing essentially full cataloging will now only gain a credit of $2.10 to upgrade nationally where it formerly would have been credited at $3.83. This cost structure would seem to undermine the tenet of cooperation that traditionally has been the foundation of OCLC. We begin to see the dilemma presented by Sercan (1994) and Erbolato-Ramsey and Grover (1994). Is there enough financial incentive to upgrade these records, or will we begin to see less full-level cataloging at the national level as budgets for technical process continue to tighten?

Conclusions

Our original questions were:

1. How much manipulation of the records must be done for vendor records to fulfill the requirements of the nationally acceptable standardized library cataloging records? Is this reflected in errors in vendor records that make it difficult to utilize them as cataloging records?
2. What level of duplication in the bibliographic databases do vendor records represent?
3. When does copy cataloging for titles represented by vendor records appear in bibliographic databases?
4. What are the costs associated with vendor records in the database?

First, we determined that the vendor records would never fulfill the requirements for library cataloging records since they will always lack classification and subject headings needed for our catalogs. We also found that there was a pattern of data-entry errors that placed a burden on the library community to correct. Accurate editing is a time-consuming and labor-intensive process in the best of settings. The types of errors found in the vendor records were not usually simple typographical ones. Rather, they were content-oriented and inconsistent, thus tended to increase the time necessary to review each record. Among the content errors was choice and form of headings. We concur with Romero (1994) that appropriate training, in this case for vendors wanting their acquisition-level records represented in the OCLC database, would help alleviate many of these problems. Further we believe that access through OCLC to the name authority files would improve the quality of the headings on the vendor records.

Second, we found that these errors led to duplication in the database. Of our sample, 37.50% was duplicated by full-level records already in the database or added during our

Table 5. Processing Costs in Dollars per Item

Full Cataloging Available	Pre-order Search	Export	Labor	Subtotal	Search at Receipt	Update	Labor	Total or Sub-total
On receipt	0.335	0.12	0.33	0.785	0.335	0.43	0.33	1.88
1st cycle	0.335	0.12	0.33	0.785	0.335		0.33	1.45
2nd cycle	0.335	0.12	0.33	0.785	0.335		0.33	1.45
3rd cycle	0.335	0.12	0.33	0.785	0.335		0.33	1.45
3rd cycle plus class no	0.335	0.12	0.33	0.785	0.335		0.33	1.45
3rd cycle plus full review	0.335	0.12	0.33	0.785	0.335		0.33	1.45

Full Cataloging Available	1st Cycle Search	Update	Labor	Subtotal	2nd Cycle Search	Update	Labor	Total or Subtotal
1st cycle	0.335	0.43	0.33	2.55				
2nd cycle	0.335		0.33	2.12	0.335	0.43	0.33	3.21
3rd cycle	0.335		0.33	2.12	0.335		0.33	2.78
3rd cycle plus class no	0.335		0.33	2.12	0.335		0.33	2.78
3rd cycle plus full review	0.335		0.33	2.12	0.335		0.33	2.78

Full Cataloging Available	3rd Cycle Search	Update	Labor	Total Cost
3rd cycle	0.335	0.43	0.33	3.88
3rd cycle plus class no	0.335		3.75	6.88
3rd cycle plus full review	0.335		21.86	24.98

study. Therefore, more experienced people would have been needed during the searching process to accurately identify appropriate cataloging copy. We determined that OCLC's duplicate-detection algorithm errs on the side of adding duplicates in order to ensure unique records are not inappropriately merged (Weitz 2000a). This means that we must rely on people to report duplication as it is found. However, Johnson and Josel (1981) concluded that this is a course of action too expensive to pursue consistently.

The corollary to the duplication issue, however, is that two-thirds of the titles in our sample were new to the database. It is valid to ask, then, would there have been more or less full-level cataloging in the database if the vendor records did not exist? Does having a base record from which to work help or hinder the production of final cataloging? It can be argued that, since materials go through the acquisition process more quickly, they are in the cataloging workflow more quickly, and therefore are attended to sooner than before vendor records were available. Further study would have to be done in this area to verify any speed of access or costs savings this might represent.

Next we looked at when full cataloging appeared for titles represented by vendor records. We could not determine full-level cataloging availability based solely on the upgrade of vendor records because often full cataloging came from records that duplicated the vendor record. However, by tracking the materials through a cycle of searching the OCLC database for cataloging copy, we were able to determine how quickly they would have been made fully accessible to the library's patrons. We determined that 81.25% of the titles had full cataloging at the end of the original study. In response to Beall (2000), we searched the titles that lacked full cataloging at the end of the original study. This revealed that a total of 95.31% of our sample had full-level cataloging available at the end of two years: a very good return.

Finally, we looked at processing costs. Do vendor records provide a cost-efficient technique for providing quick access to our collections? Our answer would be a qualified yes. If our 95% rate of full cataloging over a two-year period is representative of vendor records in the database, and as long as a handful of libraries are willing to continue to bear the full burden to upgrade these records, the rest of us will enjoy inexpensive cataloging. However, should the burden become too great and the reward remain so little, we could well begin to see a drop in the percentage of full-level cataloging available. Further study would need to be done to determine if there is any other positive impact from the presence of vendor records. In particular, more analysis of the impact on the costs of processing due to the duplication rate associated with vendor records would be useful. Also, a study needs to be done to see if there is any correlation between the addition of new and unique titles entered as vendor records and how quickly these are upgraded.

Recommendations

Vendor records will be a permanent part of the bibliographic universe for processing materials in our libraries because they provide early notification of new materials. However, beyond timeliness, high quality records are also important for both the national databases and our local library catalogs. We believe that by adopting some or all of the following recommendations the bibliographic quality of vendor records will improve and the level of acceptance in the library community will increase.

Recommendation 1: OCLC should initiate a rigorous training campaign for all vendors adding records to the bibliographic utilities. This training should concentrate on the bibliographic description standards used by the majority of OCLC's constituency.

OCLC expects members to adhere to national-level standards for the records we place into the database. It is reasonable for members to expect this quality for any record we find in the database. Sally McCallum, Chief of LC's Network Development and MARC Standards Office would also seem to support this idea:

> [I]t is generally the responsibility of sending organizations to make records conform to community developed and approved standards, and if the original sender does not do this then many to whom the original sender distributed have to take on duplicative work (McCallum 1997).

The authors sent examples of typical errors found in this study to the Library of Congress for a training program it developed in the fall of 1998. These examples could be the basis for a training program to refine all vendor records to meet the community-developed standards noted above.

Recommendation 2: We encourage OCLC to make the authority files available to vendors. Further, we encourage extensive training in the construction and application of these files and the headings they contain.

One-fifth of the errors in our sample were from incorrect headings. Having the correct form of the name will lower the cost of upgrading the record. It will also help in the deduplication process given that "author" is one of the fourteen elements used in the OCLC deduplication program (Weitz 2000b).

Recommendation 3: We strongly encourage initiation of financial credit to libraries reporting duplicates in the OCLC database.

Every library must spend precious staffing funds wading through multiple records to find full cataloging. Libraries also spend funds notifying OCLC of these duplicates. Consequently, this process is rarely done according to Johnson and Josel (1981). It is clear that any savings vendor records

may provide to libraries in terms of the keying process can easily be consumed by the searching process when there is a high rate of duplication. We have noted that further study needs to be done to determine the costs to libraries of documenting and reporting duplicates. Since OCLC's deduplication program errs on the side of adding duplicates, OCLC should offer reasonable compensation to libraries reporting duplicates based on costs determined by the recommended study.

Recommendation 4: We strongly encourage OCLC to consider increasing credits to libraries for upgrades, enhancements, and enrichments to records in the database.

As more libraries move to systems that will allow them to download records from the catalogs of other libraries, and Internet searching of catalogs becomes easier, it is possible that we will see a decrease in upgrading of any less-than-adequate records found in the bibliographic utilities. Giving significantly higher credits for upgrading and enriching the database will give incentive for more member participation. This credit needs to be concomitant with the cost of doing the work. We determined that upgrading acquisition-level records was almost $25.00 per record. This falls within the range of $15.00 to $30.00 suggested by Steinhagen and Moynahan (1998). They go on to indicate that there is a strong relationship between cooperation and economic reimbursement. To avoid undermining the cooperative spirit behind internationally shared cataloging, more realistic and appropriately priced credits need to be given to the libraries providing the intellectual content of the database.

Recommendation 5: We encourage vendors to take the opportunity to include value-added information in the records they produce.

John Riemer of the University of Georgia suggested that vendors are in a unique position to give value-added service at the point of record creation. If vendors have scanning equipment, they could readily add tables of contents to their records with minimal increase of the labor costs. This would provide additional information to prospective buyers and act as an advertising technique, thus boosting sales. The added access would also give libraries more reason to look favorably on vendor records.

Works Cited

Beall, Jeffrey. 2000. The impact of vendor records on cataloging and access in academic libraries. *Library Collections, Acquisitions & Technical Services* 24: 229–37.

Erbolato-Ramsey, Christiane, and Mark L. Grover. 1994. Spanish and Portuguese online cataloging: Where do you start from scratch? *Cataloging & Classification Quarterly* 19, no. 1: 75–87.

Ferguson, Bobby. 1991. Upgrading minimal-level records for a statewide database. *Technical Services Quarterly* 9, no. 2: 31–37.

Fox, Mary Anne, and Barbara G. Preece. 1991. Upgrading minimal level monographic records: A study and conclusions. *Technical Services Quarterly* 8, no. 4: 25–35.

Handman, Gary. 1991. I lost it (or found it) at the movies: Public service implications of minimal-level cataloging of audiovisual materials. *Technical Services Quarterly* 9, no. 2: 39–47.

Horny, Karen L. 1991. What you see is what you get (so what do you get?): An administrator looks at minimal-level cataloging. *Technical Services Quarterly* 9, no. 2: 7–19.

Intner, Sheila S. 1994. Taking another look at minimal level cataloging. *Technicalities* 14, no. 1: 3–5, 11.

Johnson, Judith J., and Clair S. Josel. 1981. Quality control and the OCLC data base: A report on error reporting. *Library Resources & Technical Services* 25, no. 1: 40–47.

Marr, Russell. E-mail to the author on Oct. 26, 1998. Marr noted that it was not clear why the vendors lack access to these files.

McCallum, Sally. E-mail to the author on June 5, 1997, as forwarded to the Program for Cooperative Cataloging list, September 30, 1999, by Ana Cristan, Acting Team Leader/BIBCO Coordinator, Library of Congress.

O'Neill, Edward T., Sally A. Rogers, and W. Michael Oskins. 1993. Characteristics of duplicate records in OCLC's online union catalog. *Library Resources & Technical Services* 37, no. 1: 59–71.

Patton, Glenn. 1991. OCLC's long association with less-than-full cataloging. *Technical Services Quarterly* 9, no. 2: 21–30.

Peet, Terry. 1999. Interviewed by L. Shedenhelm. SALALM Conference, Nashville, Tenn., June 3, 1999. Peet is the former head of the Hispanic Acquisitions Section at the Library of Congress.

Preece, Barbara G., and Mary Anne Fox. 1992. Preliminary LC records for monographs in OCLC. *Information Technology and Libraries* 11, no. 1: 3–9.

Romero, Lisa. 1994. Original cataloging errors: A comparison of errors found in entry-level cataloging with errors found in OCLC and RLIN. *Technical Services Quarterly* 12, no. 2: 13–27. Compare to Cynthia C. Ryan (1978).

Ryan, Cynthia C. 1978. A study of errors found in non-MARC cataloging in a machine-assisted system. *Journal of Library Automation* 11, no. 2: 123–32. Compare to Lisa Romero (1994).

Sercan, Cecilia S. 1994. Where has all the copy gone? Latin American imprints in the RLIN database. *Library Resources & Technical Services* 38, no. 1: 56–59.

Stamm, Andrea L. 1996. Minimal level cataloging: Past, present, and future. In *Technical services management, 1965–1990: A quarter century of change and a look to the future*. Edited by Linda C. Smith and Ruth C. Carter. New York: Haworth Press.

Steinhagen, Elizabeth N., and Sharon A. Moynahan. 1998. Catalogers must change! Surviving between the rock and the hard place. *Cataloging & Classification Quarterly* 26, no. 3: 3–20.

Wanninger, Patricia Dwyer. 1982. Is the OCLC database too large? A study of the effect of duplicate records in the OCLC system. *Library Resources & Technical Services* 26, no. 4: 353–61.

Weitz, Jay. 2000a. E-mail to the author. June 9, 11:09 A.M.

———. 2000b. E-mail to the author. June 9, 11:58 A.M.