

A Survey of Batch Cataloging Practices and Problems

PHILIP YOUNG

Virginia Tech, Blacksburg, VA

Abstract.

Groups of bibliographic records are added to library catalogs with increasing frequency. Batch cataloging requires knowledge of bulk record transfer as well as current cataloging standards. While more efficient than cataloging items individually, batch cataloging requires different skills and creates new challenges. Responses to a wide-ranging online survey document the workload, tools, practices, and problems of batch cataloging. The unique characteristics of electronic resources affect many aspects of batch cataloging. Survey respondents lack consensus on how to share improved batch records, and recent trends in the bibliographic environment seem to make a networked solution less likely.

KEYWORDS batch cataloging, vendors, bibliographic record quality, cataloging tools, sharing, cooperative cataloging

Address correspondence to Philip Young, University Libraries, Virginia Tech, 560 Drillfield Drive, Blacksburg, VA 24061. E-mail: pyoung1@vt.edu

Batch cataloging in libraries has become increasingly common in recent years. This is in part due to large sets of electronic resources that need to have bibliographic records entered in the catalog when a subscription becomes valid. Machine-readable cataloging (MARC) records, sometimes by the hundreds or thousands, are obtained, edited, and entered into the library's online catalog at one time. While problems with individual records can be identified in batch view and corrected prior to or after load by a headings report or authority control vendor, all records are not examined individually as in traditional cataloging practice. The continuing proliferation of information resources, especially in electronic form, has required this workflow, which is more efficient than older methods but presents new and complex problems.

The term "batch cataloging" as used in this article refers to obtaining (or creating), transferring, manipulating, and editing groups of MARC bibliographic records. Those performing this work must know cataloging standards as well as how to transfer records into the local catalog. While it is possible for staff in other areas of the library to load records into the catalog, they are not always trained in cataloging practices. Those performing batch cataloging are also responsible for the integrity of the metadata and are therefore usually part of the cataloging department.

Batch cataloging was added to the author's responsibilities in 2009, and since then the practice has generated numerous questions about how other libraries address related problems. How significant is the batch cataloging workload at other libraries, and who is responsible for it? What tools and methods are employed to improve records? How often are holdings set in the Online Computer Library Center (OCLC)? How is authority control addressed? What is the effect of discovery layers on batch cataloging? How are duplicate records detected? How many catalogers are creating record batches? Is there an automated way to determine which of multiple links on an e-book record are valid for a library and which are not? And perhaps most importantly, how can libraries improve the quality of record batches and share those improvements with others? As the frequency and volume of batch cataloging increases and staffing levels decrease or remain steady, answers to these and other batch-related questions will be of interest to libraries desiring to improve both workflow and record quality.

Literature Review

While a number of articles have been published about batch cataloging, this literature review focuses on the most recent articles relevant to one or more survey questions. No surveys focused on batch cataloging practices and issues in general have been conducted before, though Kemp (2008) surveyed librarians about MARC record services for serials. Respondents indicated that they used a variety of methods to manipulate and improve records, including the free software MarcEdit (<http://people.oregonstate.edu/~reaset/marcedit/html/index.php>), scripts, global update in the integrated library system (ILS), and individual record editing. The survey documented a wide range of record errors, in addition to problems associated with duplicate and brief records, and about 40% of respondents expressed concerns about record quality.

Martin and Mundle (2010) reported a case study of managing and improving the quality of bibliographic records for a large e-book collection in a consortial setting. They created a discussion list for the consortium and shared record quality improvements, though not authority cleanup. Collaboration between the consortium and the vendor "to improve records before receiving them was the most productive route to quality data in the catalog" (p. 234).

Mugridge and Edmunds (2009) relate a batch-cataloging workflow and bring to light a wide variety of issues. While a significant increase in usage results from batch loading title-level

records, the process is a complex one with technological and organizational challenges. The latter requires upgraded skills, accurate documentation, and improved inter-departmental communication.

Wu and Mitchell (2010) discuss batch cataloging of e-books using MarcEdit, as well as issues raised by vendor-supplied records and the impact of guidelines for provider-neutral e-book records. Librarians rather than paraprofessionals performed most batch cataloging due to the complexity of the process. The library's shift to an e-book knowledgebase with a MARC record service, while not solving problems of record quality and duplication, streamlines workflow so that work can be delegated.

Chen and Wynn (2009) surveyed academic libraries to determine if and how e-journals were cataloged and found that 30 of 36 libraries were using record batches. The poor quality of vendor records was a common complaint. Twenty-nine percent of libraries rejected MARC record services as too expensive, and 52% did not set e-journal holdings with OCLC.

Martin (2007), in an overview of e-book cataloging problems, discussed delays in offering title-level access in the catalog and documented several common errors in records. Martin also found that when vendor records are used, holdings are often not recorded in OCLC, resulting in a loss of reliable information for interlibrary loan and less sharing of catalog records.

Sanchez, et al., (2006) described a method for improving a large set of e-book records, including authority control, before loading them into the catalog. In addition to documenting numerous record-quality problems, they emphasized the importance of interdepartmental communication. While Sanchez, et al., used an ILS report to identify authority work needed, Finn (2009) described a process using an authority vendor and MarcEdit to complete authority work before loading records into the catalog.

Methodology

The author created an online survey containing a wide range of batch cataloging questions in order to provide a landscape view of the practice that would also indicate possible avenues for future research. The survey responses document the workflow, tools, practices, and problems of batch cataloging. Because the survey focused on larger issues associated with batches, there was no attempt to document specific errors in MARC records. While the survey was focused on e-book batch records, some questions were format-neutral. The survey did not distinguish between one-time and continuing batch cataloging.

The online survey asked 31 questions in either multiple-choice (or "select all that apply") or short-answer (free text) format. This article presents a subset of the survey (22 questions, see Appendix), with some questions omitted in the interest of space and relevance. The survey begins with questions about the respondents and their libraries, moves on to questions about tools and practices, and ends with "big-picture" questions and an opportunity for open comment. The survey was announced by e-mail on three listservs, BATCH (administered by the author, with 104 subscribers), MARCEDIT-L (694 subscribers), and AUTOCAT (about 6,200 subscribers). The online survey utilized an in-house survey instrument and was active for two weeks, from November 9-23, 2010. A reminder e-mail was sent to the same listservs a few days before the survey closed.

The survey recorded 128 unique respondents, though some respondents did not answer every question. Therefore the number of responses varies by question, and is reported with the results below. Respondents were a self-selected group. The open comments at the end of the survey are incorporated into the results wherever possible.

Results

Respondents and Their Libraries

Respondents to the survey were overwhelmingly academic librarians located in the United States: 102 of 128 respondents (80%) worked in academic libraries, while 17 (13%) worked in public libraries and 6 (5%) worked in special libraries. Additionally, two respondents worked for consortia, and one for a school library. One hundred eight of 128 respondents (84%) were located in the U.S., with the remainder from Canada, the United Kingdom, and Australia. One hundred nine of 128 respondents (85%) held librarian status (defined as possessing a M.L.S. or equivalent degree), while 19 (15%) were paraprofessionals.

The survey attempted to address the workload associated with record batches by asking questions about the number of people involved in batch cataloging at each library and the percentage of time spent on it (see Tables 1 and 2). In the comment section at the end of the survey, two respondents clarified that they spent much less than 25% of their time on batch cataloging.

Table 1. Number of batch catalogers per library.

Number of catalogers	Responses	Percentage
1	51	40%
2	36	28%
3	30	23%
4 or more	11	9%
Total	128	100%

Table 2. Percentage of time spent working with batch records (choose closest answer).

Percentage of time	Responses	Percentage
25%	96	77%
50%	20	16%
75%	7	5%
100%	2	2%
Total	125	100%

A question about types of material represented in record batches allowed for multiple selections, and the numbers and percentages are from a total of 128 respondents (see Table 3). E-books were by far the most frequently batch-loaded record format, selected by 116 respondents (91%). A wide variety of formats, however, are batch loaded. In the “other” category, the most common responses identified e-journals, various audio-visual materials (such as video, music, and audio books, both online-accessible and in a physical carrier), and government documents.

Table 3. Types of material batch cataloged (check all that apply; 128 responses).

Type of Material	Responses	Percentage
E-books	116	91%
Print books	64	50%
Online video	51	40%
Microform	37	29%
Maps	16	13%
Other	50	39%

Tools

Five survey questions asked respondents about the tools used to acquire and manage MARC record batches. While many record sets are obtained through a service such as WorldCat Collection Sets or downloaded directly from a vendor's web site, MARC records can also be obtained in other ways. Asked whether they had used the Z39.50 protocol to obtain records from another catalog, 46 of 126 respondents (37%) said yes, while 80 (63%) said no. A slightly greater number have used OCLC Connexion's batch-processing capabilities to create and export a batch, with 54 of 127 respondents (43%) answering yes, and 73 (57%) answering no. However, in a later question about setting holdings in OCLC, 8 of 95 respondents stated that their libraries were not OCLC members. Therefore, some of those who answered "no" did so because OCLC Connexion was not available to them.

Respondents were asked to identify the editing tools they use, and/or name other tools (the number of responses and percentages are not cumulative because multiple selections were permitted). Of 128 total respondents, 107 (84%) selected MarcEdit. Fifty-seven (45%) used their system's global update function, 26 (20%) used Microsoft's Excel spreadsheet application, and 24 (19%) named another tool. Of the "other" responses, 4 respondents mentioned MARC Report, while 2 respondents mentioned Perl, shell scripts, Connexion, OCLC macros, and Bibliofile. Receiving one mention apiece were MARC Magician, home-grown ruby scripts, emacs macros, TextPad, Notepad++, gedit, MARCTOOL, UNIX editor, NANO, MARC Global, Regex Buddy, Mitinet, and load profiling (the subject of another question).

Respondents were asked how often they used regular expressions, a simple computer language used to identify records or parts of records for editing or manipulation. Regular expressions can be used in MarcEdit and the global update function of some library systems. Of 124 respondents, 16 (13%) selected "frequently," 40 (32%) "sometimes," and 68 (55%) "never."

Duplicate record detection is one of the thorniest problems in batch cataloging. Respondents were asked about their methods of detecting duplicates within a single batch and for records already in the catalog. Results should be viewed with caution, because the free-text answers were sometimes hard to interpret, respondents often did not indicate which question they were answering (or both), and there was some confusion over the meaning of "duplicate." While the intended meaning of duplicate was an exact copy requiring deletion, some respondents interpreted duplicate to mean multiple e-book records for the same content but from different publishers or vendors. The author attempted to categorize most of the 115 responses and eliminated others due to lack of clarity. Fifty respondents (43%) mentioned that load tables identified duplicates based on a MARC tag of 001, 010, 020, or 035. Twenty-three respondents mentioned a headings report or other information generated by the load table. It is likely these responses were referring to the same or a similar process, and these categories could be combined for a total of 73 (63%) responses. Seven respondents used a visual check for deduplication, three used scripts, two used Excel, and ten had no deduplication method. For deduplicating records within a batch, ten respondents mentioned MarcEdit, while six had no method.

Practices

The most significant recent change in the batch cataloging of e-books is the transition to provider-neutral records (Program for Cooperative Cataloging, 2009b). Previously, a separate e-book record for each vendor or publisher was entered into OCLC, resulting in records that varied in quality and were difficult to find. The provider-neutral standard allows multiple links (MARC 856 fields) to be attached to one record and eschews vendor-specific information in other fields. This means that some batches must be edited to delete links to e-books that the library cannot access. The survey asked how records with multiple links were being addressed. There were 104

responses to the question “Since provider-neutral e-book records can have two or more links to content, do you have a way of checking which links are valid for your library, or do you simply delete all links that aren’t from that particular provider?” Seventy-two respondents said they delete the links, although it was not always clear whether they were doing so as a batch edit or individually. Four of these 72 are using MarcEdit and/or regular expressions to do so. Ten respondents have not encountered the problem because they only receive records from the vendor or WorldCat Collection Sets, which includes only the vendor link. Three respondents indicated that they use a URL checker, two respondents said they add their own links, and one uses a link resolver to coordinate links. Other responses varied from checking the links manually to moving them to another field in case they were needed later.

Almost all e-books have print equivalents, and e-book records usually have a 776 MARC field for linking to the other format. However, in a batch of e-book records it is not efficient to check for a library’s print equivalent for each record. There were 118 responses to the question “What is your library’s approach to the 776 linking field in e-book batches?” (see Table 4). The majority of respondents do not address this field, have display turned off, or use the single-record approach that negates the need for a 776. Only two respondents check individual records for print equivalents in their library. In the “other” category, one library deletes all 776 fields. Some of the “other” responses were eliminated or moved to other categories.

Table 4. 776 Approach.

Policy	Responses	Percentage
We leave them as is/no policy	79	67%
We turn/have display off	23	19%
We use the single-record approach	13	11%
We check record by record whether display should be on or off	2	2%
Other:	1	1%
Total	118	100%

Respondents were asked whether they set holdings in OCLC for record batches, and why. Of 127 respondents, 28 (22%) answered yes, 42 (33%) answered no, and 57 (45%) answered sometimes/depends. Ninety-five respondents explained their answer. Twenty-three respondents said they set holdings if they had OCLC records or record numbers; vendor records did not have holdings set in OCLC. Fifteen respondents did not set holdings for e-book records, with some explaining that this was due to interlibrary loan (ILL) restrictions, vendor restrictions, or because titles changed frequently. Nine respondents upload library holdings to OCLC at regular intervals (weekly, monthly, or quarterly). Eight respondents were from libraries that were not OCLC members. Conversely, three respondents used WorldCat Local, making accurate holdings imperative. Three respondents had holdings set automatically by vendors, and one said holdings were set only for WorldCat Collection Sets. One respondent using the single-record approach attempts to set holdings on both record formats, but with some inconsistency. One respondent generates search keys from vendor records and attempts to set holdings for as many as possible. One respondent sets holdings only for purchased e-books but not for those leased.

Asked about the relationship between electronic resources management (ERM) systems and e-book batches, the vast majority of respondents (81 of 110, or 74%) answered that they do not use ERM for e-books. Sixteen respondents (14%) use ERM for e-books but not for batch cataloging, and 11 (10%) use ERM with some or all e-book batches. In the “other” category, two respondents said they were investigating ERM for e-books. Some responses in the “other” category were moved or eliminated.

Respondents were asked whether they use customized load profiles for batches, and if so, when or why. Eleven of the 105 responses were eliminated due to lack of clarity. Of the

remaining 94 responses, 76 (81%) answered yes, and 18 (19%) answered no. Of the affirmative answers, the most common reasons given were to determine record overlay and to create item or holdings records. Others used load profiles to strip or add MARC fields, add proxy information, and adjust control fields. Those answering “no” said that customized load profiles were not possible in their system, or they felt that using MarcEdit was easier than adjusting the profile.

Authority control for batches is accomplished in a variety of ways. All 128 survey respondents answered a question about their library’s approach to authority control for batches. Fifty respondents (39%) said that authority control was done in-house. Forty-seven respondents (37%) sent batches to an authority-control vendor and then did local cleanup. Fourteen respondents (11%) do not have authority control for their catalogs. Seventeen respondents (13%) gave a text response in the “other” category. Of these, 3 stated that their authority control was done in-house, but only occasionally, or with lower quality control, or only for certain fields. Three said batches were not treated differently from the rest of the catalog, from which new records were sent to an authority control vendor periodically. Two respondents said they use a combination of in-house and vendor-supplied authority control. Two said they do no authority control on batches. Six respondents said some but not all batches were sent to a vendor. One respondent said the library’s consortium handles authority control. In the comment area at the end of the survey, four respondents cited authority control as a problem, especially for names: “Authority control of names in vendor-produced e-book records leaves much to be desired. We are always having to clean them up.”

Big-Picture Questions

The survey asked four questions that situated batch cataloging in a larger context. Fewer responses were received than for previous questions, perhaps because they raise issues that are the responsibility of other people in the library. Furthermore, since the questions solicited a free-text response, compiling and categorizing answers was often challenging.

The recent advent of discovery platforms that can recognize electronic access has the potential to affect batch cataloging. Respondents with such a discovery platform, or who were planning to implement one, were asked what changes this might mean for batch cataloging. Fifteen of 56 total respondents said they did not know, including several using discovery platforms such as Summon, Primo, and Encore. Nine respondents said that the discovery platform has not affected batch loading (including Summon, Ebscohost integrated search, and VuFind). Of these, two said it had no effect on batch loading, in part because the knowledgebase is not comprehensive enough; one did not subscribe to the e-book knowledgebase; and one said title searches were problematic. Two respondents said they would still require full catalog records with Library of Congress subject headings. One respondent said Primo had changed the library’s policy because it de-duplicated records. One respondent said e-book records are purchased from Serials Solutions (the creator of Summon), and another respondent is considering discontinuing the batch loading of e-books. One respondent said, “We are currently moving to Serials Solutions for all our e-content MARC records. It has become too much work for us to keep up with all the updates to individual record sets and all the individual editing”, but another respondent using the MARC records from that service expressed serious concerns about the quality of records supplied.

Electronic resources are available to library users immediately, and there are often delays before a MARC record can be created, distributed, and entered into the catalog. Respondents were asked whether delays between print availability, online availability, and batch cataloging have caused ordering or collection-management problems in their libraries, and how these problems are addressed. Of 68 responses, 16 experienced no problems with delays, and 7 said problems were infrequent and handled on a one-by-one basis. Seven respondents download records in advance that display to users as on order, or use a brief record for this and then overlay

with a full record later. Three respondents said their acquisitions department checks print orders to ensure that there is not already online access. Three respondents reported having to wait for records to become available for e-resources: “Our collection managers would like the records available immediately, and we sometimes have to wait for vendors to create them.” Another respondent said that the “record set for our NetLibrary subscription does not always match what is available. Also, their MARC records don’t become available right away (takes up to a month).” Other respondents mentioned cataloging items individually on OCLC when records do not arrive in a timely manner, and a need for better notification from vendors when records are ready. In some cases, delays were caused internally due to lack of communication: “Our biggest headaches with respect to e-content (and in particular e-books) has less to do with getting/maintaining decent MARC records and more to do with communications between acquisitions and cataloging. Because e-book packages can be selected and acquired so quickly, there is often a significant and frustrating gap between the point of order and the point at which catalogers know what is now available and needs to be cataloged in the OPAC.” Another respondent said “we have worked with collections to improve communication on this issue.”

One problem with batch records is that when catalogers edit them to improve their quality, they are duplicating the effort of other libraries (to clarify, these are not local edits but edits correcting series treatment or typographical errors, adding classification numbers, etc.) The survey noted this problem and asked for practical suggestions to make this process more efficient. The 84 responses primarily fell into four categories. The first category includes record sharing and the use of consortia. In this category, 10 respondents suggested a repository of corrected record sets, though some were concerned about licensing issues. Five respondents are already using their consortia to share batch-record improvements. Two respondents mentioned that they had shared records for a specific set with another library. Another respondent also suggested this kind of peer-to-peer networking among libraries.

The second category involves working with vendors to improve the quality of records. Twelve respondents suggested reporting errors in batches to the vendor, with two of these offering specific examples of vendors correcting batches as a result of communication. Other suggestions in this category included vendors creating a central place for libraries to share record sets, and vendors hiring libraries to create records. Two respondents suggested reporting errors to vendors and then uploading records to OCLC to be enhanced and have headings controlled.

The third category involves requiring better record quality from vendors on the front end. Seven respondents said libraries should petition, pressure, or require vendors to provide full and accurate catalog records. An additional five respondents thought vendors should provide record customization for libraries. Two respondents called for standardization of control numbers and e-ISBNs, and one said more vendors should contribute records to OCLC.

The latter comment leads into a fourth category of respondents desiring better integration of vendor records with OCLC. Seven respondents made suggestions such as submitting corrections to OCLC, enabling record enhancement for batches, using local-holdings records, and improved set support from OCLC.

Finally, one respondent noted that the most time-consuming part of batch loading is authority control, and two respondents said they spend little or no time correcting batch sets.

The survey offered respondents an opportunity to mention other concerns or problems with batch MARC records, and/or to clarify earlier answers. These responses have been integrated into the response summaries above or the discussion below.

Discussion

Academic librarians in the U.S dominated the survey responses. This may reflect a greater volume of batch-cataloged resources, or a greater variety of them, in academic libraries. Respondents were overwhelmingly librarians (85%) rather than paraprofessionals (15%), perhaps

lending support to the view that the coordination required for batch cataloging often makes delegation difficult (Wu & Mitchell, 2010). As one respondent commented, batch loading reduces the number of materials that would otherwise go to copy cataloging, but “the work and follow up are done by higher level staff.” Librarians may be better able to facilitate communication with other departments, have skills related to record retrieval such as Z39.50 and OCLC Connexion batch processing, have experience with manipulation tools such as regular expressions, scripts, global update, and MarcEdit, and have access to load table configuration. However, batch cataloging is likely to be delegated as the process becomes more stable and routine at each library. Since each record batch is unique, librarians will likely continue to evaluate record quality, and both librarians and paraprofessionals will need to be able to manipulate large groups of records in addition to their cataloging skills.

Evidence of batch workload has not been previously available. This survey shows that a majority of libraries have two or more catalogers responsible for loading and/or editing batch records, but the majority spends 25% or less of their time on the process. These results may indicate that as a greater variety of resources become available in batches, the workload becomes more distributed, but the inherent efficiency of the process means that each person spends a small amount of time on it. However, batch processing for print or government documents may account for the number of personnel involved. At the author’s institution, for example, all or most records in these batches are checked individually, but that is not the case for e-book record batches. Record transfer itself takes little time, but large differences in time can occur depending on whether metadata quality is checked for each record individually or as a record batch. Additionally, the relatively small amount of time spent on batch loading and editing may not include time for policy and planning, maintaining documentation, and authority work.

While e-books were by far the most commonly batch-loaded material type, the survey also indicates that virtually every type of resource has been cataloged in batches. Batch processes will likely become more frequent due to the ever-increasing volume of resources, the transition from print to electronic, the marketing appeal of MARC records for vendor packages, and above all, the efficiency of the process.

It is hardly surprising that MarcEdit was by far the most popular tool for batch catalogers, since the MarcEdit listserv was solicited for survey responses. Of greater interest was the wide variety of tools listed by respondents, especially those involving scripting. This provides evidence of the usefulness of programming skills for cataloging functions and perhaps helps explain the predominance of more broadly skilled librarians over paraprofessionals as batch catalogers. While a majority of respondents never use regular expressions, a significant minority frequently or sometimes does. Scripts are also used for duplicate record detection, though respondents are largely dependent on load or headings reports for this purpose.

The relatively new provider-neutral standard for e-books raises the issue of how the matching of e-book records from different vendors will occur in local catalogs. Guidance on the provider-neutral approach defers to libraries whether to use single or separate e-book records (Program for Cooperative Cataloging, 2009b). However, because many libraries use the OCLC control number as an overlay point, it may be difficult to maintain a separate record approach for OCLC records. The provider-neutral policy does not suggest how the ILS might recognize library subscriptions to various vendor e-book packages. One scenario is that a load table will recognize the OCLC control-number match, and simply add the 856 associated with the specific vendor. However, this situation presents two opposing problems. On one hand, the original record in the catalog is “protected,” and a new 856 is added to the record, authority control and other record editing is preserved. On the other hand, the incoming bibliographic record may have updated fields that the catalog record does not, such as a summary, contents notes, or corrections. In this case it may be possible to adjust a load table to accept only new fields in addition to the 856 field. If a vendor is not an OCLC participant and uses its own control number, a different match point will be needed. ERM is a possible solution to this problem, but only 10% of survey

respondents are currently using it for e-book batches, and none mentioned its use in this regard. This question sought to discover automated means of checking which links offer valid access, and it appears that only those using a URL checker or link resolver are in this category.

E-book records commonly provide a linking field to the print format (MARC 776). Reciprocal linking fields on the print and electronic record versions allow library users to select their preferred format. With e-book batches of hundreds or thousands of records, however, it is often not possible to quickly determine instances in which the library also offers the print version. Doing so manually would undermine the efficiency of batch loading. Automated methods usually depend on ILS reports showing matches on ISBN, call number, or other number. Still, determining which 776 fields should display requires viewing individual records as well as adding a reciprocal 776 on the print record. For the 13 respondents using a single-record approach, this is not a problem, but it appears that most respondents have been unable to provide navigation between formats in their catalog. One concern is that print and e-book records for the same content might differ, such as in contents or summary notes, which might result in retrieval of only one of the records by a keyword search. A process to link formats via the 776 in this situation has been described (Simpson, Lundgren, & Barr, 2007) but may be too time-consuming or beyond the capabilities of some libraries. Some catalog discovery layers collocate resources available in different formats and may make manipulation of the 776 unnecessary.

Recording holdings in OCLC for batches is affected by several factors. Some vendors do not use OCLC records, and without OCLC control numbers, setting holdings for these records can be more time-consuming. A significant number of libraries do not set holdings if they perceive that the only purpose in doing so is for ILL, because most electronic resources are not eligible for loan. However, setting holdings would benefit library patrons using WorldCat as a means of discovery and would be of benefit if the library decides to use WorldCat Local as their catalog. Some libraries do not set holdings for particular sets because their catalog records are uploaded to OCLC on a regular basis. Setting holdings based on whether content is owned or leased reinforces the importance of communication between cataloging and acquisitions or collection development.

When libraries use vendor records for non-lending resources, they often have little or no interaction with the de facto union catalog, OCLC. The potential for large-scale cooperation on record quality, including authority control, is thereby lessened. Other record sources exacerbate this situation. Z39.50 is used to obtain records by 37% of survey respondents, and MARC record services and new bibliographic utilities are emerging as record sources. Additionally, three survey respondents were concerned about the licensing of MARC records as a barrier to sharing.

Discovery platforms have the potential to influence batch cataloging practice due to their ability to collocate resources automatically, and through their associated knowledgebases. As the volume of resources increases, and staffing in cataloging departments remains stable or decreases, outsourcing of tasks to knowledgebases and discovery services will likely increase. Knowledgebases serve administrative functions by tracking subscribed online resources, and by “recognizing” when new resources come online. The latter function is difficult to perform otherwise. Given the hundreds or thousands of resources available in a set, and the delays in the creation and distribution of MARC records, it is currently difficult to know about new resources or identify those not yet represented by a record. However, there are several problems with these services. As one respondent commented, knowledgebases are often incomplete, perhaps due in part to lack of vendor cooperation. And another respondent felt that the record quality of such services seems to be no better, and sometimes worse, than records obtained previously. Additionally, record licensing poses a barrier for record sharing and results in duplicate record creation by different entities. The cost could be a barrier to smaller libraries, with the MARC record service an additional cost above that of the knowledgebase.

While the survey question about the delay between the availability of online resources and the availability of MARC records for them was focused on ordering issues in the library, this

delay can potentially decrease catalog use. The online environment reverses the traditional cataloging workflow for print. When a subscription or a new e-book comes online, it is immediately available to library users, whereas print resources have traditionally been cataloged before they are made available. Therefore the timeliness of cataloging has increased in importance. If library users know that a vendor's newly available e-books are slow to be reflected in the library catalog, there is added incentive to bypass it. This is a difficult problem for libraries to address, because they are dependent on a vendor or utility for record delivery. While knowledgebases partly help address delays by offering brief records so that library users can discover resources, they come with many problems, as noted above. Additionally, as some respondents noted, delays are caused or exacerbated by lack of internal communication. More frequent use of print records as a basis for e-book records might minimize delays as well as improve record quality.

If one were to measure delays until the time a MARC record was more or less "finished," then authority-control work would likely be the most time-consuming part of batch cataloging, especially considering the number and quality of the records. While this survey did not solicit specific record-quality problems, a few respondents offered examples, and common issues are well-documented in the literature (Kemp, 2008; Sanchez et al., 2006). Vendor batches with poor or nonexistent authority control create a tremendous amount of work for libraries, especially for the 39% of survey respondents doing this work in-house. Some record sets may be excluded from authority control due to some combination of size, quality, or set changeability. For example, problems might be caused by patron-driven collections in which the majority of e-books have not been purchased by the library, or if a set needs to be deleted and then re-loaded. Applying authority control in these situations would expend additional time, money, or both. Large efficiencies would likely be gained by performing authority work at a more networked level, for example by a vendor or bibliographic utility.

Reducing or eliminating the duplication of effort in correcting batches of records is a difficult problem because it would involve a high degree of coordination between libraries, utilities, and/or vendors. Improvements to record batches, including authority control, seem to be infrequently shared, and even then rarely at a networked level available to all. Sharing within consortia makes workflow more efficient, but does not benefit libraries outside the consortia. Direct sharing of records can be prohibited by the licensing or record use policies of vendors or utilities. Libraries can make edits to the "master record" on a bibliographic utility. But some libraries cannot afford OCLC membership, and making corrections on OCLC would greatly reduce, if not eliminate, the time savings of batch processing. Corrections to record batches, including edits to individual records, take place in record-editing tools or in the library catalog external to OCLC. After cleaning up the local catalog, there is little incentive to repeat this work in OCLC. Batch upload into OCLC is possible but problematic. It is open only to PCC libraries, and record overlay depends on algorithms that do not always work properly or do not replace the fields needed. Sharing authority control would greatly enhance the effectiveness of batch cataloging, although it would likely add to delays in obtaining records. However, this may not be significant if records are improved in a multi-step iterative process, beginning with a brief record to register availability, to an initial MARC record, to a "finished" MARC record with authority work and other quality control completed. Working with vendors to improve records is successful in some cases but, in the author's experience, not in others. As one survey respondent advocated, vendors need to be convinced that record errors, as well as delays, affect the findability of the resource. Vendors have an incentive to enhance discovery, particularly for patron-driven collections. Guidance for vendors creating MARC records is available, which includes the suggestion that authority control be implemented (Program for Cooperative Cataloging, 2009a). The quality of many batch sets, however, seems to indicate that these guidelines are not followed. Requiring vendors to provide high-quality records seems problematic due to poor communication between collection development and cataloging departments in some libraries, in addition to the

time needed to fully examine a subset of records. Subscription decisions in libraries are primarily based on the content offered but need to incorporate the quality of metadata available as well. Better integration of vendor records with OCLC is dependent on the OCLC-vendor relationship, and while allowing for record improvements, still does not eliminate the duplicate work required. Greater effort toward cooperative cataloging at the “network level” for the benefit of all could make batch cataloging far more efficient.

Conclusion

Practices related to batch cataloging of MARC records are extremely variable and are influenced by differing record batches, editing tools, and local policies. Librarians are usually responsible for this workflow, probably because they are in the best position to facilitate communication with other departments, determine policies, and perform system functions such as adjusting load tables. As the workflow becomes established, batch cataloging will likely be delegated more frequently, especially where ERM and MARC record services are used. Those performing batch cataloging will need to build on their cataloging knowledge to learn new tools and methods of manipulating large groups of records. Increases in the number of electronic resources make it likely that batch cataloging will become more frequent. Automation of some batch-related tasks such as collocating formats and combining links for online access on a single record is currently low but seems likely to increase.

The unique characteristics of electronic resources—immediate availability, varied acquisition models, and lack of lending permissions—have implications for batch cataloging. Delays in the delivery of MARC records may lead more libraries to move to an incremental metadata model beginning with brief records for immediate access, followed by overlay by a full MARC record, and finally, record editing and authority control. Some libraries choose not to address record quality for sets of electronic resources that are leased, changeable, or part of a patron-driven acquisitions plan.

Survey respondents are divided on the best way to share upgraded batch MARC records. Greater use of consortia, peer-to-peer sharing, vendor communication, requirements for vendors, and involvement by OCLC are suggested. Batch cataloging of vendor records often reduces interaction with OCLC, due to the perception that holdings do not need to be set and the inefficiency of transferring improved records there. This trend, along with the use of Z39.50, a proliferation of other record sources such as MARC record services and new utilities, as well as concerns about MARC record licensing, seems to make large-scale sharing of record improvements less likely. A new emphasis on cooperative cataloging, including authority control, will be necessary to make batch cataloging more efficient.

Received: June 8, 2011

Accepted: July 22, 2011

References

- Chen, X. & Wynn, S. (2009). E-journal cataloging in an age of alternatives: A survey of academic libraries. *The Serials Librarian* 57(1), 96-110.
- Finn, M. (2009). Batch-load authority control cleanup using MarcEdit and LTI. *Technical Services Quarterly* 26(1), 44-50.
- Kemp, R. (2008). MARC record services: A comparative study of library practices and perceptions. *The Serials Librarian* 55(3), 379-410.

Martin, K. E. (2007). Cataloging eBooks: An overview of issues and challenges. *Against the Grain* 19(1), 45-47.

Martin, K. E. & Mundle, K. (2010). Cataloging e-books and vendor records: A case study at the University of Illinois at Chicago. *Library Resources & Technical Services* 54(4), 227-237.

Mugridge, R. L. & Edmunds, J. (2009). Using batchloading to improve access to electronic and microform collections. *Library Resources & Technical Services* 53(1), 53-61.

Program for Cooperative Cataloging. (2009a). *MARC record guide for monograph aggregator vendors*. (Includes revisions to September 2010). Retrieved April 11, 2011 from <http://www.loc.gov/catdir/pcc/sca/FinalVendorGuide.pdf>

Program for Cooperative Cataloging. (2009b). *Provider-Neutral E-Monograph MARC Record Guide*. (Prepared by Becky Culbertson, Yael Mandelstam, George Prager, includes revisions to September 2010). Retrieved April 11, 2011 from <http://www.loc.gov/catdir/pcc/bibco/PN-Final-Report.pdf>

Sanchez, E., Fatout, L., Howser, A. & Vance, C. (2006). Cleanup of NetLibrary cataloging records: A methodical front-end process. *Technical Services Quarterly* 23(4), 51-71.

Simpson, B., Lundgren, J., & Barr, T. (2007). Linking print and electronic books: One approach. *Library Resources & Technical Services* 51(2), 146-152.

Wu, A. & Mitchell, A. M. (2010). Mass management of e-book catalog records. *Library Resources & Technical Services* 54(3), 164-174.

Appendix

Batch Cataloging Survey

What type of library do you work in?

- Academic
- Public
- Special
- Other:

My library is located in:

- U.S.
- Canada
- U.K.
- Australia

Other:

I am a:

- Librarian (i.e, have a M.L.S. or equivalent)
- Paraprofessional/staff member

How many catalogers in your library are responsible for loading/editing batches of MARC records?

- 1
- 2
- 3
- 4 or more

How much of your time do you spend working with batch MARC records (choose closest answer)?

- 25%
- 50%
- 75%
- 100%

I use the following tools to edit MARC record batches (check all that apply):

- MARCEdit
- Global Update in my ILS
- Excel

Other:

Do you use regular expressions (RegEx) to identify and edit MARC data?

- Frequently
- Sometimes
- Never

Since provider-neutral e-book records can have two or more links to content, do you have a way of checking which links are valid for your library, or do you simply delete all links that aren't from that particular provider?

What is your library's approach to the 776 linking field in e-book batches?

- We turn/have display off
- We check record by record whether display should be on or off
- We leave them as is/no policy
- We use the single-record approach
- Other:

What material types are batch loaded by your library (check all that apply)?

- E-books
- Online video
- Microform
- Print books
- Maps

Other:

Do you set holdings in OCLC for record batches?

- Yes
- No
- Sometimes/Depends

Explain your answer above:

Have you used OCLC Connexion's batch processing to create and export a batch?

- Yes
- No

Have you used Z39.50 to retrieve records from another catalog?

- Yes
- No

What is the relationship between ERM and e-book record batches at your library?

- ERM for e-books is linked with some or all batches we load
- ERM for e-books is separate from our batch loads
- We do not use ERM for e-books

Other:

Do you have a uniform practice for control numbers (001, 035, etc.) used in your catalog? Describe.

Do you use customized load profiles/programs for batches? If so, when or why?

How is authority control for batches addressed?

- Sent to a vendor followed by local cleanup
- Authority control is done in-house
- We do not currently do authority control
- Other:

What are your methods for detecting duplicate records (1) within a single batch, and (2) for records already in your catalog?

If you have or are planning to purchase a discovery platform like Summon, with a knowledgebase that recognizes access to e-books, how does/will this affect batch loading of e-books?

If delays between print availability, online availability, and the loading of records into the catalog have caused ordering/collection management problems in your library, please describe how these are addressed.

Often many libraries are locally correcting the same batches, which results in a tremendous duplication of effort. In your view, what would be the most practical way to make this process more efficient?

Please feel free to add other concerns or problems with batch MARC records, and/or clarify answers above: