Cleanup of Netlibrary Cataloging Records: A Methodical Front-End Process

Elaine Sanchez
Leslie Fatout
Aleene Howser
Charles Vance

ABSTRACT. Electronic resources and ebooks in particular, have become a very important source of information for library patrons. When our library was given access to more than 20,000 ebooks, we were faced with bibliographic records of unknown quality. To provide high-quality records in a timely manner, we identified as many potential problems as we could, worked with reference staff to create the best PAC displays, and created efficient record-editing methods to address these issues prior to loading the records in our database. This article documents that process and describes the MarcEdit, Word, and Excel strategies used to methodically correct and improve these records. It also offers practical solutions and procedures for database maintenance and quality control for NetLibrary or any outsourced cataloging records. The future of ebooks and other related cataloging issues, including authority control, are also discussed as points that remain to be addressed. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com> © 2006 by The Haworth Press, Inc. All rights reserved.]*

Elaine Sanchez is Monographs Cataloging Librarian/Unit Head; Leslie Fatout is Library Systems Coordinator/Circulation Librarian; Aleene Howser is Head Monographs Cataloging Assistant; Charles Vance is Database Management Services Librarian/Unit Head, all at Alkek Library, Texas State University-San Marcos, 601 University Dr., San Marcos, TX 78666-4604.

INTRODUCTION

NetLibrary (www.netLibrary.com), a division of OCLC, is the major provider of electronic books (ebooks) to the library community,[2] with more than 60,000 ebooks from more than 400 publishers, covering all subject areas, and serving more than 6,700 libraries.[3] Ebooks are published works such as research materials, reference books, and textbooks that have been converted into digital format for electronic distribution. They are an important supplement to print materials as they provide instant access for patrons in remote locations.[4] Ebooks offer other advantages such as full-text searching; instant linking to related resources; no risk of theft, damage, or loss; potential savings in processing costs; and no physical space requirements.[5]

Furthermore, while the debate rages about the future roles of printed books versus electronic resources, patrons are growing increasingly more reliant on electronic resources for their information needs. Given this fact, it is imperative that the library offers access to ebooks, ejournals, and all other remote electronic information.

In 2002, through TexShare (www.texshare.edu), Texas' statewide resource-sharing program, a NetLibrary collection of more than 20,000 electronic books was made available to member institutions. The ebooks were a welcome addition at a time when the enrollment at Texas State University-San Marcos (www.txstate.edu) was growing rapidly (from 22,471 in 2000 to 26,375 in 2003) and there was a strong emphasis on using technology to improve and expand access. However, only the most curious and adventuresome library users were aware of the ebook collection because the catalog, beyond which many users never venture, did not disclose it. Therefore, upon learning that TexShare offered NetLibrary MARC records to its member libraries at no cost, our university librarian initiated a high-priority project to load bibliographic records for the ebook collection into our catalog.

In the summer of 2003, reference librarians met with the system librarian and catalogers to plan for the addition of ebook records to the public access catalog. Reference staff wanted to ensure that these

cataloging records would clearly distinguish electronic from print resources, and provide simple, direct access to the ebooks.

Cataloging staff were concerned about the quality of the NetLibrary records; this would be the first time that bibliographic records which we had not cataloged would be loaded into our database. Our cataloging standards are high, with full encoding level, AACR2, ISBD, careful authority control, consistent series authority work, and complete classification and subject access. We wondered how NetLibrary records would compare to our own, and how they would affect the integrity of our carefully constructed bibliographic and authority databases.

We found the richest source of information to be Autocat, where many problems were listed and discussed. We reviewed published literature that covered the cataloging of ebooks and discussed fields and standard values to include, but at the time did not address cataloging problems in the set of NetLibrary records we were going to acquire.

At the same time, the system librarian communicated with the Data Research Associates (DRA) Classic user group to inquire about system-specific problems and solutions other sites might offer. Points that were mentioned included dealing with the lack of physical items, hard copy records for the same titles, and variations in cataloging rules and practice. There are a couple of programs that can create or message pre-existing MARC data; however, several of the users recommended Terry Reese's MarcEdit program as a gem of a tool which made batch editing records a breeze. Web site citations that describe the other programs are included in the bibliography at the end of this article.

Learning about the problems within the records, we knew we had to find a way to get them corrected. How would we deal with this task? Would there be a simple way to identify the different types of problems? Who would be responsible for finding and correcting the problems?

It made sense to approach the cleanup systematically instead of dealing with records on an "as-found" basis. First, having the records isolated from the entire database allowed the task to proceed more quickly and efficiently. Second, doing most of the cleanup work before the records were loaded made them more accessible and useful to the patrons as soon as they appeared in the catalog. Third, the tools that are available with our current system (DRA Classic) are limited in scope and functionality. It would be more efficient to use programs that had greater search, sort, replace, and edit capabilities: MarcEdit, Microsoft Word, and Microsoft Excel.

We have succeeded in creating a methodical front-end process that raises the quality of the NetLibrary records, as nearly as possible, to our

own local standards. A link to detailed procedures appears in the notes at the end of this article.

SURVEY OF THE LITERATURE

A survey of the literature was performed on OCLC, the Internet, and in the *Library Literature & Information Science* full-text database regarding the quality and cleanup of NetLibrary records. The results were

- some information, articles, and Web sites on MarcEdit and two other MARC-editing software programs
- a small number of functional applications of these utilities in use by libraries, explaining in broad terms how certain features could be used to perform various types of edits, such as globally editing the 049 field
- a few articles that had brief mentions of non-traditional software applications used to perform editing prior to loading MARC records in catalogs.

We found no article or information on the overall process of using non-traditional, non-ILS supplied editing utilities to correct MARC records prior to loading them in the catalog. Neither was there any information on the actual procedures for these workflows, nor any discussion on the quality controls and editing standards required to bring the records up to the quality level of records already existing in the catalog. This article presents the entire process, as experienced at Texas State University-San Marcos, which can be emulated for any outsourced MARC record cleanup project.

PROBLEMS AND RAMIFICATIONS

Reference staff were guardedly positive about adding bibliographic records for NetLibrary titles to the catalog, provided the following issues were addressed:

- The OPAC display must be patron-friendly and unambiguous.
- The link to the online titles must be as simple and direct as possible.
- Users must be able to search by keyword and limit the result to ebooks.

Reference staff were also concerned about the absence of item records that caused the system-generated note: "The library currently has no holdings for this title."

Cataloging staff, on the other hand, were cautious about loading NetLibrary records, sight unseen, into the database. This was our first experience with accepting someone else's cataloging en masse and without local cataloging oversight. We wanted to learn everything we could about the quality and potential problems of these records. After the first set of records was loaded in August 2003, we had the search capabilities to evaluate the records. Records were loaded at night, when staff was not cataloging, so the records appeared in sequential database control number (dbcn) order. This enabled staff to identify NetLibrary records that were retrieved when performing random searches for possible problems and errors. One way to directly retrieve NetLibrary records was by the dbcn as we knew the range of numbers for these titles. Another simple way was by the subject "Ebook," as this had been added to all the NetLibrary records. Both of these searches allowed staff to limit searching to only NetLibrary titles. Our initial searches in the records uncovered various issues such as

- A title search for initial articles in any language brought up a few records with incorrect filing indicators.
- Searches for authors, corporate bodies, and other access points showed that these were not linking to the assigned authority records in our database because of minor typos, capitalization discrepancies, diacritics, incorrect subfield codes, punctuation, or other problems.
- Limitations in our system caused series with initial articles not to link with their authorized headings.
- NetLibrary records with call number formatting problems; we used the 050 and 090 section of OCLC's *Bibliographic Format and Standards* as examples.

We also wanted to learn what other cataloging agencies had found. We searched Autocat archives to see what messages had been posted on the topic of NetLibrary records. There were disturbing problems with records created in 2001:

- Duplicate records
- 7xx fields stripped from records
- Acceptable subject headings stripped from the full OCLC record, sometimes leaving only one subject heading

- MESH subjects retained, but LCSH deleted
- Errors in 245s; non-English language cataloging description
- Main entry vs. title entry errors, etc.

Cataloging discussion lists detailed NetLibrary record problems in 2002 including the following:

- Treatment of single serial issues and single volumes of multipart monographs that varies from our local practice
- Records without call numbers
- Print format had accompanying CD or software and a note or 300 $e accompanying material text reflecting this, and the NetLibrary record retained this note
- Duplicate call numbers for records in classed-together series
- The possibility of 126 duplicate records in the NetLibrary set released by SOLINET
- Authority headings of all types conflicting with OCLC and LC authority files.

Since 2002, the NetLibrary records have been corrected and redistributed to address the 2001 problems and some of the 2002 problems. Later records, made available in 2003 and 2004, demonstrate continued improvements in quality. Duplicate records are rare, and there are fewer records lacking call numbers. However, the following problems still remain:

- Formatting errors in call numbers
- Treatment of single serial issues and single volumes of multipart monographs differs from our local treatment of print serial and monographic set equivalents
- Print format had accompanying CD or software and the 300 field $e accompanying material text was retained in the electronic format record
- Duplicate call numbers for records in classed-together series
- Authority conflicts with OCLC, LC, or local authority files.

In our review of NetLibrary records we have identified continuing problems (see Figure 1). We are working on our processes to detect and fix these errors before they get into our database where they are harder to find and fix. It is important to address these errors because they create barriers to consistent and correct access.

FIGURE 1. NetLibrary Problems and Ramifications

Problems	Ramifications
No LC call number in an 050 or 090 field, or multiple call numbers	We classify all electronic resources so that patrons or staff searching by call number can browse the shelflist for similar materials, no matter what the format of the work. Any record lacking a call number, or with multiple call numbers, or with an incomplete or incorrectly formatted call number, is unavailable for accurate browsing. Additionally, ebooks lacking a call number will not be represented in accreditation reports for collection development which select titles based on the call number field.
050 or 090 field with multiple $a's or lacking subfield b	
Spacing and formatting problems in 050s and 090s	
Authority conflicts of all types	Authority conflicts create multiple headings for a single access point, diluting the patron's search and causing confusion in search results
Records are not updated to reflect most current AACR2 rules, LCRIs, including the GMD change from computer file to electronic resource	AACR2 changes, MARC tagging errors, and descriptive cataloging errors and typos make access points on the record index incorrectly and display in a manner inconsistent with other new materials, or misidentify the ebook, hiding it from the patron
Various descriptive cataloging typos in indexed fields such as 245	
Tagging errors in subject fields, especially 650 and 651 errors	
Accompanying material in 300 subfield e	Ebook records that purport to have accompanying materials mislead the patron into thinking that physical materials are also available with the ebook
Annual serials cataloged as individual monographs, often with the serial title entered in the 246 rather than the 730, often with incorrect call number formatting	Annual serials and parts of monographic sets cataloged as individual works conflict with existing classed together treatments in our online catalog, causing duplicate entries, variant call numbers for the same works, and patron confusion
246 fields with initial articles	All initial article entry problems hide the actual access point from user searches
No method to isolate and search only for ebooks except for the 710 NetLibrary, Inc. heading	NetLibrary is a commercial name, not a typical type of qualifier for format. As more ebooks from other suppliers are added to the online catalog, a broader format search is needed
Presence of 653 and 655 fields	These are not authorized access points in our online system and can cause conflicts with existing LCSH
Problems that are specific only to DRA because of its system limitations, including: series 440 with initial articles and changing the tag 776 to 876 to force correct PAC display	In DRA, series in 440 fields with initial articles do not index with their authorized form, creating multiple index hits when the patron searches. Another drawback in DRA PAC display causes 776 fields to display as part of the 260. For this reason we edit them to become 876's which display correctly as related title entries
856$z, which is a public explanatory note, does not display in our catalog, and the 856$3 text "Bibliographic record display" is not meaningful	This text is intended to provide the link to the ebook. Without it, there is no clear way for the patron to make that connection

Figure 1 lists only the *basic* ramifications of the NetLibrary errors we identified. While some of these errors are not as critical, those that impair patron identification and use of desired materials *are* critical.

AUTHORITY CONTROL

In addition to the bibliographic description issues, authority control was a particular area of concern. Large batches of records were being loaded into our database with no systematic method to verify headings. Had the records been cataloged locally, we could have felt secure that between the efforts of our staff and the maintenance procedures we have in place, the headings entered would be up to our standards. In this case, however, we were dealing with authority records of unknown quality.

The DRA system provides us with reports of headings ("index dumps") that have been loaded into the database but not authorized. We use these reports to perform routine authority control, and with some minor adjustments to our normal procedures, we determined that they could be used for the NetLibrary record loads as well.

The index dumps are set to run at the end of the work week to catch the headings entered during the course of normal cataloging. Because the NetLibrary records are loaded when other cataloging work is not being done, we can run a special dump to isolate the NetLibrary records and their corresponding unauthorized headings.

This special index dump created a large set of authority headings which were imported into an Excel spreadsheet. Headings were then divided among cataloging units and are being searched in our catalog and OCLC's authority file for any necessary conflict correction, export of authority records from OCLC, or creation of local authority records.

SOLUTIONS IDENTIFIED

Our system librarian headed up the NetLibrary record load project. In February 2003, she initiated the record-review process by asking catalogers what kinds of problems they found with these records and how the records should be edited to best serve the patrons and match the quality of the records in our existing database. Working with printouts of some of these records, we were able to determine how they were cataloged, which fields and texts were used, what types of errors we could

FIGURE 2. Solution and Standard Text with Rationale

Solution and Standard Text	Rationale
Adding an ebook extension to the call number Standard text used: E-book	To appear on the first level hitlist to alert the patron that these are ebooks, not physically available in the collection
Adding a local 690 access point for ebook as a material type Standard text used: Ebook	To enable patrons to retrieve ebooks or use this access point as a way to limit a keyword search on a given subject or with a specific title search to only ebooks
Setting up an authority record for Ebook to include variant form cross references and a scope note	To explain the difference between Ebook and the LCSH heading Electronic books
Determining what the encoding level should be in order to differentiate them from our locally created records and identify them as NetLibrary records Standard text used: Encoding level of "3" which local system reports indicated had only been used 2 times in our database	So that the bibliographic records would not be purged in the future when the DRA bibliographic record purge program is used to delete records without holdings
Changing the 245 $h [computer file] in these records to the newer AACR2 GMD form of [electronic resource]	To match all existing records with AACR2 GMD of $h [electronic resource]
Deleting 653s, 655s, and all other 6xx fields with second indicators other than 0 or 1	Our catalog does not have an authority control structure for these, and they often conflict with LCSH
Not adding a 538 technical specifications note, but keeping it if it is already in the record	These notes add no new information to the record that is useful to the patron or that describes the work
Changing the material type from "a" (language material) to "m" (computer file)	To allow NetLibrary records to have the right material display icon in the Web catalog
Adding the Eproxy server prefix before the url for the ebook	Authorized patrons would automatically be logged on to NetLibrary
Isolating all unauthorized headings from these records into an electronic log	To maintain a discrete file of NetLibrary record headings requiring future authority control work
Changing 856$3 from "Bibliographic record display" Standard text used: Online version	The meaningful text provides a clear and direct hyperlink to the actual ebook
Providing separate cataloging records rather than one record with multiple holdings for different formats	The GMD and "E-book" after the call number, which display on the first-level hitlist, clearly show when the title is an electronic format

identify, and how they should be modified. By June, the following solutions and standard texts had been identified (see Figure 2).

We also noted that these records were in OCLC-MARC format, not MARC21, as there were certain variable and fixed fields in these records

that were not included in MARC 21. Also, we observed that the 049 field for these records would have the holding code of the cataloging agency IKMN, rather than our own TXIM holding code, which we anticipated would cause no problems.

The pros and cons of having holdings for ebooks were also discussed, and expediency required loading the bibliographic records without including a holding record for each title to contain call number and barcode information. The absence of holdings records causes the text "The library currently has no holdings for this title" to display in our Web catalog; Reference staff were concerned this could confuse patrons. Our system allows for customization of this message, but to date it has not caused problems. We have agreed that we will revisit the item record and holdings issue in the future, as our ejournals have holdings records, but our ebooks do not.

Our serials cataloging unit had been cataloging ejournals since 1998 and already had established cataloging parameters. We reviewed these procedures and determined that their policies did not relate as closely to ebook cataloging needs as we had thought. Their electronic resource cataloging procedures did, however, reinforce the decisions we had already made:

- Add text after the call number to alert patrons that the title is an ebook
- Modify the text "Bibliographic record display" in the 856 $3 field to "Online version."

Finally, we determined parameters for the bibliographic load process:

- Splitting the large file of records into sets of more manageable size so that pre-editing could be done more easily
- Running the keyword index program on the split files of records rather than all at once, so that keyword indexing proceeded more quickly
- Trying to use the load program to identify duplicate print titles for their NetLibrary counterparts (unfortunately, our bibliographic load program lacked this capability)
- Tracking the database control numbers for each file in case we needed to isolate these records later for global updating.

TOOLS AND STRATEGIES

MarcEdit

MarcEdit (*http://oregonstate.edu/~reeset/marcedit/html/*) is a free MARC-editing utility developed by Terry Reese, Assistant Librarian at Oregon State University's Map and Aerial Photography Library. It includes a tool that "breaks" MARC records into an easily readable, tagged text file, and another which restores broken records to MARC format. It also includes a powerful editor which provides the ability to find and replace text; edit fields, subfields, and indicators; and count the fields and subfields in a file of records.

The system librarian downloaded MarcEdit and experimented with a sample of the NetLibrary records to familiarize herself with the program's capabilities and limitations. She then met with cataloging staff that had been reviewing the records and identifying problems. It was clear that MarcEdit would be effective in fixing several of the problems.

First, MarcBreaker was used to convert the records to display as tagged text (see Figure 3).

With the records broken, various MarcEdit editing tools were used to fix several of the problems which had previously been identified (see Figure 4).

Field Count

MarcEdit includes a tool to count the fields and subfields in a file. This proved useful in determining various problems that affected access and accurate description, and certain descriptive cataloging practices that we do not use, including

- call numbers lacking $b
- 050 call numbers with multiple $a's
- records without call numbers
- 245 fields with $n and $p that need review and revision of duplicate call numbers, along with other problems
- 300 fields with $e's indicating accompanying material, usually CDs or software
- 440s with initial articles, which our DRA system does not link to the authorized series heading
- 653 and 655, which we do not use in our online catalog
- 6xx fields with indicators other than 0 or 1, as we use only LCSH

FIGURE 3. Pre-Cleanup MARC Record Converted to Tagged Text
(.mrk Format)

Example:

```
=LDR  01467cam  2200349La 45{92}0
=001  ocm42328389\
=003  OCoLC
=005  20010825074918.0
=006  m\\\\\\\\u\\\\\\\\
=007  cr\cn-
=008  970916s1998\\\\nyu\\\\\s\\\\\001\0\eng\d
=010  \\$z   97040153
=040  \\$aN{dollar}T$cN{dollar}T$dOCL
=020  \\$a0585000255 (electronic bk.)
=020  \\$z0814403875
=050  14$aHD30.2$b.P477 1998eb
=082  04$a658.4/038$221
=049  \\$aIKMN
=100  1\$aPeterson, Brad L.
=245  14$aThe smart way to buy information technology$h[computer file] :$bhow to
maximize value and avoid costly pitfalls /$cBrad L. Peterson, Diane M. Carco.
=260  \\$aNew York :$bAMACOM,$cc1998.
=300  \\$axiii, 258 p. ;$c26 cm.
=500  \\$aIncludes index.
=533  \\$aElectronic reproduction.$bBoulder, Colo.
:$cNetLibrary,$d1999.$nAvailable
via the World Wide Web.$nAvailable in multiple electronic file formats.$nAccess
may be limited to NetLibrary affiliated libraries.
=650  \0$aInformation technology$xManagement.
=650  \0$aInformation technology$xEquipment and supplies$xPurchasing.
=655  \7$aElectronic books.$2local
=700  1\$aCarco, Diane M.
=710  2\$aNetLibrary, Inc.
=776  1\$cOriginal$w(DLC)   97040153$w(OCoLC)37695172
=856  4\$3Bibliographic record
display$uhttp://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=1910
$
An electronic book accessible through the World Wide Web; click for information
=994  \\$a92$bIKM
```

- 7xx fields with $e relator terms, which we do not use and which conflict with our authority records.

The field count report was also very useful because it provided a sure method to review the contents of the MARC record tags and subfield

FIGURE 4. MarcEdit Tools and Fixing Bibliographic Record Problems

Problem	ToolUsed
245$h - change [computer file] to [electronic resource]	MarcEdit/Edit Fields
Add 690 E-book	MarcEdit/Edit Fields
Delete 653 and 655 fields	MarcEdit/Edit Fields
856$3 - change "Bibliographic record display" to "Online version"	MarcEdit/Edit subfields
Add proxy server prefix to URL	MarcEdit/Edit subfields
Change 776's to 876's	MarcEdit/Replace

FIGURE 5. MarcEdit Field Count Report Examples

Label	No. in Fields	No. in Subfields	Meaning and Use
LDR	100		Every record has one, and only one, leader
040	100		No problem. This count of 100 040's can be used as a sure count of the number of records in this batch as each record should have one 040
040$a		100	No problem – there are as many 040 $a's as there are 040 fields
050	97		There are 100 records but only 97 050 fields. There are 3 records without an 050 which need call numbers
050$a		103	There 3 too many occurrences of 050 $a which we need to identify and correct
050$b		96	There are perhaps 4 records without a $b which we need to identify and correct
099	1		We do not use 099s, so we would need to identify and correct the record
105	1		This is not a valid MARC field. We need to identify and correct the problem record
245$a	100		Every record has one, and only one, 245
245$n		37	Records with $n are often single issues of a serial, and may have duplicate call numbers, along with other problems. We need to review these
300	65		May not be a problem, as born-digital resources do not have 300 fields
300$e		7	Ebooks do not have accompanying materials. Identify and delete $e's

codes for any other unidentified problems. For example, we found that we could use the overall number of certain required fields, such as the 049, to learn the exact number of records in the batch. We could then compare this number to other fields that should have the same number, such as 050/090, to determine if we had records lacking a call number. Figure 5 shows selections from a MarcEdit Field Count report and its usefulness in identifying problems in the content of the records. Many of these specific problem records can then be isolated by using the spreadsheet strategies described in the next section.

Microsoft Word Macros

A different solution was sought for a second group of problems because we were unable to fix them using MarcEdit (solutions may be

available in new versions of the software). Because the records were converted to display as text using MarcBreaker, Microsoft Word macros were developed to handle this group of problems (see Figure 6).

Microsoft Word Find and Replace

The next task presented a challenge that sent the system librarian into Word's online help. Our local cataloging policy dictates removal of 6XX fields with a second indicator of anything but 0 or 1, excluding 690. Word's find/replace function, with the wildcard option, provided the solution (see Figure 7):

Find what: =6[!9][0-9] ?[!01]*=
Replace with: =

Spreadsheet Strategies

Importing MARC records into a spreadsheet enabled grouping fields to examine their contents for error identification. We initially arrived at the idea of using a spreadsheet to identify a record lacking a known field, the 050, and then later found it useful for pinpointing several other problems. The steps are outlined below:

1. Load the file of tagged records into an Excel spreadsheet.
2. Use Excel functions to number the lines and the records.
3. Sort the records to bring like tag numbers together.
4. Then:
   (a) visually inspect for missing or incorrect data;
   (b) visually inspect for missing sequential record numbers; or
   (c) select a group of records and perform a search for text within the selected group.

This was an invaluable cleanup method, unavailable in our system's traditional database maintenance programs. There were many other areas for which we were able to use the spreadsheet. Depending on the type of error we found, we had two cleanup options:

1. Immediately edit the text-file copy of the records with the identified corrections. These were types of problems that were relatively straightforward, and required little or no cataloging judgment (see Figure 8).

FIGURE 6. Microsoft Word Macros and Fixing Bibliographic Record Problems

Problem	What the Macro Does	Macro

Change encoding level	For each record:
	
	1. Find the record leader
	2. Delete the 20th character in that field and replace it with "3"
		```Sub NetLibrary_change_encoding_level()```
		```Dim Task As Long, NumberofTasks As Long```
		```NumberofTasks = InputBox("Perform edit how many times?",```
		```"Number of Repeats", "1")```
		```For Task = 1 To NumberofTasks```
		``` Selection.Find.ClearFormatting```
		``` With Selection.Find```
		``` .Text = "=LDR"```
		``` .Replacement.Text = ""```
		``` .Forward = True```
		``` .Wrap = wdFindContinue```
		``` .Format = False```
		``` .MatchCase = False```
		``` .MatchWholeWord = False```
		``` .MatchWildcards = False```
		``` .MatchSoundsLike = False```
		``` .MatchAllWordForms = False```
		``` End With```
		``` Selection.Find.Execute```
		``` Selection.MoveRight Unit:=wdCharacter, Count:=20```
		``` Selection.Delete Unit:=wdCharacter, Count:=1```
		``` Selection.TypeText Text:="3"```
		```Next Task```
		```End Sub```

Change record type	For each record:
	
	1. Find the record leader
	2. Delete the 9th character in that field, to replace it with "m"
		```Sub Netlibrary_fix_type()```
		```Dim Task As Long, NumberofTasks As Long```
		```NumberofTasks = InputBox("Perform edit how many times?",```
		```"Number of Repeats", "1")```
		```For Task = 1 To NumberofTasks```
		``` Selection.Find.ClearFormatting```
		``` With Selection.Find```
		``` .Text = "=LDR"```
		``` .Replacement.Text = ""```
		``` .Forward = True```
		``` .Wrap = wdFindContinue```
		``` .Format = False```
		``` .MatchCase = False```
		``` .MatchWholeWord = False```
		``` .MatchWildcards = False```
		``` .MatchSoundsLike = False```
		``` .MatchAllWordForms = False```
		``` End With```
		``` Selection.Find.Execute```
		``` Selection.MoveRight Unit:=wdCharacter, Count:=9```
		``` Selection.Delete Unit:=wdCharacter, Count:=1```
		``` Selection.TypeText Text:="m"```
		```Next task```
		```End Sub```

FIGURE 7. Explanation of Word Find/Replace Using Wildcards

Part of the Expression	Explanation
Find what:	
=6	All tags begin with an equal sign, and the first character of the tag to find is "6"
[!9]	The second tag character can be any digit except "9"
[0-9]	The third tag character can be any digit
2 spaces	In MarcEdit's tagged text display format, there are 2 spaces between the tag and the indicators
?	The first indicator can be any character
[!01]	The second indicator can be any digit except 0 or 1
*=	Any number of characters up to and including the equal sign that signifies the beginning of the next tag
Replace with:	
=	Removes all characters described by the "Find what:" expression, then restores the "=" at the beginning of the next tag
Example: The first two 650 fields below will remain but the third and fourth will be removed because the second indicator is not a zero or one. =650 \0$aWomen in motion pictures. =650 \0$aViolence in motion pictures. =650 \6$aFemmes au cin{acute}ema. =650 \6$aViolence au cin{acute}ema.	

FIGURE 8. Spreadsheet Examples: Problems to be Corrected Before Records are Loaded

Problem	Example
050 call numbers with multiple $a's	=050 14$aZ7204.S44$bB88 2003eb$aBF632 E-Book
Incorrect indicators in 245 fields	=245 10$aThe smart way to buy information technology (second indicator should be 4)
300 $e	=300 \\$a1 v. (various pagings) :$bill. ;$c28 cm. +$e1 computer disk (4 3/4 in.)
440 initial article	=440 \4$aThe series on school reform =440 \4$aThe series on social emotional learning
700 $e	=700 1\$aBeinborn, Birgit,$d(1962-),$eaut. =700 1\$aHerv{acute}as Garc{acute}ia, Juan,$eil.

2. For errors which were more complex or required additional cataloging tools, we extracted subsets of the spreadsheets containing those records. These we saved and printed for correction after the records were loaded into our catalog (see Figure 9).

FIGURE 9. Spreadsheet Examples: Problems to be Corrected After Records are Loaded

Problem	Example	Corrections Needed
Missing subfield b in call number	=050 14$aPR9619.3.M23 E-book	Main entry and shelflisting required
Record lacks call number	=050 14$aRC280.B8$bR48 1997 E-book =245 02$aA review of the Department of Defense's program... =050 14$aTK7874$b.N333 1997 E-book =245 10$aMicroelectromechanical systems... * =245 00$aAdequacy of the Comprehensive Clinical Evaluation... =050 14$aBL2525$b.E426 1997 E-book =245 14$aThe fifties spiritual marketplace... * Records are sorted by tag and all lines are deleted except 050 and 245. Resorting by line number pairs 050s with their respective 245s. Consecutive 245 fields with no intervening 050 indicates a call number is missing from the second 245.	Classification and shelflisting required
Record has multiple call numbers	=050 14$aHN801.A8$bF78 2003eb E-Book =245 10$aPolitical identity and social change... * =050 14$aD126.5$bT73 2003eb E-Book * =050 14$aDS102.95$b.B66 1988 vol. 6$aDS126.5eb E-Book =245 00$aTraditions and transitions in Israel studies... =050 14$aHD58.8$b.C317 2004eb E-Book * Consecutive 050 fields with no intervening 245 indicate that the record containing the following 245 has multiple call numbers	Classification and shelflisting required
245 $n and $p	=245 00$aApplied developmental science.$n[Vol. 6$nNo. 4] =245 00$aApplied developmental science$n[Vol. 6.$nNo. 2] =245 00$aEducational studies.$n[Vol. 33?] =245 10$aWiley CPA examination review.$nVolume 1,$pOutlines and study guides... =245 10$aWiley CPA examination review.$nVolume 2,$pProblems and solutions...	Review and revision of duplicate call numbers, along with correction of other problems

TRANSFORMING CLEANUP TASKS INTO A METHODICAL FRONT-END PROCESS

Library staff had already determined the specific edits that were needed and had the tools to make the changes, namely

- MarcEdit to identify field and subfield anomalies and perform global edits
- Word macros and find/replace editing tools to retrieve problem texts and perform global edits
- Excel spreadsheets to perform data sorts which identified and grouped other anomalies and problems.

With these tools, we had the ability to upgrade all NetLibrary cataloging records to our standards *before* we loaded them into our

bibliographic database. This was a breakthrough in our method of bibliographic record cleanup, which had previously been done *after* the records were in our catalog.

Because these were new bibliographic record cleanup processes that used new editing tools, we must

- Establish file-naming conventions and report parameters
- Set up workflows and specific tasks for staff performing the work
- Create procedures that detail a step-by-step approach to editing and revision tasks.

Cataloging staff from the monographs cataloging unit and the database management services librarian created new workflows and corresponding documentation. As we created a cleanup process and procedure, we tried it out on a copy of our existing set of NetLibrary records, using the new cleanup tools and honing the procedure as necessary until its steps were correct and in the correct order. The result is a methodical front-end record cleanup process that is efficient, robust, and effective.

After the cleanup procedures were completed and tested and revision steps documented, we began the work of implementing our newly established cleanup methodology on the actual NetLibrary records. The system librarian, who introduced us to the new tools, rejoined us at this point to boost confidence and provide insight as we put them to use.

FUTURE CONSIDERATIONS

While we have made every effort to identify and correct as many errors as possible before loading the records, it is likely that we will continue to encounter new and different problems. As we do so, we will look for ways to incorporate new solutions into our pre-load cleanup procedures.

There are also future considerations regarding NetLibrary records for which we are unable to provide definitive solutions; these include the following:

General Issues
- Permanence of collection and future viability of NetLibrary itself; NetLibrary has already been in financial trouble but was saved by OCLC.
- Will TexShare funding be continued for NetLibrary titles?

- In the mix of ebook providers (Project Gutenberg, Million Book Project, Internet Archive, etc.), what will NetLibrary's role be in the future of electronic resource dissemination?
- How will our arrangement with TexShare be affected by Baker & Taylor's partnership with OCLC to provide NetLibrary titles?

Technical Issues
- Ongoing authority issues: Outsourcing vs. in-house cleanup
- Item records for ebook titles: Yes or no?
- Can this cleanup process be used for other outsourcing projects?
- Quality of future NetLibrary records: Certain types of problems appear to have been corrected in recent batches; will this trend continue?
- Will there be a mechanism put in place to allow error-reporting to the agency who catalogs NetLibrary records?
- Monitor MarcEdit for functionality enhancements, and identify other potentially useful software or strategies
- How to handle the relationship between print and ebook manifestations of the same title
- Will this front-end cleanup process of vendor-supplied bibliographic records become a regular database maintenance function?
- Will metadata description of ebooks assume a larger role in the future, perhaps replacing MARC as a communication format?

With this project we have achieved a high degree of quality control over cataloging records from one specific source of electronic resources. However, with the proliferation of ebook sources that use very basic cataloging or none at all, we will face larger issues of how, or if, we can continue to provide consistent, quality cataloging and authority control for these titles. If some entity does not provide cataloging for the universe of ebooks, will other methods such as basic Internet search engines be sufficient to provide access?

CONCLUSION

Cataloging, reference, and the system librarian worked together as key players in the NetLibrary record load process. Our desire to have front-end quality control over the vendor-supplied records required that we look outside of the traditional database maintenance tools available in our integrated online system. Our literature survey revealed no com-

prehensive information available on the process that we envisioned. However, our system librarian found the tools and initiated the process. MarcEdit, Word, and Excel were identified as the software applications that would fill this role. They have given us more flexibility and power in our database maintenance work than we had ever imagined possible. We will continue to use them in the future in order to assure the quality of any other vendor-supplied records *before* we load them into our catalog.

NOTES

1. These procedures are employed in the Alkek Library Cataloging Department of Texas State University-San Marcos, to perform cleanup of NetLibrary records prior to loading them into the database. They encompass a variety of tasks that assure the quality of NetLibrary records and provide a clear and consistent OPAC display. *http://www.library.txstate.edu/cat/netlibrary/procedures/index.htm*

2. OCLC PICA: NetLibrary Ebooks Available. *http://oclcpica.org/?id=1012&ln=uk*

3. DA Information Services–Electronic Media: To Sample Some eBooks, Go to NetLibrary. *http://www.dadirect.com/Emedia/emediatitle1.asp?id=3*

4. FAQs: NetLibrary: MINITEX eBooks Collection: CPERS: Programs and Services: MINITEX. *http://www.minitex.umn.edu/ebook/netlib/faq.asp*

5. Hyatt, Shirley, "netLibrary," *Ariadne*, Oct. 10, 2002. *http://www.ariadne.ac.uk/issue33/netlibrary/*

BIBLIOGRAPHY

MarcEdit

Bigwood, David, "MarcEdit," *Catalogablog,* Jan. 11, 2005. *http://catalogablog.blogspot.com/2005/01/marcedit.html*

Kentucky State University Libraries, *Department Manual: Using MarcEdit to Edit Large Numbers of Bib Records. http://www.lib.ksu.edu/depts/techserv/manual/general/marcedit.html*

MarcEdit Homepage: Your Complete Free MARC Software. *http://oregonstate.edu/~reeset/marcedit/html/*

Palermo, Natalie, *Using the MarcEdit Program, LOUIS Users Conference 2002, Louisiana State University. http://www.nsula.edu/watson_library/acrl/Using%20the%20MarcEdit%20Program.ppt*

Other MARC-Editing Tools

MITINET/marc Library Services: MARC Magician. *http://www.mitinet.com/ Products/p_cleanup.htm*

The next link takes you to search results for the MARC subsection of Perl scripts within CPAN.ORG, Comprehensive Perl Archiving Network. *http://search.cpan.org/search?query=marc&mode=all*