

Author Bio: Mary Finn, MLIS is Head of Expedited Cataloging at Virginia Tech, University Libraries. She can be reached at maryfinn@vt.edu. She presently serves as Vice-Chair of ACLTS Creative Ideas In Technical Services Discussion Group, as well as on the ACLTS Council for Regional Groups. She was previously Chair and Co-Chair of the Virginia Library Association Chapter of ACRL. The author gratefully acknowledges the input of Jana Doyle, and Terry Reese as creator MarcEdit.

Title: Batch-load Authority Control Cleanup Using MarcEdit and LTI

Author Name: Mary Finn

Abstract: Vendor sets of bibliographic records are readily available and must be loaded into our catalogs to provide patron access to the items they use the most. However, batch load authority control can be an arduous process since bibliographic standards vary between sets. Librarians at Virginia Tech have implemented a process that includes authority control clean up before batched bibliographic records are loaded into the catalog. This article describes the process used.

Keywords: batch-loads, LTI, authority control vendors, MarcEdit, vendor supplied records.

The importance of batch-loading vendor supplied records has become obvious in libraries whether academic or public. At Virginia Tech, our most recent statistics show that door counts of students and faculty entering the library, as well as usage of reference services, books, paper journals, and circulation have fallen dramatically, while use of electronic resources has risen dramatically each year.¹ Many of the electronic resources to which patrons need access are sold in batches of bibliographic records which must be loaded into the catalog. Authority control of these resources has been time consuming for catalogers and much of it has been done after the fact, when a cataloger happens to run into a problem. At Virginia Tech, all the clean up is done on the front end so that by the time the batches are loaded into our catalog, authority control is finished. Some examples of batches we have loaded using the process described at the end of this article are Early English Books Online, American Periodicals Series Online, World Bank e-library, Safari Books, NetLibrary, Oxford Reference Online and government documents.

An August 2007 search of both the Web and of the *Library Literature & Information Science* full-text database using the terms LTI, authority control, MarcEdit, and batch-loads found many discussions of authority control and the websites for OCLC, several Technical Service departments web pages, and authority control vendor sites. Only one article was found specifically dealing with batch record loads and authority control. However, at Texas State University-San Marcos, where this procedure was implemented, they were not using an outside authority control vendor. No articles were found which used LTI and MarcEdit for pre-batch clean up of vendor records. This article will discuss only the pre-processing of batches using LTI and MarcEdit, as the authority work can be done using these programs alone, before loading the records into any library catalog as a batch.

The University Libraries at Virginia Tech Technical Services department has been using Library Technologies Incorporated (LTI) as our authority control vendor since September, 1998 when the first copy of our entire database was sent to them for machine processing. This process ensures that our existing database was brought up to current authority control standards. In order to ensure newly added bibliographic records meet current authority control standards, we use LTI's Authority Update Processing (AUP) service which gives libraries notification of new, changed, and deleted authority records which have been distributed to the National Authority File (NAF), vendors, and utilities after initial record loads. We also use Authority Express (AE) which provides ongoing authority control for new bibliographic records added to the Integrated Library System (ILS).² In order to use the second service, we extract records monthly from our local ILS and electronically send them to LTI. LTI processes our records immediately and sends

back an “Unlinked Headings” (UHL) report, which provides us with all headings that are not linked in either the NAF or the LTI supplemental authority file.

BATCHLOADING AUTHORITY CONTROL PROBLEM

When batch-loading first began in 2002, all clean up of batch-loaded headings had to be done after they were loaded into our catalog. There was not much point in batch-loading bibliographic records if the records had to be edited one by one. At that time our ILS was VTLS. In VTLS, headings inaccuracies were reported in a buffer report which ran each night. The headings could not be separated out by cataloger, so the responsibility for cleaning up these inaccurate headings fell to the Authority Control Coordinator. It became obvious that our practices needed refining so that we could keep our database to the standard which was maintained before batch-loading was imperative. The Authority Control Coordinator could be burdened with up to thousands of these headings changes after batches were loaded. We needed a less labor intensive way to clean up our batch-loaded headings.

In February 2004, our Technical Services Department reorganized, creating an Expedited Cataloging Unit. This unit is responsible for all batch-loads, including MARCIVE, all other batched vendor records, Promptcat, and copy cataloging. Expedited Cataloging consists of one librarian and six copy catalogers and from two to four student workers depending on needs. When this unit began production, we decided to clean up the authority problems in batch-loads *before* we loaded the batches into our catalog. As Sanchez, et al. found “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.”³

To do this we used our regular authority control vendor (LTI) and Terry Reeses’s MarcEdit Program to pre-edit the batches. There are many sources of batches now, and both our patrons and reference librarians want these records in the catalog because most offer electronic access to the resources. The quality of these batches varies greatly, and like most records added to the catalog, they must be edited by catalogers. They obviously cannot be looked at and edited one by one, but maintaining our standards as closely as possible for the best possible record retrieval is important. Virginia Tech employs three programs for editing. LTI provides machine editing, and MarcEdit enables us to make changes in the batches that we need to make based on LTI reports. We now use Innovative Interfaces Incorporated as our ILS, and do a last pass on the records, if needed, using III’s Global Update and Headings reports.

BATCH AUTHORITY CONTROL SOLUTION

Using our authority control vendor for batches provides the advantage of focusing work on potential problems rather than the whole batch. Upon downloading a batch file of records from OCLC collection sets, or any vendor, the records batches can be opened in MarcEdit, and FTP’d to LTI with an “.in” extension and then received with an “.out” extension within a few minutes depending on the size of the file. When the records are returned there is an “Unlinked Headings” report which comes with them. The unlinked headings report helps guide us to what needs to be fixed. This process costs us 10 cents a record to do authority matching and have a matched record sent back from LTI. Also, any records that get put in NAF in the future will be sent to us

and added to our authority file when it is added to the NAF. It seems unlikely that this work can be done less expensively than this.

There are also over 60 types of other automatic changes made to the batch of records by LTI when the batches are sent. For example, the indicators in the 245 field are standardized, and initial articles are stripped out of subfield “t” or 130 fields. Punctuation is standardized in all headings fields. In series fields the form of numbers for numbered series are standardized according to the authority record. Obsolete fields are converted – 261 to 260; 301 to 300; 840 to 830 ; subfield ‘b’ is changed to ‘n’ in conference headings; field indicators are updated – 710 20 to 710 2_ ; title subfields are inserted – subfield “l” before “. English”; GMD’s are updated – “computer file” to “electronic resource,” for example. To view a complete list and explanation of standard and optional bibliographic record cleanup, please see our vendor’s website at <http://www.authoritycontrol.com/>.

The LTI process is a machine matching process. Errors that may be obvious to the Cataloger’s eye may prevent LTI from making a match. These can easily be spotted by skimming the Unlinked Headings report. As a quality control tool, the Unlinked Headings report can be used to quickly spot problem headings. Problem headings can lead to problem bib records. Because of this, work can be focused on problems made obvious by this report. Below are some excerpts from a report to demonstrate:

Typographical error:

X30-0-aFood safety and food security. lSpanish.
X00-1-aFrense, Amina.

X00-1-a**Gadrey, J.p0s579539.**

X00-1-aGerstenberg, Frauke Johanna.

X00-1-aGiandrea, Michael D.

(Extraneous data prevented LTI match to NAF. Should be: **Gadrey, J.)**

Below, two suspiciously similar name headings appear on the list. A look at the bibliographic records confirms that these are variants of the same author's name. Name variants within the batch are easily spotted on the ULH report. Although this name was not established in the NAF, we can choose a single form to use in our catalog, or submit an authority record via NACO.

X00-1-aJeffers, Elizabeth Ann.

X00-1-a**Johnson, Kathryn E.**

X00-1-a**Johnson, K. E. q(Kathryn E.)**

X00-1-aJohnston, Christopher Owen.

X00-1-aJolliffe, Dean.

X00-1-aJolly, Vineet Kumar.

(Same person; name not established in the NAF. Determine single form for use in catalog.)

In this excerpt a corporate name is seen in pre-AACR2 form. Again, this name is not in the NAF, but we can identify the error and upgrade the heading to the correct AACR2 form. If the name is established in the NAF at a later date, LTI would identify the match and send the authority record if you subscribe to an authority update service.

X00-1-aWei de Meyer, F. W.

X00-1-aWiegand, Douglas Michael.

X10-2-a**Woman's Hospital Association, Cleveland.**

2 X00-1-aWon, Jongsun.

X10-2-aWuste Filmproduktion.

(Name not established in NAF; AACR2 form would be: **Woman's Hospital Association (Cleveland, Ohio)**)

Below, both of the highlighted headings contain typographical errors that prevented a match to the authority record, Denki Gakkai, or the Institute of Electrical Engineers of Japan.

X00-1-aDawkins, David.

X00-1-aDazzini, Monica Mabel.

X10-2-a**Denkai Gakkai (Tokyo, Japan)**

X10-2-a**Denki Gakkai (1988)**

(Variant form or typographical error. Should be: **Denki Gakkai (1888)**)

Unlinked series headings are listed separately on the report. Here is an example of a variant series title.

Y30-0-aCollectibles (Narberth, Pa.)

2 Y30-0-a**Easy step by step guide.**

Y30-0-a**Easy step by step guides.**

2 Y30-0-a**Easy step-by-step guides.**

Y30-0-aSpon research.

Y30-0-aLibrary of international relations ;vv. 28.

Below one can see that subject headings are also listed separately. This part of the list will include topical subject headings, as well as name and titles used as subjects. In our controlled subject index, we use only LC subject headings, so all 650 headings should be valid. We investigate all that are on the ULH report.

- **650- -aData compression vCongresses.**
(Should be: **Data compression (Computer science)**)
(this is a topic which is lacking a qualifier)
- **630-0-aJava.**
(Should be: **650 Java (Computer program language)**)
(this is a title as topic that should be topical heading)
- **650- -aProfessional xStudy and teaching (Elementary)**
(Should be: **Professions**)
(this is a typographical error in main topic)

MarcEdit Manipulations based on LTI Report:

When the file is returned from LTI, headings can be manipulated based on the LTI report using

MarcEdit. MarcEdit is available free on the web

(<http://oregonstate.edu/~reese/marcedit/html/index.html>), and was developed and is constantly updated and maintained by Terry Reese at Oregon State University. MarcEdit provides a simple method for breaking text into different formats. We break MARC files into MarcEdit's mnemonic text format, which can then be edited using the editor. When we are done editing batches we remake the text files into MARC communications format, so they can be loaded into our catalog.

Our process is the following:

1. Download file from provider
2. Open file in MarcEdit
3. Name file with "YOURINSTITUTION" prefix and ".in" suffix
4. Send to LTI
5. Open LTI ".out" file (input file) in MarcEdit. Type file name into "Input file dialog." Type file name in "Output file." If an output file with the same name currently exists, MarcEdit will attempt to overwrite it.
6. Choose "MarcBreaker" to open the file into eye-readable text.
7. Run MarcEdit field count report by clicking "Reports" → "Field Count" → "Generate Report." Save report as a .txt file, print if needed.
8. Based on MarcEdit field report, remove unwanted fields (such as certain 9xx's, etc.) To do this, go to "fields" → "edit field data" and choose the fields or subfields you need removing throughout the batch, and choose "delete field."
9. To add a field to every record, follow instructions #7, but choose "add field" at the end.

10. There are numerous operations you can do once the file is in eye readable form. You can edit field data, indicator data, subfield data, swap field data, Find and replace information, use the wizard to create scripts.
11. Using information in unlinked headings LTI report, edit headings (names, series, and subjects) that are reported as “unlinked” by looking in NAF for correct headings, and replacing incorrect headings in your MarcEdit file with the correct NAF headings.
12. Check headings to make sure they are in correct AACR2 form. This way, even if record is not in NAF, LTI will match the name if the heading *is* established and will send you the correct updates.
13. When you are done editing, save your edits.
14. If you have completed operations, you want to click on the whirlwind icon to turn the file back into raw MARC.

REFERENCE NOTES:

Eileen Hitchingham, email to author, 31 July 2007

Ashmann, A. 2003. The Lowdown on Automated Vendor Supplied Authority Control .

Technical Services Quarterly 20 (3): 33-44. doi: [10.1300/J124v20n03_03](https://doi.org/10.1300/J124v20n03_03)

Sanchez, Elaine, et al. 2006. Cleanup of NetLibrary Cataloging Records: A Methodical Front-

End Process. *Technical Services Quarterly* 23 (4): 51-71. doi: [10.1300/J124v23n04_04](https://doi.org/10.1300/J124v23n04_04)