2015

Updating Alma Records with the WorldShare MARC Update Service

Barbara Anderson
bjanders@vcu.edu

Follow this and additional works at: http://scholarscompass.vcu.edu/libraries_present

Part of the Collection Development and Management Commons

Downloaded from
http://scholarscompass.vcu.edu/libraries_present/35

This Presentation is brought to you for free and open access by the VCU Libraries at VCU Scholars Compass. It has been accepted for inclusion in VCU Libraries Faculty and Staff Presentations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
When I arrived at VCU in the 1980’s, we were just going live with the old DataPhase ALIS system. Since then we migrated to NOTIS and then to Aleph, and we’ve been using Alma since October 2012. Our local systems have come and gone, but we’ve always relied on OCLC as a source for most of our cataloging records and we have tried to be a responsible contributing OCLC member library, too.

As OCLC users we have migrated from old Beehive terminals to desktop Passport and Connexion applications, and we’ve used a number of OCLC services along the way, including WorldCat collections sets, PromptCat (WorldCat Cataloging Partners), and the Bibliographic Notification Service.
This was a no-fee service that allowed you to use EDX or the OCLC Product Services Web to retrieve records with your institution’s holding symbol that had been upgraded in some limited way (encoding level changes, 505 fields). Along with the daily files of records there were daily reports identifying the records in the files and the nature of their upgrade.

It was a handy way to ensure some degree of currency for the records in your online system, and at VCU we used this service to enhance our Aleph records

We retrieved our bib notification files from the Product Services Web, applied perl scripts to sort the records into categories and loaded them into Aleph with a carefully crafted combination loading and fix routines – trying to make sure that we were adding specific fields (e.g., 520 or 505) that weren’t already in our records without overlaying and removing fields that were of special local interest.

We did this from the fall of 2006 right up until October 2012, when we froze our cataloging activities to begin our migration to Alma.

When we emerged on the other side of Alma implementation, we had higher priorities to address before we could think about how to use Alma’s box of tools to work with Bib Notification records.
Meanwhile, RDA happened.

Of course, it had been happening all around us for a while, but March 31, 2013 was designated “Day One” for Library of Congress’s RDA implementation.

Other national libraries followed suit, and OCLC announced its policy for accommodating RDA records, including plans for effecting batch changes to OCLC records.
Machine Manipulation of Existing WorldCat Records

“OCLC will begin changing existing WorldCat legacy records after 31 March 2013 to incorporate various useful RDA practices. Anticipated changes include, but are not limited to:

• Adding 336, 337, 338 fields
• Spelling out non-transcribed abbreviations in 255, 300, 500, 504 and other fields
• Converting Latin abbreviations to English equivalents in 245, 260, and other fields
• Converting dissertation notes in 502 field to multiple subfields
• Removing GMDs (after 31 March 2016)
• Heading changes in accordance with RDA
• Heading changes are anticipated to spell out abbreviations in headings like "Dept." and to change Bible and Koran headings in accordance with RDA practice.”
Enriching WorldCat with FAST
(Faceted Application of Subject Terminology)

“Beginning in September 2013, OCLC is systematically adding FAST headings to WorldCat bibliographic records. “

<table>
<thead>
<tr>
<th>FAST Facet</th>
<th>MARC 21 Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological</td>
<td>648</td>
</tr>
<tr>
<td>Corporate Names</td>
<td>610</td>
</tr>
<tr>
<td>Events</td>
<td>611</td>
</tr>
<tr>
<td>Form/Genre</td>
<td>655</td>
</tr>
<tr>
<td>Geographic Names</td>
<td>651</td>
</tr>
<tr>
<td>Personal Names</td>
<td>600</td>
</tr>
<tr>
<td>Titles</td>
<td>630</td>
</tr>
<tr>
<td>Topics</td>
<td>650</td>
</tr>
</tbody>
</table>

A few months later, OCLC began another major initiative – to add FAST headings to existing WorldCat records in batch mode.
OCLC RDA Policy Statement

“The policies and planned manipulation of legacy records in WorldCat ... are intended to position WorldCat for an optimum transition to a post-MARC metadata structure in the future. They also attempt to balance the dual roles of WorldCat as a catalog and as a repository of bibliographic data.”

“Legacy records (non-RDA records) need to be made as functional as possible in the future RDA environment to benefit catalogers, systems, and catalog end users. As a post-MARC future becomes a reality in years to come, greater consistency of practice present in existing WorldCat records will allow easier migration to another metadata format.”

OCLC articulated its reasons for their decision to apply global RDA and FAST enhancements to WorldCat records.
At VCU, we began to worry about how we would be able to keep up with these changes. We weren’t sure exactly how we would use the new RDA fields or the FAST headings, but we knew we couldn’t make effective use of them unless they were present in a critical mass of our Alma records.

So, while we were thinking that maybe we needed to re-instate our use of the bib notification service and get in on all of this record enhancement activity, we missed seeing another OCLC announcement.
Bib Notification Service was going away Nov. 1 2013 and was being replaced (and improved) as a part of OCLC’s new WorldShare Metadata Collection Manager services.
At this point we had no idea how we were going to make this work with Alma, but we were optimistic that we could find a way, and we figured we’d better get moving if we were going to take advantage of the global enhancements that were already underway in WorldCat.
Log in to the WorldShare interface

So we set up our WorldShare Collection Manager account and followed online instructions to
Enable MARC Record Delivery

Configure our MARC update record delivery.
We identified fields to delete from delivered MARC records.
We identified fields that would trigger MARC update record delivery if they were added, updated or deleted in the master record.
We set up the frequency and methods for delivering updated MARC records ...
And we did the same for the accompanying reports.

All of this was pretty easy to set up, and in late September 2013 files of updated MARC records started to arrive in our designated ftp site. We stockpiled them in a local network drive until we could figure out how to load them into Alma.

Up to this point, we had experimented a little with Alma normalization rules and we had set up a few import profiles, but we had done very little with merge rules.
Almost immediately we hit some bumps in the road. For one thing, there was the confusion about the significance of the merge rule Name and Description.
We set up a new merge rule called *VCU merge* for our WorldShare MARC Update records, and we set up an import profile to begin testing our first few records, but we couldn’t get past the Merge method part of the import profile, because we couldn’t find “VCU merge” in the drop down list of merge methods. Turns out, with merge rules, it’s not the Name of the merge rule that counts as much as the Description.
Once we got past that little speedbump, we hit a brick wall when we got a little deeper into the Alma merge rule syntax. We wanted to make sure we could keep the fields that were important to us and NOT keep the fields that were no longer relevant. For example, it was particularly tricky, we found, to set up a rule that would add a 264 field from an incoming RDA record and would also remove the 260 field of the previous pre-RDA version of that record. Same with records in which the original / old version of the record had a 1XX field and the new version did not.

We opened a Salesforce case about this in late September, and we went back and forth for months, literally, trying to come up with an acceptable combination of merge rules, normalization rules and import profile settings for our situation. (To be fair, during this same time period we were also engaged in a massive pre-construction weeding project, so we didn’t always give this our full and undivided attention)

The upshot was that from late September of 2013 until late April of 2014 we were accumulating files of updated records with no reliably safe way to load them into Alma. Finally, in April 2014 -- with the help of Ex Libris support (thanks to Liz Best for her patience) and a lot of good suggestions from Alma-L -- we came up with a
workable combination of normalization and merge rules, and we began, very slowly, and cautiously, to load our files – just a few records at a time (looking closely at each one) -- then a few hundred at a time (divvying them up among several catalogers to look closely).

We tweaked our rules and tested again and re-tested before we finally opened the flood gates --- almost exactly one year ago. So, here’s how this is working for us now:
We are still not caught up with accumulated files, so we have a weekly routine in which we retrieve our MARC files from the ftp server and copy them into a shared network drive, sorted chronologically.
For each file that gets posted to the FTP server, we receive an emailed spreadsheet report of the records in the file and which specific updates were made to them.

These spreadsheets get stored in the same shared network drive with the files of MARC records.

By the end of the summer, I hope to be caught up and operating on a current basis only, but while we are playing catch-up, we have a tag team of catalogers working on loading and troubleshooting these update files – in chronological order.
The assigned cataloger locates the next available file of MARC records in the network drive and loads it into Alma using a special Worldshare MARC Update Import Profile, set up to load WorldCat records in binary MARC.

Note that the import profile is currently set up to handle manually uploaded files.
Once we are current with our update files we will change the import profile so that Alma will retrieve and upload these files automatically. We have used this type of automatic retrieval and loading in other contexts and we are confident that it will work.
The import profile uses a named normalization process (MARC21 Bib normalize for Worldshare updates) which is made up of a group of individual normalization rules to modify the MARC records as they are imported into Alma.
The import profile matches incoming records against existing Alma records strictly on the OCLC record number.

Incoming records that match existing IZ records are merged according to a specific merge method - Overlay all fields except VCU local (or VCU merge as it appears in the Metadata Editor).

It ignores CZ records with matching OCLC record numbers.

And it doesn’t try to load records with matching problems.
The merge rule is set up in the metadata editor to remove all MARC fields from the existing record *except as noted* (these are fields that we *might* want to keep, and we’ll address them later in the merge rule)
For example, it removes all 2XX fields except the 246 field.
Then it issues a blanket rule to REPLACE all MARC fields except as noted.

replace MARC.XXX excluding "001,035,246,500,510,541,583,590,655,710,740,773,856,9XX"
Then it gets into more detail about specific fields - replacing them *except as noted* (these are all fields excluded from replacement in the previous statement)

```plaintext
replace MARC."035" when MARC."035"."a" does not contain "NotRecd" excluding MARC."035"("9"," ")
replace MARC."246" when MARC."246"."i" does not contain "Medline" excluding MARC."246"("9"," ")
replace MARC."500" when MARC."500"."5" does not contain "VRC" excluding MARC."500"("9"," ")
replace MARC."583" when MARC."583"."5" does not contain "VRC" excluding MARC."583"("9"," ")
replace MARC."655" excluding MARC."655"(" ","7")
replace MARC."700" when MARC."700"."4" does not contain "drt" excluding MARC."700"("9"," ")
replace MARC."710" when MARC."710"."5" does not contain "VRC" excluding MARC."710"("9"," ")
replace MARC."740" when MARC."740"."a" does not contain "English Short Title Catalog" excluding 
replace MARC."856" excluding MARC."856"("4","1")
```
We have added Medline abbreviations in 246 fields of many of our medical journal records, and we don’t want to lose that work.
And, finally, it adds specified fields (some of which will actually not be added because of normalization rules in place – for example, the merge rule says to add MARC 9XX fields, but our norm rule says to remove field 936 and 938, so they are removed from the record before the merge even takes place)

```
add MARC."505" if does not exists
add MARC."520" if does not exists
add MARC."59"X
add MARC."655"
add MARC."773"
add MARC."9"XX
```
Once the file has been loaded into Alma, we troubleshoot the problems identified in the job report.

No matches

- Maybe our WorldCat holdings settings have added our holdings symbol to a WorldCat record for an electronic title, but in Alma that record is represented in the CZ without any OCLC record number at all.
- A shelfready project from the distant past (we received vendor records without OCLC record numbers and submitted records to OCLC to have holdings set in batch)
- Maybe our holdings symbol was added to a WorldCat record incorrectly at some point in the past
- Maybe we moved our holdings from monograph records to a serial record and forgot to go back and remove our holdings symbol from the WorldCat monograph records

Multiple matches

- OCLC record merges (we had our holdings on each or two different OCLC records and they were merged onto one, or one record used for ordering and another one used for cataloging)
- Actual duplicate OCLC records in Alma (??!)
Earlier I mentioned that for every file of records, we also get a spreadsheet reporting the individual records in the file and the reasons they were included in the file. After loading a file, the cataloger du jour pulls up the spreadsheet and does some sampling to make sure

- The incoming record matched and overlaid the correct record
- The changes indicated in the spreadsheet report are reflected in the Alma record
- There are no weird additions or transformations that will require additional tweaking or that could be deal breakers going forward.

WorldShare MARC Update report for contents of one file:
Here’s a closer view of portions of one of the lines on the spreadsheet (showing the addition of 040 $e rda, changes to the 100 field, 33X fields and 505 field)
And here’s the Alma record before/after
In the 6 years that we were using the bib notification records in Aleph, we upgraded/enhanced almost 160,000 records. Recently we started getting monthly reports of our WorldShare MARC update activity – we received an unusually large number of updated records in files from last August, but since then, we’ve been getting almost as many updated records each month as we processed during our entire 6 years of working with bib notification records.
In the 1 year that we have been loading WorldShare MARC update records, we have modified over 1.4 million records – nearly 10 times as many.
In fact, in the space of a year, WorldShare MARC update records have enhanced more than 3/4 of all of our active institution zone records (the ones with actual OCLC record numbers)

So far so good. The quantity of records we’ve touched in the past year has exceeded my expectations. It’s not an entirely automatic process by any stretch of the imagination (although it will become more automated when we get caught up later this year), and it’s not 100% foolproof, either. In our checking we have found a few problems, but they are few and far between. And we’re convinced that the benefits far outweigh the risks.
Barbara Anderson
804-828-6358
bjanders@vcu.edu