Report on Authority Vendor Assessment

Lloyd Chittenden

Marmot library network is evaluating three different vendors to help us manage out authority records, LTI, MARCive and Backstage. We sent a set of 4929 bibliographic records to each of the three vendors to see how each one would process the records so we can evaluate differences in how each one processed the same set of bibliographic data.

The three vendors have different methods to achieve the same goal of keeping a library's data clean. One primary option is whether the vendor supplies replacement bibliographic records when headings change or depends on the library system's Automatic Authority Control Process [AACP] to change the headings based on the authority records the vendor sends. Both Backstage and MARCive provide the option of using either method. LTI only uses the process of providing new bib records to change your headings. LTI strongly discourages their customers from using their AACP.

DATA PROCESSING

1. Backstage

The response from Backstage was disappointing. There were some serious flaws in the data they returned.

A major issue I noticed was how Backstage processed the term 'score'. It is not uncommon for catalogers to put the word score in a 650 field on a musical score. This is not correct cataloging, but it was common particularly before we had 655 genre fields available to use, so our system has many such records. The problem is that there are four different organizations that have 'score' as part of their name or an acronym for their name, so the word score is in a see reference in the authority records for these organizations. The records returned from Backstage had flipped our 650 score fields in musical scores to four separate 610 fields in each record, one for each of those organizations. When I pointed this problem out to Backstage, they responded that they didn't have a solution to that, but they would work with us to develop a solution.

The other serious error I found in the data returned from Backstage was how they dealt with topical subheadings that could also be genre subheadings, such as Juvenile literature. They flipped all of them so they were coded as genre subheadings even they had been correct topical subheadings for those items. So for example, there was a bibliography of children's literature of Spain. It had this subject heading:

651 $a Spain $x Juvenile literature $v Bibliography

Backstage changed it to:

651 $a Spain $v Juvenile literature $v Bibliography

as if it were itself a children's book rather than about children's books. They did this for all subheadings that could be either topical or genre. When questioned, they pointed out that most of the time these headings are applied incorrectly and these changes usually correct an error. I did a spot check of our

data and found that this was true. Out of 20 examples of the use of a topical subheading Juvenile literature or Juvenile poetry, 16 of them were themselves juvenile literature and the topical subheading was applied incorrectly. However, there were also 4 examples of the subheading being applied correctly. If we change all of these indiscriminately it becomes impossible to correctly use a topical subheading on such an item. It will always get changed to a genre subheading. Again Backstage's response to my complaint about this was that they would work with us to develop a solution.

I was also unhappy with Backstage's responses to technical questions. At one point I realized that I had misunderstood their process. An email from them suggested that we would be using our AACP, and I had thought their process included them sending replacement bib records rather than having our AACP flip our headings. I asked about this, and they suggested I contact Martha Rice Sanders at III about that rather than give me their own recommendation. Martha was helpful, but in the end she didn't have a complete understanding of the options from Backstage. In Rhode Island she had used the form of their service that required the use of AACP, and didn't seem to realize that they also had a process option where they send you updated bib records.

Also, the file of authority records they sent us was not formatted correctly for a III system. III needs the name and subject authorities in separate files. Backstage sent a single file, which I believe is the format used by a Sirsi system. When I pointed out the error, they did very quickly produce the correct files, but they should have sent the correct type of file for our system in the first place. I had told them we were on Sierra.

Because of these failures, I do not recommend Backstage as our authority vendor.

2. MARCive

MARCive also has two different options for how they will manage library data. Their Standard Service involves sending us only authority records when headings change and depending on our AACP to flip the headings in our bib records. Their Comprehensive Service includes sending us new bib records to overlay ours when headings change. The Comprehensive Service is brand new and they are still working out what it will be and how much it will cost.

One problem we would encounter if we used the Standard Service is that Sierra's AACP does not take advantage of authority records for free floating subheadings. When free floating subheadings changed we would have to fix them manually. This could be a significant and unpredictable amount of work.

The second option from MARCive is what they call Comprehensive Service. This provides us with new copies of our bibliographic records when headings change. My analysis found several cases where MARCive fixed errors the other vendors missed in the bibliographic records they sent for the Comprehensive Service. All the cases of errors that MARCive fixed and others missed fell into the same category. They were situations where there were more than one very similar authorized name, and our data had the wrong one. The other vendors did not catch these errors, I assume because those other names were also correct LC headings, they just didn't happen to be the right ones for our items. The MARCive standard service would not flip these either because they are all authorized headings, but the Comprehensive Service would.

For example, we have an item where one of the creators is Walter Becker of the rock band Steely Dan. It had an access point 'Becker, Walter'. There is an authority record for someone named 'Becker,

Walter', it just doesn't happen to be the guy from Steely Dan. The guy from Steely Dan is 'Becker, Walter, $d 1950-'. MARCive flipped our heading to the correct one.

There were several similar situations where they did this. They changed 'Garner, Alan' to 'Garner, Alan, $d 1934-'. They changed 'Kilvert, B. Cory' to 'Kilvert, B. Cory, $c Jr.'. They changed 'Alan, Ray' to 'Alan Ray, $d 1926-'. They changed 'Adams, Graham' to 'Adams Graham, $c Jr.' and 'Emenhiser, JeDon A.' to 'Emenhiser, JeDon A. $q (JeDon Allen)'. All of those were correct.

However, I also found an error MARCive made where they changed 'Bond, Michael' to 'Bond, Michael, $d 1926-'. While the Michael Bond who wrote the book we own was born in 1926, he wasn't the one with a date in his authority record, he was the other one.

Regarding the serious error Backstage created with the word 'score' in authorized access points, MARCive avoided that by their policy of not flipping any headings that are matched by more than one see reference in other authority records. They would provide us with an error report and leave it to us to figure out what to do in those cases.

Regarding Backstage's error of flipping topical subheadings to genre subheadings, MARCive did this as well, but not every time. Their process appears to be more selective about which topical subheadings that could be genre subheadings to flip. They told me that they have a list of these terms that are frequently used incorrectly, and they only flip those rather than changing all that could be changed. While this is preferable to flipping them all indiscriminately like Backstage seems to, they still flipped several of these incorrectly. They were willing to work with us to improve the algorithm that determines which of these to change. All of the examples I could find where they flipped the topical subheading incorrectly also included a subheading of '$v Bibliography'. Based on my complaint they are already working on changing their programming to prevent the subheadings from flipping it if they are accompanied by '$v Bibliography' or '$v Bio-bibliography'.

Another unique aspect of the MARCive service is that they have a mechanism to allow us to delete records from their system. LTI does not have an easy way to remove bib records. When you delete a bib record from your local system, LTI simply continues to send you updated versions of that record. They don't load into the local system because there is nothing to overlay, but they create an error in the record load, which is an annoyance. LTI only removes these records when you do a complete overhaul of your database (for a fee). MARCive is still developing the Comprehensive Service, so they haven't got all the processes nailed down yet, but they are now saying these records can be deleted quarterly at a reasonable price.

Overall, I can recommend MARCive as an acceptable option for our authority vendor.

3. LTI

LTI only has a single process option. They would send us all new bib records to replace our bib records when headings change. They strongly recommend that their customers turn off the AACP in their library systems. This is because they actually go beyond the Library of Congress authorities when they clean up headings for library data. They set authorized forms of name for people who are not in the National Authority File [NAF], and fix the headings in your data so they are uniform even though there is no official authority record. If our AACP system were turned on with LTI data in the system, it would undo many of the improvements they make to our bibliographic records.

They also have other mechanisms to catch errors in your data beyond what you would get from processing just with the authority records. For example, we had a bib record with the name 'Kilburn, Smauel Smith'. The LTI process caught the misspelling in that name and changed it to 'Kilburn, Samuel Smith'. Just flipping records based on the see references in the NAF would not fix that error because there are no misspellings in those see references.

Another example of LTI's data clean up was a record we had where someone had entered the name 'Dudley Moore' in direct order rather than 'Moore, Dudley' in an authorized field. LTI caught this and inserted the correct authorized form of name even though there are no names in direct order in see references in the NAF. The other vendors did not catch this error.

LTI also changed 'Kaplan, Sanford S.' to 'Kaplan, Sanford Sandy' which is the correct authorized heading. The others didn't catch this because there is no cross-reference in the authority record.

LTI also fixes series headings. If there is a series heading in a 490 field with first indicator 0 as if it were an untraced series, but it is actually a heading that could be traced, LTI will change the indicator to 1 and insert an 830 into the bib record.

LTI creates their own authority records for people who are not in the NAF. For example, we had a record for 'Jones, H. Bolton' and LTI changed it to 'Jones, Hugh Bolton, 1848-1927'. This sort of clean up would help us keep our data uniform even for names that are not in the NAF.

LTI's system does have some errors as well. I found they replaced 'James, D. Clayton' with 'James, Dorris Clayton, $d 1931-' in our data set. These are the same person. The correct LC form of name is 'James, D. Clayton'. My guess is that LTI had their own authority record for him before LC created one, like they do for Hugh Bolton Jones above, but when LC created a heading LTI didn't catch the change and left the old heading in their system. When I pointed the error out to them, they changed it in their database.

LTI fixes many errors found in names for events and meetings. They changed 'EUROMECH-MECAMAT 2008 $d (2008 : $c Torino, Italy)' to 'EUROMECH-MECAMAT Conference $n (11th : $d 2008 : $c Turin, Italy)'. I found several examples where authorized meeting names were fixed like that.

In contrast to MARCive and Backstage incorrectly flipping topical subheadings, I found that LTI's fixes for these were all correct. They flipped several incorrect topical subheadings to genre subheadings. For example, they changed 'Teachers $x Diaries.' to 'Teachers $v Diaries.' This item was not about the diaries of teachers, it consisted of the diaries of teachers, so $v was correct. LTI's process led to correct headings, where the others created errors.

Regarding the 'score' error that Backstage produced, LTI has two separate mechanisms to avoid this problem. One, like MARCive, they won't flip a heading if it matches multiple see references in authority records. Also, they won't flip a heading with five or fewer characters because they say these frequently create problems. They keep a list of such short headings that are okay to flip such as State abbreviations and codes, but most short headings in records will not be flipped.

LTI does not have a method to remove deleted records from their database. They expect that customers will periodically completely replace their entire database with them. It is entirely up to the customer how often to do this. My impression is that some libraries do this as often as every year and some wait as long as 10 years.

Overall, I would also recommend LTI as an acceptable option for our authority service. They do cost more and the group will have to decide if they think the improvements they make to the data are worth paying for.

COST

MARCive has the better price, even with the Comprehensive Service, and the extra cost for deleting records. Our rate of adding records to Sierra continued to accelerate last year. If you include all the sets of subscription electronic resources, we added 883 records per day on average. Without the e-resources it was a much lower 229 per day.

This chart is an estimate our cost based on the rate of cataloging since our last LTI gap file in 2015:

2017 Cost Estimate			
LTI		MARCive	
Without e-subscriptions	With e-subscriptions	Without e-subscriptions	With e-subscriptions
Gap up to 2/28	Gap up to 2/28	Gap up to 2/28	Gap up to 2/28
$2,261	$8,669	$3,940	$12,270
Update existing headings		Update existing headings	
$6,000	$6,000	$1,650	$1,650
Add new records excluding e-subscriptions	Add new records including e-subscriptions	Add new records excluding e-subscriptions	Add new records including e-subscriptions
$5,925	$22,716	$3,555	$13,629
Total		Total Standard Service	
$14,186	$37,385	$9,144	$27,550
		Comprehensive Service (including deleting)	
		$15,194	$33,600

2018 Cost Estimate			
LTI		MARCive	
Without subscriptions sets	With subscriptions sets	Without subscriptions sets	With subscriptions sets
Update existing headings		Update existing headings	
$6,000	$6,000	$1,650	$1,650
Add new records excluding subscriptions	Add new records including subscriptions	Add new records excluding subscriptions	Add new records including subscriptions
$8,394	$32,336	$5,036	$19,401
Total		Total Standard Service	
$14,394	$38,336	$6,686	$21,051
		Comprehensive Service (including deleting)	
		$12,736	$27,101

Each of these prices is the extreme end of the spectrum of cost depending on how many electronic items we include or exclude.

I recommend that we include everything in the gap processing, because the price for gap processing is so much cheaper that we really don't save that much money. However, at the full on-going price it makes a bigger difference. An argument can be made that it is not worthwhile for materials we only have subscription access to, and could eliminate or change vendors easily. On the other hand, assuming the authority work improves searchability, it's not unreasonable to want to apply that improvement to all materials in the system.

It appears to me that we have several categories of e-resources to consider:

- Electronic government documents
- Purchased e-resources
- Items on 3-year lease that the vendor calls a 'purchase' (i.e. Kanopy, Bullfrog)
- Subscriptions sets that the library intends to keep permanently
- Subscriptions sets that the library may not keep permanently
- Patron Driven Acquisitions discovery records

The high end prices in my chart include all of those types of records (even if they had a 'No A.C.' field), but it is the big sets like Springer and Naxos that have the greatest effect on the price.

Under our current system any one library will be able to add a subscription set of 100,000 records and have it get authority control. However, that would cost the group $10,000 (or $6,000 at MARCive rate), so it seems like we might want to give some thought to in the future.