

Marmot Union Cataloging Committee
 Wednesday, February 26, 2019
 Minutes

	Jamie Walker [Chair]	CMU		Betsey Dick	Mesa Co
X	Mary Walsh	Adams State		X Mary Paladino	Montrose
X	Steven Weiler	Basalt		Shelly Fratzke	SD#51
X	Robin Trehaeven, Juliana Van Pelt, Carol Harvey, Cindy Young, Mary Miller	Bemis		X Wendi Weinstein, Lydia Wacasey	Pine River
X	Nina Darlington	Bud Werner		X Carol McArdell	Pitkin
	Melissa McCain	Buena Vista		Debbie Barnes	Rampart
X	Oliver Schulz	CCU		Becky Nelson, Kristin Trombo	Salida
	Becky Kramer, Chris Cook	CMC Glenwood		X Sarah Nordholm	Summit Co
	Liz Miller	CMC Leadville		Jo Norris	Vail
	Jonathan Beam	CMC Steamboat		X Nicole Becwar	WSCU
X	Diane Levin, Jolanda Durbin	Eagle Valley		Amy Sieving, Alison Farnham	Wilkinson
X	Selene Gardner	Englewood		Tammy Poquette	Marmot
X	Tyler Dunn	FLC		X Lloyd Chittenden	Marmot
	Amy Shipley	Garfield Co		Ashley Sneesby	Marmot
X	Tallie Gray	Grand Co		Pascal Brammeier	Marmot
	Jill McKinney	Gunnison Co		Brandon Cole	Marmot
	Stacy Contreras	Lake Co		Adam Murray	Marmot

Announcements

- Pika cataloging cooperative meets on Friday at 11 to talk about subject facets in Pika
- Lloyd may be able to create load notes that identify the date of each load or overlay
 - This would require that every record updated or loaded the previous day be exported, changed and reimported overnight. There is a way this could be done, but we don't know all the effects. That would be a lot of record exporting and reloading.
 - Marquis plugin already puts this date in. That has been really useful. It has allowed us to track down problems.
 - It would be valuable to have this information on all record loads.
 - Jamie asks how the dates are added now. Lloyd does a manual global update at the beginning of each month that changes the note in the 995 from "this month" to identify the month.
- Duplicates team report
 - Discussed DVD grouping again
 - 10-minute grace period seems best
 - Based on a sample of 100 cases where we have more than one record for films with the same name, the 10-minute grace period lead to an 86% success rate for grouping. Longer grace period did not improve the success rate.
- WOLFcon report
 - This is the conference about Folio and ReShare.
 - The big news is that China is moving into Folio in a big way. They are talking about implementing it in the city of Shanghai. They sent about 10 people to the conference. Shanghai would be by far the largest implementation of Folio yet. They have over 4 million patrons. They also support colleges. They would have to build it into a very robust system. They would have to create all the features we would need like scoping. They are hiring 2 software development companies to work on it in addition to developers working for the city itself. One problem is that it is very difficult to collaborate on a project like this with people in China. They are not allowed to use Zoom, or any Google products. It is likely they will just take Folio and create their own software fork. Maybe they will make their code public, but even if they do, they may change it so much that the main project can't use it. Another issue is, if we use their code, can we trust that our data is secure on software written by the government of China? There are many uncertainties, but maybe opportunities as well.
 - Folio is not ready for us yet. For one thing, there is currently no system for migrating to Folio from a generic system. Everyone who is migrating onto Folio is building a unique tool to migrate from whatever system they are on. So, Texas A&M is building a tool to migrate from Voyager. University of Chicago is building a tool to migrate from OLE. Chalmers built a tool to migrate from Sierra, but we will need the ability to add new members who could be on any system, so we will have to be able

- to migrate from any system. It turns out that ByWater is working on a generic migration tool because they have the same problem. Their tool will be open source, so that is very promising.
- The system is flexible because sometimes you can build your own tools to do what you need, like ByWater is doing with the migration tool. Marmot is unique. We may need functions nobody else needs. We may be able to create them.
 - There was a meeting of consortia. There are two consortia that are going live soon. FLO and 5 Colleges. Both are academic consortia in Massachusetts. However, we would need functions that neither of them need. For example, the 5 Colleges are really one institution. A student at one of them is really a student at all of them. They don't have to worry about HIPPA, so they don't need or want patron scoping and they are not interested in item scoping either.
 - It was troubling to talk to people who are going live this year. Many of them said they are not getting the functions they think they need to go live. There were people who felt like they were committed to going live before the system is ready for them. They made the commitment on the assumption that there would be functionality that is not ready. This means we should probably wait until the functions we need already exist before we commit to anything, unless we are willing to build them ourselves. Committing to going live does not mean you will get the functions you need created.
 - Jamie asks, can't they create the code they need? That is not clear. They can add more coders to the project, but unless they create their own software fork, or develop something that functions outside the core software, they don't get to pick what their coders work on. There is a product council that decides what all the coders will work on. The product council is made up of Ebsco, Index Data, and the Open Library Foundation. A library can join the Open Library Foundation, and maybe have some influence on one of the votes in the product council, but it would be very hard to really control the direction the software takes. They could develop tools that work outside the core product, like ByWater's migration tool.
 - It is being built in a very modular way. Each small part is a separate module, so maybe we could fork a module to do something different. But it would be difficult to make sure our module continued to work with new versions as the other modules changed over time.
 - It is very messy and complicated, but promising.
 - Next year's WOLFcon is in Germany.
 - Shelly asks if there were many international people there. Yes, there were people from Germany, China and Sweden.

Completed action items

Automate loading of Garfield birth dates	Lloyd	This is finally working. Let Lloyd know if you are interested in the same functionality, or something similar.
Check sample of DVDs for optimal running time grace period for grouping	Lloyd	86% success with 10-minute grace period.
Take up Pika faceting issue with Discovery Team, recruit group to investigate	Lloyd	Meeting on Friday.
Investigate new Pika format facet for "Read-Along eBook"	Pascal	This is not possible because of limited metadata from Overdrive.

Discussion Topics

- What happened to .b59885233? Is that a pattern?
 - Nina wanted us to look at this record. She says her cataloger has been finding these records that look like nice OCLC records in Sierra, but when you look them up in OCLC it is the same title, but it is a very brief record. She wants to know how these robust records in Sierra have minimal records for the same books on the same OCLC number in WorldCat. This would be a problem if our good record gets overlaid by the bad record from OCLC.
 - Jamie asks if there is also a good record in OCLC. Nina says there is and when she finds these, she moves her items to a good record that matches OCLC, but she doesn't want to move other people's items.
 - Nina says they have only noticed these in the past few months.
 - Robin says they have seen this many times ever since they migrated a year ago.

- We decide to start a Google Sheet that people can add to as they find these. Lloyd will start a Google sheet that everyone can add examples too.
- We investigate further and notice that the record in Sierra was edited by someone with the code TnLvLS and that code is not in the OCLC record. We search who that code belongs to and discover that it is Ingram. So it seems likely that Ingram is pulling these records from OCLC before they are complete. Ingram then fills out the records and sells them to customers. This leaves the bad record in OCLC to possibly overlay our good record someday. This business practice does not fit well with our consortium. We will continue to look for examples of this problem to try to get an idea of how big this problem is.
- **FOLIO discussion**
 - Following WOLFcon Lloyd starting thinking about how we could make use of Folio if we decide to use it. The Folio system is probably going to be very flexible and we would have many options for how we set it up. The first decision we would have to make is whether we want to operate as a single tenant or multi-tenant. Single tenant would be like what we are doing now, a single server with everyone sharing records. Multi-tenant would be a set up were everyone had their own server with their own separate database. Each option has advantages and disadvantages. Also there are parts of Folio that don't yet exist that would be required for us to implement each option successfully. Another possibility would be a hybrid system where we split into two or more servers. That could have advantages of both, but would require even more parts of the Folio to be built before it would be possible.
 - One option is a multi-tenant set up. Each library would have its own Folio system. This would have the advantage of not requiring scoping because all the records would be on separate servers. This would require several things:
 - A system for unmediated requesting in other library's catalogs. ReShare may be able to perform this function, but it is not yet ready either.
 - It would also mean we could not share authority work. Each library paying for their own authority service would be a significant cost increase. Also each library would have to be responsible for their own authority work. Current Marmot staffing would not allow us to load authority files for all the different libraries.
 - It would mean the OPAC would have to be able to recognize identical items in different systems and manage requests like Inn-Reach does now.
 - We would want title level requests. Now Folio only does item level requests.
 - Jamie asks if this could be done with Prospector. Inn-Reach currently performs the unmediated requesting functions we would need. Maybe we could use Prospector. Although the Alliance is also interested in Folio and ReShare, so they may go this route as well. We would need to use the Prospector catalog rather than Pika and we like Pika. Jamie suggests that maybe a link could be created to allow requests to go through Prospector from the Pika interface.
 - A second option would be single tenant. This is like what we are doing now. We would continue to share cataloging on a single server. We could continue to use a shared authority service. We would continue to have the same problems we have with record duplication and such. This would require that Folio had scoping functionality that does not exist. We would need patron scoping, item scoping, and separate accounting units. The advantage would be that we might be able to build tools that make things work better, such as automatic deduping.
 - Another option would be multi-tenant with a central cataloging app. This would mean that everyone would have their own Folio server for circulation, but we would share cataloging functions on a central app. This would allow us to continue to share authority service and we would not need Folio to develop scoping because the records we would need scoped would exist on the separate servers. We would still have problems of duplication and bad overlays. The main problem here is that this central cataloging app does not exist. However, that might be the sort of thing Marmot could create as an add-on to the Folio software. Also, we might be able to use the Libris system from Sweden that Lloyd saw at WOLFcon. Libris is a central cataloging app already in use by the entire country of Sweden, and Chalmers University has it working with their Folio system now. Lloyd believes it could be adapted for our purposes. It is open source. However the interface is in Swedish. Also, it does not natively use MARC. It is already operating entirely with BIBFRAME linked data. It can import and export records in MARC, but the Chalmers implementation is using BIBFRAME in Folio as well.

- We could also break up our data on to two or more Folio servers based on some affinity, such as one server for OCLC libraries and another for SkyRiver libraries. Such a structure would make it easier to manage the duplicates problem. We would still need a separate authority service for each server, but that would be cheaper and easier than 30 such services. We could also divide the data in some other way such as splitting academics and publics.
 - Jamie asks what about Z39.50 users. Maybe they could have their own Folio too, but most of the records they use are actually ones they get from other Marmots, not actually from Z39.50. Shelly points out that maybe they could still search the rest of Marmot for records and pull them into their own server without adding Z39.50 records to either of the other servers.
 - Such a set up would require creation of scoping, unmediated requesting, and deduping in the OPAC.
- Another possibility would be multi-tenant Folio servers with 2 or more central cataloging apps divided along the same lines. This structure might be the best of all worlds, but would also require most the different structures that don't exist to be created. Scoping would not be required because records would be kept on separate servers. It would require unmediated requesting, deduping on the OPAC and creation of a central cataloging app.
- Jamie suggests the option of separate Folio servers for publics and academics. There are a variety of options for how we could organize this.
- At this point we don't have enough information to know which of these options would be optimal. We can start exploring these questions when we get a Folio sand box up and get software development looking seriously at the code.
- We don't know yet which way Folio itself will go. We don't know yet if they will create scoping or cross-tenant unmediated requesting. Which way they decide on may determine what we choose.
- Jamie asks when do we think it could be viable for us to us?
 - Lloyd thinks that is a question for the software development team. They would need to get into the code to understand where it is really at and how challenging these various options would be to implement.
- **Illegal Aliens heading update**
 - The report from the CCDA Subject Analysis Committee came out at ALA Midwinter. Lloyd sent the report out to committee members. This report looks at all the different things that various libraries are implementing regarding the Illegal Aliens heading. The most interesting is what Villanova did with VuFind. Without changing any of their MARC records, their public catalog will respond with the same results for a subject search on "Illegal Aliens" or "Undocumented Immigrants." Also it displays "Undocumented Immigrants" as the subject heading despite the fact that the MARC records still say "Illegal Aliens."
 - Since VuFind can do this we think that probably Pika can do it too, but we don't know if Pika may be too different from VuFind at this point. This is a question for Software Development.
 - We test this in the Villanova library catalog.
 - This should be easy to implement in CMU's VuFind.
 - It would also have to deal with all the other variants, Women Illegal Aliens, Child Illegal Aliens, etc.
 - The committee asks Lloyd to bring up the issue with Discovery Committee. << **ACTION ITEM**
 - If it is too complex to implement in Pika, then we do have the option of actually switching the headings in Sierra with fake authority records. This would not have advantage of still being able to search on the old term.
- On chat Mary brings up the question of bad language codes in Sierra.
 - Jamie points out that we discussed this at duplicates team. Pascal determined that he can look for language in the 041, if there's no code there, then it can look in the Sierra fixed field. His problem is really that other discovery partners don't have that fixed field to check. We realized that it is not actually a big problem for Marmot.
 - Mary reminds us to look at list 21 with mismatched language codes. This is a list of bibs with 'eng' in the 008, but not 'eng' in the Sierra fixed field. The Sierra fixed field is what Classic catalog uses for language limiter. The problem is we should distinguish between things that are not coded from things that should be coded 'no language.' If we don't code them correctly, things that should be in the no language group will get in the English group.
 - Jamie suggests we split this list into subgroups that would be easier to figure out. For example, music titles within this list are much more likely to be 'no language.'

- A library who was not concerned about getting the 'no language' code correct, could just global update them.
- Jamie suggests rerunning list 21 now and then as some get fixed.
- Then we realize that if one library does a global update they would also change records for libraries that are more concerned about accuracy.
- Jamie suggests we could create subgroups of ones that are easy, then we can get the list smaller and more manageable. You could break down the list by material types.
- Robin says that they have already fixed the records they were attached to and they found that in most cases the MARC leader had been corrupted, so they reloaded the record from OCLC to overlay the record there. We could do a batch reload of all the OCLC records, at least for those libraries that are on OCLC. Shelly says, go ahead on her records. It could be a problem because some of our 001 fields are not reliable.
- Lloyd will try to create a list of records that are music and should be set to 'no language' and fix those. Lloyd could do searches within the music list to find things that likely have language << ACTION ITEM
- Lloyd will try to create lists of records for OCLC members and overlay them from OCLC. << ACTION ITEM

New Action Items

Action	Responsible parties
Bring up Illegal Aliens heading question with Discovery Committee	Lloyd
Look for ways to find music with no language in list 21 language problem list.	Lloyd
Investigate creating lists of records from list 21 language problems, for OCLC members to overlay in batch, maybe using title check on the load profile	Lloyd

Ongoing Action Items

Action	Responsible parties
Experiment with creating a file for EDS extract without the OCLC prefix.	Lloyd
Create test MARC record and fake authority record to see how Sierra behaves	Lloyd
Figure out how we can control Prospector display with ITYPE	Lloyd
Develop cataloging training materials	Tammy/Lloyd
Develop flow chart for how to use the volume field	Lloyd
Figure out a process for authority control for FLC's discovery records	Lloyd
Investigate a new Tableau utility for finding bad volume field use	Lloyd/Brandon
Develop documentation for Marquis macro	Lloyd/Tammy
Create training for the duplicate checker	Lloyd/Tammy

Next Duplicates Sub-committee meeting: March 11

Next UCC meeting: March 25