Shared Print Infrastructure Discussion
Meeting Notes
Monday, 5 November 2018, 8:45am-3:45pm EST

Attendees

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Job Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaron Krebeck</td>
<td>WRLC</td>
<td>Digital Services Librarian</td>
</tr>
<tr>
<td>Alison Wohlers</td>
<td>California Digital Library/WEST</td>
<td>Shared Print Operations and Collections Analyst</td>
</tr>
<tr>
<td>Amy Wood</td>
<td>CRL</td>
<td></td>
</tr>
<tr>
<td>Cathy King</td>
<td>OCLC</td>
<td>Executive Director, Product Management</td>
</tr>
<tr>
<td>Cheryle Cole-Bennett</td>
<td>ASERL</td>
<td>Project Manager</td>
</tr>
<tr>
<td>Kristin Ewig</td>
<td>OCLC</td>
<td>Senior Product Manager</td>
</tr>
<tr>
<td>Marie Waltz</td>
<td>CRL</td>
<td>Special Projects Manager</td>
</tr>
<tr>
<td>Mary Sauer-Games</td>
<td>OCLC</td>
<td>Vice President, Global Product Management</td>
</tr>
<tr>
<td>Matt Barnes</td>
<td>OCLC</td>
<td>Director, Sustainable Collection Services (SCS)</td>
</tr>
<tr>
<td>Matthew Revitt</td>
<td>University of Maine &amp; EAST</td>
<td>Maine Shared Collection Librarian &amp; EAST Shared Print Consultant</td>
</tr>
<tr>
<td>Mike Garabedian</td>
<td>Statewide California Electronic Library Consortium (SCELC)</td>
<td>Shared Print Program Manager</td>
</tr>
<tr>
<td>Morag Boyd</td>
<td>The Ohio State University Libraries</td>
<td>Acquisitions and Discovery Strategist</td>
</tr>
<tr>
<td>Rebecca Crist</td>
<td>Big Ten Academic Alliance</td>
<td>Project Manager</td>
</tr>
<tr>
<td>Ryan Jacobs</td>
<td>CRL</td>
<td>Systems Development Manager</td>
</tr>
<tr>
<td>Susan Stearns</td>
<td>Boston Library Consortium &amp; EAST</td>
<td>Executive Director (BLC) &amp; Project Director (EAST)</td>
</tr>
</tbody>
</table>

Review of Agenda and Introductions

Introduction
- OCLC and CRL are working together to get shared print commitments into OCLC
- OCLC’s goals are to improve the monograph workflow, identify registration discovery options and develop the API
- OCLC will develop batch registration for shared print serials.
- OCLC Collection Manager will be used by OCLC members to register retentions and load serials holdings data.
- Data will be retrievable from OCLC through an API.
• Additional discovery improvements will include retrieving records through Connexion and Record Manager.
• PAPR will be a tool to support these efforts
• PAPR enhancements include adding new fields, and creating import/export functionality that will synch PAPR and OCLC’s WorldCat.

OCLC is currently in the Product Strategy/Requirements phase, and specifically today at the Requirements Validation/Charleston point.

Comments
• OCLC hopes to finish and release the new product by June 2020, though some of it may come sooner.
• OCLC’s Shared Print Registration Service (SPRS) went live for single-part monographs June 1, 2018.
• SPRS holds over 250,000 retention commitments registered through WorldCat.
• The planned serials registration workflow is based on the existing single-part monograph workflow, with the addition of the 583$3 to record holdings information.
**Agenda item 1: WorldCat Registration Workflow (OCLC)**

Kristin Ewig walked the group through current process of registering shared print monograph commitments in WorldCat. Slides and conversation notes included below.

**Points of note from Kristin/OCLC:**

- Functionality for batch loading serials commitments is building on existing functionality for batch loading monograph commitments.
- Profiles using OCLC’s Collection Manager must be established before batch loading holdings.
- Profiles will include elements common to all records in the batch: program, library, holdings location, action date, commitment expiration date. Today, action is limited to *committed to retain*. ILL policies are also documented.

**Points of note from discussion:**

- Concern about inability to register gaps and condition problems through batch process if only one action, *committed to retain*, is allowed.
- Concerns about creating exceptions to profile rules: different date of action or commitment expiration.
- Need more than LHR format for downloaded holdings. Innovative systems do not store holdings data in LHRs.
- Concern that dates of action and dates of expiration are in the profile. There is a need to customize these per batch load or to update already contributed records.
- Confusion about *Merge field information or Do not update* when adding commitments to records with existing LHRs. Need examples from OCLC.
- Configuration of ILL rules is inadequate for shared print. Need more nuanced rules.
- Location information needs to allow for multiple locations.
- Need to make condition information actionable—used for discovery or generating reports.
- Concern that Shared Print Collection managers do not have access to WorldShare Collection or Record manager because they are not associated with a single OCLC member.

**Follow up**

Progress on these will be communicated via email, virtual updates and/or additional posts on CRL’s webpage.

- OCLC will investigate expanding the batch loading spreadsheet to allow for more optional information that could override profile information. (Ex. Dates or location information).
- OCLC will provide examples to make merging, updating, or not updating more understandable. Note: Please see OCLC webpage that calls out the specific merging rules for each field
- OCLC will investigate creating an account or username/password for Collection Managers that do not have access to WorldShare.
Part 1: Establish a collection/create profile

Step 1 – Create the Collection

To access the registration process, in WorldShare select *Data Sync Collection* for *Collection Type*, and *Local Holdings Record* for *Data Sync Type*.

Commentary

- There’s a need to be able to edit retention commitment data. Item by item or in batch mode?
- OCLC doesn’t include validation in the existing workflow. But will make this possible.
Step 2 – Complete Properties

On this screen you enter the Collection Name. Indicate whether or not you are a WMS library. Select Non-MARC for the Original Data Format. Select Yes for Shared Print.

Comments

- Campus librarians will input this information.
- Campus librarians will be able to download fully formed LHR's through this process. This will not benefit Innovative users, who cannot use LHR's.
Step 3 – Complete LHR Information

On this screen users enter the Initial Estimate for Record Count. This tells OCLC what the file size required.

![LHR Information Screen]

Slide 12
Create LHR Information

Step 4 – Complete Retention Commitment.

There are separate tabs for the next five steps. The Retention Commitment tab is required. Here you specify Treatment of Existing LHR’s (Merge field information or Do not update), and provide information for the 583 subfields. Multiple values are allowed for the 583$f (Archiving Program Name). However, this process does not support different commitment types. Different values for the 583$c (Action Date) or 583$d (Retention Commitment Expiration Date) will require separate collection profiles.
Comments

- You can register under multiple program names
- For serials, fixed dates are set when you select them.
- In the Optional Information control, you can add additional 583 subfields, including $j–Site of Action, $u–URL to Documentation for Archiving Program, $z–Public Note, and $5–Archiving Institution.
- We’d like to be able to pull up all the holdings with foxing or other condition problems a program would want to replace. Currently, you can create a query collection to help you download what you need.
- The 583$u with a link to the terms of a program is in the Optional Information dropdown.
Step 5 – Complete ILL Policies

This screen lets users specify whether items can be loaned or reproduced, or not.

Slide 14
Complete ILL Policies

Comments

- There are limits on lending in most shared print programs.
- We might need a separate policy for each ILL
- We might need to distinguish between service and preservation copies at the item level.
- Having a mechanism to determine if something is at risk is also important for ILL policy

Step 6 – Complete Location

The information on this screen populates based on the OCLC symbol and Holdings Library Code of the current location, in the 852$a and 852$b.

Slide 15
Complete Location
• A library’s holdings can be split between facilities or different locations in the same library.
• The OCLC system only allows one location per profile.
• It would be easier for OCLC if programs could make the location item-level rather than profile-level.
• In some cases separate locations have different retention commitments.
• ALMA winds up overwriting some shared print information when using OCLC.
• When there are two different locations PAPR requires two separate LHR’s
• The 583$3 can be utilized to disclose a different location. Anyone not using a separate OCLC symbol for shared print should be able to use this technique, but no one actually does this.

Step 7 – Complete Linkage Information

In the Treatment of Existing LHR options, you can specify Merge field information or Do not update.

<table>
<thead>
<tr>
<th>Retention Commitment</th>
<th>ILL Policies</th>
<th>Location</th>
<th>Linkage Information</th>
<th>Ownership History</th>
</tr>
</thead>
<tbody>
<tr>
<td>The MARC Organization Code is for the institution whose control number is contained in 014 $a. (For example: OCoLC.)</td>
<td>If you do not enter your library’s MARC Organization Code, then OCLC will automatically set this subfield value to your OCLC symbol.</td>
<td>Treatment of Existing LHRs</td>
<td>MARC Organization Code (014 $b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Merge field information</td>
<td>Find your code at the Library of Congress Marc Code List for Organizations [1].</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do not update</td>
<td>0au</td>
<td></td>
</tr>
</tbody>
</table>

Slide 16
Complete Linkage Information

Comments
• The information input here helps OCLC with internal matching.
• The general consensus was this option is not well understood, and the wording is confusing.
• OCLC will look into this some more and provide some examples for the community.
• A suggested use case was Colby College. Colby has 583’s for multiple commitments. They would not want to overwrite some of these.
**Step 8 – (Optional) Ownership History**

This step includes an *Ownership and Custodial History* note.

![Ownership History Table]

*Include Ownership and Custodial History Note*  
- Yes  
- No

*Treatment of Existing LHRs*  
- Merge field information  
- Do not update

*OCLC Symbol of Previous Owner or Location (581 $a)*  
- OCWMS

*MARC Organization Code of Previous Owner or Location (581 $s)*  
- Find your code at the Library of Congress [MARC Code List for Organizations](#).

**Comments**

- This process is meant for registering a new batch upload. It is not for querying collections. Querying is part of discovery and is at the record level, not the batch.
- This information needs to be filled out every time you submit a batch, the last profile is available and you can reuse that information and add or change it as needed.
- A record of ownership is useful for tracking ownership.
- A program might need this information if a library withdraws their content from a program.

**Step 9 - MARC Record Output Information**

Enable MARC record delivery.

![MARC Record Output Information]

*Enable MARC Record Delivery*  
- Yes  
- No

*File Delivery*  
- Deliver a file of local holdings records interleaved with my bibliographic records

**Slide 18**

*MARC Record Output Information*
### Step 10 - Complete Contact Information & Submit Collection

Libraries can add multiple contacts. This is the final step before submitting the collection.

![Contact Information](image)

**Comments**
- A discussion about recording gaps and conditions in 583's.
  - When materials have condition notes they are at the title level.
  - Can OCLC add condition information as part of the batch registration process?

### Part 2: Submit shared print retention commitments

In part 2 of the monograph registration workflow librarians submit their commitments. The data file is output as CSV. The headers are renamed. The file is renamed to allow a successful upload. Future columns can support serials information. For instance, 583 subfield 3 will ensure that holdings data gets loaded. Barcodes are optional, for WorldShare Management Services (WMS) libraries only.

#### 1. CSV File

<table>
<thead>
<tr>
<th>WorldCat OCLC Number</th>
<th>Bib Record Number</th>
<th>Barcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1107603</td>
<td>1188423e</td>
<td>3934024905925</td>
</tr>
<tr>
<td>20298765</td>
<td>19695152</td>
<td>3934016183922</td>
</tr>
<tr>
<td>34887420</td>
<td>1981790x</td>
<td>3934016435132</td>
</tr>
<tr>
<td>2480318</td>
<td>25976045</td>
<td>393406232127</td>
</tr>
<tr>
<td>5219401</td>
<td>10878488</td>
<td>393401653127</td>
</tr>
<tr>
<td>11978271</td>
<td>19520967</td>
<td>mq1211696</td>
</tr>
<tr>
<td>62180478</td>
<td>1019404X</td>
<td>393402290273</td>
</tr>
<tr>
<td>8432569</td>
<td>28988233</td>
<td>393402491296</td>
</tr>
<tr>
<td>36438790</td>
<td>b14237702</td>
<td>3934015435474</td>
</tr>
<tr>
<td>2046008</td>
<td>b2151438</td>
<td>393407280232</td>
</tr>
<tr>
<td>4090162</td>
<td>b14936901</td>
<td>3934006913148</td>
</tr>
<tr>
<td>28222328</td>
<td>b36047041</td>
<td>393402912779</td>
</tr>
<tr>
<td>6322007</td>
<td>b3129908</td>
<td>3934024195496</td>
</tr>
</tbody>
</table>

#### 2. Rename Headings

<table>
<thead>
<tr>
<th>OCLC Number</th>
<th>LSN</th>
<th>Barcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>3976667</td>
<td>3128692x</td>
<td>393402493292</td>
</tr>
<tr>
<td>20249576</td>
<td>19695352</td>
<td>393401643952</td>
</tr>
<tr>
<td>34887420</td>
<td>1981790x</td>
<td>3934016435132</td>
</tr>
<tr>
<td>2480318</td>
<td>25976045</td>
<td>393406232127</td>
</tr>
<tr>
<td>5219401</td>
<td>10878488</td>
<td>393401653127</td>
</tr>
<tr>
<td>11978271</td>
<td>19520967</td>
<td>mq1211696</td>
</tr>
<tr>
<td>62180478</td>
<td>1019404X</td>
<td>393402290273</td>
</tr>
<tr>
<td>8432569</td>
<td>28988233</td>
<td>393402491296</td>
</tr>
<tr>
<td>36438790</td>
<td>b14237702</td>
<td>3934015435474</td>
</tr>
<tr>
<td>2046008</td>
<td>b2151438</td>
<td>393407280232</td>
</tr>
<tr>
<td>4090162</td>
<td>b14936901</td>
<td>3934006913148</td>
</tr>
<tr>
<td>28222328</td>
<td>b36047041</td>
<td>393402912779</td>
</tr>
<tr>
<td>6322007</td>
<td>b3129908</td>
<td>3934024195496</td>
</tr>
</tbody>
</table>

#### 3. Rename File

<collectionID> <OCLCsymbol> sharedprint <optionaltext> _<YYYYMMDD> <optionalfile#>.csv

Example:
- 1028071.ndlv.shareprint_20170731.csv

**Future:** Potential columns added here to support serials information
Comments

- Local record number discussion
  - There should be another record number column, in addition to the bib record number.
  - If there is no barcode you could use that field as an item-level identifier.
  - Innovative systems don't have LHRs, which makes this less useful. The number needs to be meaningful to the tracker. It would not always be a static number.
- Other columns could also be added.
  - Location might go into this part of the process, rather than in the profile.
  - When it comes to the flow of data, this afternoon's discussion may impact this spreadsheet.
- Can we incorporate item level validation too?
- This could be a solution for authority control too.
- We could add Program terms too. The 583 standard allows approximately 20 terms, but they are not used consistently. Making this data actionable would be helpful. The current subfield l/subfield z format is awkward. These details are worth figuring out.

On this screen librarians will upload their data file, under My Files.

Slide 23
Upload Data File

Comments

- Two types of batch load reports will be available, a Summary Report and an Exception Detail Report. There may also be an option to download MARC records.
- When it comes to submitting commitments, are there opportunities to enhance the existing process, as well as the proposed support for serials?
- Can we add multiple 583's to an LHR, in a single batch? For example, is there a need for a separate 583 to specify "date audited on."
- Useful additional columns that match 583 information are: location, commitment date, record number, condition reviewed, preservation copy vs. a circulating copy (or just multiple copies) and last audit date. This would allow adjustment at the item level, rather than only allowing this during batch processing.
- OCLC would like to include these columns for monographs, too. Would there ever be a use case where serials and monographs were combined? For example, if batches are by subject area?
• Having separate workflows for monographs and serials takes more time. Our members want to send us everything, not just monographs or serials.
• Retention models for monographs and serials can be very different. Action notes can be different for the two.

Registration Discovery & API Development

Here there was a general discussion of registration commitment visibility, metadata services, discovery, the API and if shared print commitments should go into WorldCat.

Comments

• OCLC probably won’t include shared print commitments in WorldCat.org. This is because WorldCAT.org is public-facing. Also librarians use First Search for discovery more often than WorldCat. Is there an important use case for WorldCat.org?
• Shared print symbols are not visible in WorldCat.
• Not all Shared print program staff have a WMS subscription.
• Having records publically visibility helps to show members the value of their participation in shared print programs.
• First Search is easier to use because WorldShare groupings are collapsed, but they’re not in WorldCat.
• OCLC will look into giving consortia groups WorldShare subscriptions so they can access FirstSearch products and put records into WorldCat.org.

This is the Record Manager Display in WorldShare.

Slide 29
Record Manager Display

Comments
• Our member libraries search at the individual title level. They mainly use the information for resource sharing.
• Our member libraries need to know about shared print commitments. They use the information for reallocation of space and preservation. Access is secondary right now.
• WEST asked members where they look for retention commitments, and the answers were AGUA, PAPR, and WorldCat Discovery.
• Retention information can be found with a query collection download from OCLC that’s compared to PAPR data.
• WorldShare is pretty powerful. Can WorldShare be tied into this new product? This may be out of scope for this discussion.
• We want to know what we have uniquely retained.
• Also our members use JRNL, they need more training on other tools.
• It would be great if we could get gap information. Although, this is difficult with title changes and publication gaps.
• All of this variability will make it harder for the API. Any uniformity this product can introduce will really help machine learning. Can we use machine learning to read holdings and detect gaps?
• Aggregating serial holdings is difficult. This is why JRNL gaps are entered manually.
• Publisher’s data isn’t any better. They normalize their data differently and have imperfect knowledge of what was published.

Here's another view showing 583 detail. The LHR flag determines whether you can drill deeper to an LHR.

![Record Manager LHR Display](slide31.png)
Here’s a Connexion display. The SP indicator is shown in yellow.

![Connexion Display 1](slide32)

Here’s another view showing 583 detail. It’s the Connexion display and raw fields for a test record.

![Connexion Display 2](slide33)

Comments

- This display doesn’t have a 583 subfield 3. I would also want to see the 86x fields.
- OCLC will add the 583 subfield 3 to the Connexion and Record Manager Displays.
- These tools are used by cataloging staff. OCLC can convene a group of catalogers and collection development staff.
- ASERL has a working group made up of these people that could participate.
Here's a mockup of a FirstSearch results page.

![FirstSearch Mockup](image1.png)

**Comments**

- This is not used by a lot of folks at the moment. It tends to be used by collection managers, researchers, and library staff. It’s not a user-friendly view so it’s not for most patrons. For precision searches you can get title-specific results, or results based on other facets. We could add a filter so you could specify the number of commitments to retain.
- Click on the Libraries Worldwide link to see, then you can see the Shared Print Commitments to Retain.
- OCLC will investigate including a second symbol
- You can set your Library as a priority, also the Location (e.g., US, MA) to see those holders first.
- In the Group Holdings box, you can put in an ILL abbreviation, like EAST, to let you go beyond one preferred library.

If you click on the Local Holdings link you’ll get a detailed local holdings record. The Shared Print Committed to Retain section is not currently exposed, but OCLC plans to change this, in order to make the screen more user-friendly.

![Detailed Record](image2.png)

**Slide 38**
*FirstSearch Mockup (continued)*
Comments

- Can users see the summary holdings in FirstSearch?
- Can we add an OCLC symbol to a title without holdings? Then users could see who was committed to retain the title, even if the individual holdings weren’t ready yet. It would be useful for Collection Managers/
- Use cases in FirstSearch
  - ILL tracking
  - What programs already retaining a title
  - Who’s weeding a title
  - Commitments beyond your program, group or state.
  - Subject or collection level commitments.
  - A report with a scope broader than my program would be useful. We could be more confident in the copies retained than relying only on WorldCat holdings in general.
  - We would want to match our holdings against those of other programs at a macro level. What we get from PAPR is great, but it’s not always a direct OCLC match, and it’s going by titles, not by copies.
  - GreenGlass is a better tool for this but not for serials.
  - A deduped report in PAPR would be helpful. If the data is in sync with WorldCat people can check what others programs hold.

API Discussion

The goal of the API is to provide bulk access to data in OCLC. This will facilitate decision making across a wide number of libraries. The developers know they won’t be able to anticipate every use case, so it will be useful for them to know the types of decision making that might be needed before they start to build the API. For example, a user could harness cross-library decision-making to decide, “Hey, I have two copies, I need them as widely diversified geographic locations.” This also addresses the “one source of truth” problem. It would be helpful for OCLC and GreenGlass as well.

Comments

- The developers want to know if there can be two separate APIs: one for what’s held vs. one for what is not.
- For the Big 10, I would want a batch list of materials. WorldCat tells us what isn’t held, PAPR doesn’t provide “wants” data.
- I would want all the data to be in one place.
- Regarding the format of API results, OCLC is thinking MARC XML or Dublin Core. Is that a question for your teams, or can you answer?
  - JSON.
- Not all libraries have developers with API expertise.
- Can the programs offer their API developer services to smaller libraries and programs?
- The API will be available to everyone with an OCLC cataloging/and or holdings subscription, or the skill level to use the API.
- For folks who don’t have one of these options, OCLC wants PAPR to be in sync with WorldCat.
- In the past we’ve had misleading results with OCLC’s API scripts. We need the expertise to work that out.
CRL discussion of PAPR updates

This part of the meeting discussed the PAPR registry. Marie Waltz reported on prior advice on shared print profiles from a CRL advisory committee. The major recommendation was to make data from the profiles available as data points in the PAPR directory to compare across programs. For example, funding and support. What is important about funding: cost, membership levels, number of people in the project? Discussion focused on what data points were most important to include in the PAPR directory.

Comments

- For me will depend if the information is for recruiting members, reliability or sustainability of the program?
- As a program I might want to structure the information differently.
- It is difficult to identify a list of funding categories. So many things go into a program, different tiers, membership fees, etc. Also, they are reevaluated over time. I would rather see a link to the full documentation.
- A link to a program’s full documentation would be helpful for recruiting.
- Some programs are closed, and so there is no need for recruitment.
- Libraries use local information when deciding whether to join a program.
- With monographs there are different levels of commitment. EAST has a matrix that shows whether monograph and serials programs are self-funded or grant-funded, and their storage models. We will supply this to you so you can understand what we do.
- Sometimes stakeholders aren’t important to know about. For example, eleven of our schools have state network resource sharing agreements, but these commitments are not considered important to our program unless they are holders of the last copy in the state. The last copy holders will be more important in the future.
- It would be useful for us to have a record of program decision points over time, how a collection model has changed. Perhaps a timeline
- Review periods are important to record. For us they’re built in and there’s one in 2020 – we’ll need to know what to do. For example will we need to change the LHRs?
- In our program we also have a review period coming up in May of next year. We may want to weed or reduce our net commitments for monographs like textbooks. We aren’t sure how things would evolve.
- In JRNL we have a facilities template partially based on PAPR’s. It’s a spreadsheet with broad options and a description that includes terms like ‘remote storage’, ‘Harvard-style stacks’, and ‘closed stacks’. We will share this with you.
- In the future, facility type will be more important.
- Technologies are constantly changing in some areas, like lighting and fire suppression. On the other hand, date of last fire inspection is available from most institutions.
- We’d want to know if the program is local or distributed. That information is easy to get. Also how often the material circulates.
- If a program can’t supply a particular type of data, it’s hard to compare it with others. Temperatures and humidity levels are almost always available.
- If you’re interested in what collection managers look for, our working group has a storage [data] section we can share with you.
- Would programs want to link this information at the title level?
- One of our members has twelve different library facilities.
• I would like to see links to disaster-preparedness plans or preservation plans if they are publicly available. Or just information as to whether they exist.
• Risks and security levels, vary a lot from building to building.
• It would be useful to know if people have de-accessioning plans.
• We weed monographs. For out-of-date textbooks, for example, we don’t need to check with other libraries. Most shouldn’t have been retained to begin with.
• If we could make PAPR filterable by location it would answer a lot of questions. But as useful for conditions at a location that change over time.
• I like the idea of assigning a risk level, but it would need to be applied fairly across programs. Would it make a difference if people knew the three copies were at risky locations? Would they require four or five locations in that case? I don’t think people will bother to check whether some locations are, say, earthquake-prone.
• Length of run and conditions are important. But the information is really difficult to get. Especially if you don’t have a master list of all the issues that were published.

Q: Collaborations are a big issue for updating PAPR. Linking programs up to titles and to other programs and institutions is a big task. It will take a lot of development work. Do you want to see how every title links up with all of the institutions that hold it?
• We have Phases. Some schools only committed for certain phases. Some programs are simpler. Should we be looking at these subsections?
• Maybe life cycle is the thing to track.
• There may be use cases that involve institutions holding for others, and maybe they’ll want to charge them back.
• In the UC model we have projects that behave differently, especially over time. For example, the IEEE is now inactive and has been absorbed into WEST through UC, so it no longer needs to be separated out. JSTOR has different behaviors from WEST.
• The big question is how can this information be supplied to CRL for PAPR? Currently there is no place for it in the 583. WEST uses the 583 subfield f for this, but this is not a perfect fit. Should we change the data standard, or obtain this data externally?
• EAST has project phases where new partners join at different points. This has no bearing on retention rules, it was just something we were concentrating on at certain times.
• In EAST we retain based on the number of holders. For example 4-6 holders would be considered medium rare. Titles can move up and down in risk categories. “Phase” could be a misleading term in our case.
• Gold, Silver, and Bronze communicate multiple categories that can change over time, plus they communicate a baseline validation level. Users can see that when they download from PAPR.
• So is it fair to say that over time these projects/phases even out and get absorbed into the larger programs? For example, right now COPPUL is archiving different kinds of materials in separate phases but maybe it doesn't matter over time.
• Our program is the Big 10, the project is shared print, and there are two distinct phases. Phase 1 was Indiana University; phase 2 is the University of Illinois.
• We do want to be able to distinguish material type (print vs. electronic). Maybe that can build in a calculation of risk on its own.
Break

Originally, the next item on the agenda was for attendees to discuss their needs. The group decided to skip to the data sharing section.

Data sharing among programs, CRL, and OCLC

We assume the data flow will always be one of these three:

- PAPR-to-WorldCat to other library/other group
- WorldCat-to-PAPR-to-library/other group
- Other group/library to both CRL and OCLC.

The order is important and has implications for the development. For example, PAPR doesn’t allow serial holdings on a monograph record. So if you submit those monograph holdings to WorldCat first then the record will be rejected when it reaches PAPR. How can we ensure our databases are all in synch when we don’t all have common rules?

What will be the common data elements? There are some data elements we don’t collect in JRNL. However, they are needed for PAPR and OCLC. What will we need to ask libraries to provide?

Comments

- People will need to know the direction of the workflow.
- The big data needs are action statement, validation level, date of action. We’ll also need to know the date format. If CRL puts the dates into program records, we would want a step where someone from the program checks that we put them in accurately.
- Our members want to use PAPR because not all of them are in WorldCat, so we’re less concerned about being in synch with OCLC.
- OCLC is investigating a workflow that would allow CRL to batch-register commitments from a number of non-member libraries at one time, if they were cleaned up and normalized. Then they could go directly into WorldCat, registered on the libraries’ behalf by CRL. Would this make sense? Would local libraries go for this? It’s a question of control vs. efficiency.
- That would not be a problem for EAST libraries. Cleanup is not a problem; it is seen as introducing good safeguards.
- A separate submission to OCLC would be a burden. No one has said they would have a problem with data going through CRL.
- Illinois just has some questions; they’re still thinking about it. Indiana is very comfortable with CRL doing it; they’ve said “They do good work”, which is rare praise.
- Scholars Trust members will talk about whether they’ll put data into WorldCat.
- CRL would need to add more staff. First-time submissions have the most errors. With subsequent loads users understand the rules better. Sometimes we have to say the holdings don’t make sense outside of the context of their own catalog.
- PAPR serial records have been normalized, that’s a separate workflow. It seems like a burden.
- Maybe the data goes to OCLC first and is just forwarded to CRL by OCLC? But then who is the contact person when something is wrong with the record?
- Also, what about needing to make a one-off update? Would that be done through OCLC or CRL? And how would the corrections actually get made? Possibly through CRL?
- It would be a major change for OCLC to adopt the same requirements as PAPR. Error checking for the upload of monographic data into WorldCat is very much high level. It’s would be hard to
apply that to the holdings level, or to automate it. We don’t have to worry with single-piece monographs.

- CRL could make it work if data goes into WorldCat first. We would just need to outline the workflow and who would be notified if there are problems. Either way works as long as the communication points are defined.
- Data going into OCLC first would be problematic for WEST.
- We want the program to register commitments for individual institutions into OCLC, but Rebecca and Cheryle want the individual institutions to be able to do this on their own.
- OCLC calls it the agent registration project. We would want to add a group ID and an agent ID to registration form upload workbook. We want to set up permission levels. Anyone would be able to upload as long as an agreement is in place.
- In that data flow how does data get back into local systems?
- I don’t know enough about it yet. Will we give LHR’s back to the libraries, but add in individual ILL- or ILS-specific information?
- How do we make sure that important data in local records doesn’t get overwritten?
- Whoever submits data to PAPR has to give us the right data. I’m not sure if the current batch load process can take that information.
- Right now OCLC batch loads allow only commitment to retain.
- Keeping PAPR and OCLC in sync is a major part of this grant. How can we overlay only this shared print data and not irrelevant local data?
- There may be data going through one system that’s not in the other. How do we ensure one source of truth? We don’t know how yet, but OCLC is committed to doing this.
- Does PAPR allow monographic sets? If not, how can we get those into OCLC if they can’t go through PAPR?
- OCLC is expanding monographic registration. It’s still only OCLC, but development will allow CRL into the registration process.
- Does EAST first use the OCLC interface, then your data goes to PAPR, and then you get validation reports? No, I’m not visualizing the OCLC start – so far we have always started with CRL. But that will change. There might be a need for different workflows for individual libraries.
- This dataflow would be a lot more work for CRL. We’d have to keep track of what’s new and what’s different. Sometimes there’s cleanup that’s needed to make sure holdings are really the same. We can tell what’s already been sent to OCLC but what if the data was changed in OCLC and we weren’t notified? We wouldn’t know to pull the new data.
- We also have to keep track of how JRNL and AGUA are updated.
- We’re updating holdings for retention commitments. There might be a four-month lag. Can it be overcome with a shorter update cycle? And how would this impact the ILL?
- A process for single-part and multi-part monographs starts with OCLC. A separate flow for serials commitments would go through CRL.
- I would need to talk to the member libraries about this. I don’t know how often they update WorldCat. I have to beg them to update PAPR, and I don’t know about updating JRNL. They tell me what they’re going to do.
- The best thing is to get more safeguards into OCLC. Especially making the start and end dates of holdings actionable. This will ensure they’re on the appropriate title and material type records. Short term this will require the most development, but in the long term it will provide the best results. It would be be less mediated than our current process, which is not sustainable unless all of the groups come together and decide who will pay for what. Otherwise there can only be on-the-fly mediation.
There seem to be several options. Can we present them to members?
CRL will write up the pros and cons and ask the programs to present them to their members.
Members will say, yeah, let CRL do it. But we need to say what about this and this situation. They might be willing to fund some activities.

Program representatives discuss their requirements, survey answers or pose issues for discussion

For the first five years CDL maintained PAPR. CRL took it over a couple of years ago. All this time we have never surveyed users about how they use PAPR. We wanted to do that as part of this grant. Some survey questions are functional, regarding how people use PAPR now. Bigger questions are should we add a material type, or title families? Currently in PAPR electronic and print holdings get mixed together. Those are really broad overall concerns. What else haven’t we covered?

- This group has touched on most of my concerns and requirements. It would be good to have one workflow. I’ve been engaged with our lending statistics, which are hard to keep track of with shared print symbols. I was surprised by how few of our members actually use PAPR. I want to make that change.
- It would be helpful if CRL could provide their requirements for validation. A self-check interface that returns a report would be useful.
- Data discussions are good, but we need to look at workflow, at how users will actually use OCLC and CRL data. We must focus on bulk processing. If programs are doing things right, it will be efficient at scale, requiring fewer individual corrections. Are the data and processes more available in OCLC vs. CRL’s PAPR, and which are current users more familiar with? It makes sense for the program to gather statements for PAPR, but if OCLC makes batch processing available we will need to talk to users.
- Many of our members are large libraries and have their own systems, but BTAA is small. At the end of the year, Indiana will be doing ingesting content except for gap-filling. Illinois will soon only be doing gap-filling. In answering the survey I got hung up on the questions about title families because I wasn’t sure about Illinois. I also got hung up on completeness questions. BTAA caps serial runs at 2010. Complete for us is different than for other programs. [Of the three systems we use]:
  o JRNL is most agile
  o PAPR is publicly available, so it’s useful for inter-program comparisons
  o WorldCat has the records and everyone knows how to use it

Methods for improving data creation and sharing

CRL wants to make PAPR more relevant by not replicating what has been done elsewhere. It’s not easy to get program information from WorldCat. As we know from the microfilm preservation projects, only the metadata lives on after people leave and organizations change or cease. We want to assure members that the data will persist into the future.

Commentary
• There wasn’t much discussion of analytics on the survey, such as geographical distribution of retention commitments, or number of copies.
• CRL does provide a little-known service: give us a list of OCLC numbers and we’ll tell you if anyone has retained those titles.
• Many in this group were involved in a “GreenGlass for journals” exploration. It quickly got too involved and too expensive. After this project is completed we do plan to revisit that. Cleaner data will really help. Going back to the API solution, we think that will provide one analytics approach.
• CRL asked if it’s too early to ask these questions. I won’t make accession decisions based on whether a particular institution holds a title, but I do want to know about its programs. In my experience, having gone through a long Digital Public Library of America (DPLA) project, it was clear to me they should have asked some questions they didn’t ask. There are seven million digital objects with four million different rights statements! This could have been resolved if different questions had been asked. So now is the time to ask program-related types of questions in the shared print project.
• We’d like to know which institution’s policies a program thinks are best.
• We don’t use the PAPR database – we’re early on – but our participants have asked similar questions. We just got our second cohort. Cal State is joining soon. There are still lots of question marks, and we’re only doing multi-part monographs. Our hope is that as different kinds of institutions join us, we’ll gain more expertise and we’ll know more.
• We participate in the Big Ten and we’re also building a new repository to become the last copy holder. Our in-state nature makes us different – we’re going to take more formats. Also, these materials are being added to our catalog. Looking at the survey, the family question interested me because e-journal archives with different publishers are a similar problem. We’re also interested in capturing item level data that will help us in the future. It almost seems like we’ll need to apply machine learning. We want to put the data in once and then have it propagate to where it needs to be. There is a need to move a bundle of metadata together, not separately.
• A real pain point is reprints, because the cataloging community is so removed. I want to expose that a bit more. When CRL recognizes a title change and JSTOR doesn’t, often it has to do with reprints. If they digitize a reprint, often the title page is different, resulting in different records. We’re also early on − but our participants have asked similar questions.
• Is there room in the bib-level metadata to indicate whether there is a digital equivalent for a title?
• And if there is a digital equivalent, does it require subscription access? If so, is that really a digital equivalent? Also, Google digital records are not really equivalents. And this data will need to change over time.
• The 583 might not always be the best place to record something. It was originally for preservation treatment. We don’t want to store data in a field for which it was never intended.
• So there will be a follow-up for institutions to discuss the CRL vs. OCLC workflows? What else should we ask them about?
• One question CRL had was whether to bring title families together in PAPR. That doesn’t seem as necessary after this discussion. Is it worth it if only a few people would use it?
• Currently the initial PAPR search results are sorted by title. There is no sort mechanism by institution. We may be able to add that, and also sorting by format (print vs. electronic).
• Could you add a sortable validation level field to the first page of results, as well as program and facility type (closed, open stacks, remote facility)?
• It was agreed that it’s really more about being able to filter the results, rather than how they are sorted.
• CRL will get more input on the pain points that people experience when navigating the PAPR interface. This could help to identify low-hanging fruit that will improve the interface. Perhaps a survey or WebEx. When identifying people who will be willing to participate, keep in mind that there are different use cases at the program vs. the institutional level.

Methods for improving data creation and sharing

• It would really help if people can provide the beginning and ending volumes and years they hold.
• I’d love to see publication history in PAPR.
• Like OCLC’s title history visualization tool.
• We would also need to know how the holders interpret ‘9999’ in an ending publication year. Likewise, there would need to be a common understanding of ‘unknown’ digits, such as ‘uuuu’. At the piece level it is extraordinarily difficult to do this, even for a small dataset. Electronic vendors normalize the data differently. Even different JSTOR archives store their data differently.
• CRL tried crowdsourcing with a common subset of BTAA data, but couldn’t get cooperation.
• Libraries create records and never go back to change them.
• With ReCAP data, the best we could do was indicate “there may be a gap” if the holdings contained semicolons or commas.
• Publisher websites data isn’t always useful because they use different numbering systems. Also, they have started to clamp down on sharing that information.
• PAPR accepts gov docs if they are cataloged as serials, but not if they’re called out as such.

Wrap up and next steps

• OCLC’s next steps will be to hold discussions with catalogers and collection managers regarding discovery, workflow, and future and current requirements of the API. Specific topics will include:
  o Pros and cons of a separate CRL workflow
  o “Do not update” vs. “merge” options
  o Risk levels of the various workflows and options
  o Spreadsheet columns that will be needed
  o Barriers caused by second OCLC symbol legacy data
  o Group batch registration
• CRL will put together a focus group to discuss the PAPR data elements. We’ll share the data elements in advance and ask the participants which ones they need.
• OCLC/CRL will coordinate a virtual update at the end of January to follow up with the group on project status, and decisions and improvements made based on feedback from this workshop and additional conversations.