

Outsourcing Cataloging at the University of Maryland, College Park: Problems and Opportunities

by Benjamin Bradley (Discovery Librarian, University of Maryland Libraries) <bbradle1@umd.edu>

and Beth Guay (Continuing Resources Librarian, University of Maryland Libraries) <baguay@umd.edu>

Background

For the **University of Maryland Libraries** (the Libraries), a major outsourcing initiative began in late 2011 following an earlier implementation of WorldCat Local as a discovery tool. The Libraries transitioned from MARC record set loads of e-resources collections into the local catalog to the creation and activation of e-resources collections in the WorldCat knowledge base (WCKB) using WorldShare Collection Manager. WCKB collections automate holdings maintenance on WorldCat catalog records and provide direct linking from those catalog records to the Libraries' e-resources in the local discovery layer.

Cataloging of individual ebooks purchased on approval plans or firm orders transitioned from review and enhancement of vendor MARC records loaded to the local catalog to title activations in appropriate knowledge base collections. Catalogers of monographic e-resources not sourced from vendor records but received on standing orders or subscriptions, licensed in perpetuity and not available in WCKB collections, discontinued exporting catalog records to the local catalog, and instead began creating and updating local WCKB collections for access provision. With the cessation of use of SFX by staff in the then Technical Services Division of the Libraries in December 2015 and the completion of the data migration to WorldCat Discovery (WCD) and the WorldCat link resolver in July 2015, electronic resources cataloging transitioned from a traditional to a highly automated and outsourced environment.¹ From yet another perspective, with 77% of the Libraries FY2013 expenditures among print and electronic books and journals going to electronic resources, WorldCat Discovery had effectively replaced the local catalog.^{2,3}

In June 2013, seven Metadata Services Department staff members' duties included e-resources cataloging, among other assignments. Following staff retirements, departures, reassignments, and a March 2017 reorganization, the number of catalogers working on electronic resources, among other assignments, had fallen from seven to three.⁴ Thus, cataloging of some monographic e-resources received on standing orders or subscriptions, licensed in perpetuity, and not available in WCKB collections was suspended.

The reorganization resulted in the formation of four new units within Collection Services (formerly the Technical Services Division): Acquisitions and Data Services, Continuing Resources and Database Management (CRDM), Discovery and Metadata Services, and Original and Special Collections Cataloging. This change brought relief in the form of a new Discovery Librarian position. The Continuing Resources Librarian, one of the three remaining e-resources catalogers, and the Discovery Librarian began collaborating to distill the benefits of the outsourcing experience.

Outsourcing Problems and Opportunities

Problem: Transportation research record

SFX had been customized to push local catalog record URLs to the link resolver services menu for catalog records lacking an ISBN or an ISBN in the first ISBN field of catalog records that matched those in the SFX knowledge base. This functionality was of course lost in the migration to the WorldCat link resolver. In these cases, from the local catalog, the WorldCat link resolver returns a canned response, "We were unable to find direct full text links for this item." A good number of e-resources local catalog records produce this response, for example, many among the 124,885 "legacy" e-resources records that had been loaded in MARC record sets.

As intended when the Libraries first implemented WorldCat Local, URLs from local catalog records are accessible in WCD. OCLC created a Groovy Script for making this service possible.⁵ With this fact in mind, the Discovery and Continuing Resources librarians had discussed the

issue of duplicate working links displaying on WorldCat catalog records from local catalog records and from the WCKB, concluding that the inconvenience to the patron would be negligible.

In situations in which domain and/or path names for URLs within e-resource records that had been loaded into the local catalog changed, local catalog URLs would present an inconvenience to WCD users. One such problem did surface, through staff investigation, on two separate occasions, for titles in the series, *Transportation research record, or "TRR."*⁶ In the first instance in 2018, a STEM librarian was considering withdrawing print versions of resources for which the Libraries' had purchased perpetual access to corresponding e-versions. The librarian consulted with CRDM's Library Services Supervisor, who determined the extent of the Libraries' perpetual access rights; the Library Services Supervisor then consulted with the Continuing Resources Librarian, because e-version records for titles in the series were in the local catalog. The staff then discovered the URL changes. An additional factor for consideration was that the cataloging of these e-version resources had been suspended due to the previously discussed staffing issues. The Continuing Resources Librarian and Library Services Supervisor consulted with the Discovery Librarian, who created a local WCKB collection derived from the OCLC numbers of the local catalog records for the resources, and globally updated the collection URLs. In addition, the Discovery Librarian identified local catalog records for URL field deletion by the Consortial Library Applications Support (CLAS) unit (staff responsible for the ILS database) to prevent the incorrect links from appearing in WCD.

To create the local collection, the Discovery Librarian worked with CLAS to receive the MARC records for the series and then used MarcEdit to transform the MARC records into a KBART file. The MARC 2 Kbart Converter MarcEdit plugin pulls the necessary data from MARC records (such as the URL, title, and standard number) and creates a tab-separated values file using that data. The Continuing Resources Librarian alerted the Discovery Librarian of the previous two cataloging policies for e-resources: the earlier "single record approach" under which e-resources holdings were added to print version records, followed by a "separate record approach" to print and e-version resources. WCKB records need an OCLC number to place holdings on the corresponding record and for linking. Because of the two different cataloging policies, the Discovery Librarian needed to find which OCLC numbers in the local MARC records referred to a print master record and which referred to an electronic master record.⁷ Using the OCLC numbers from the local records, the Discovery Librarian ran a Z39.50 batch search of the WorldCat database using MarcEdit. Once MarcEdit returned the set of master records, the Discovery Librarian exported the 776 (additional physical form entry) fields to check what other formats were listed. If the 776 subfield \$i referenced a print version, the librarian could then deduce that the record was an electronic version and the OCLC number was correct, and if an electronic version was referenced, the Discovery Librarian could replace the OCLC number from the local record with that OCLC number. Finally, the URLs needed to be updated. This was a minor change, as only the domains needed to be changed, which could be accomplished using find and replace, correcting all the URLs at once. Then the 881 records in the KBART file could be uploaded as a local collection to WorldShare Collection Manager, adding the Libraries' holdings to the correct records and linking directly to the resources.

The second incidence occurred about a year later, when the Acquisitions and Data Services Graduate Assistant reviewing the TRR license found that links to the resources via WorldCat Discovery were again failing. She reported the problem via the Libraries' help desk ticketing system, and within minutes, the Library Services Supervisor contacted

continued on page 18

the Continuing Resources Librarian.⁸ Once again, they consulted with the Discovery Librarian, who quickly resolved the URL access problem. In the previous incident, the base URLs needed a small update, but in this case, the series changed platforms so the changes to the URLs were more significant. However, the Discovery Librarian found that the URLs for each issue followed a similar pattern; the domain and path for each title's URL contained the same domain and path followed by a pattern using the title's issue number. Using find and replace in a text editor, the Discovery Librarian used regular expressions to automate the URL corrections. This time, the Continuing Resources Librarian suggested contributing the Libraries' local WCKB collection to OCLC's global collections, since the staffing issue remained, and the cataloging of this particular collection had been suspended. If contributed as a global collection, there would be potential for other catalogers, outside of the Libraries, to add to it.⁹ This time, the Continuing Resources Librarian took steps to remove local catalog records for the e-resources, in consideration of several factors, including differing past e-resources cataloging practices, new practices, and that e-resources in this series were no longer being cataloged. One effect from this action was that call number searching for these e-resources, lacking local catalog records, is no longer an option in WCD.

Problem: Discoverability of PMLA in WorldCat Discovery

While WorldCat Discovery and the associated systems offer efficiencies, they can come at the cost of local control over the metadata presented to users and the user experience in general. One such example was brought to the Discovery Librarian's attention by a Humanities Librarian. The librarian was having trouble finding the Libraries' print holdings for a journal. The Discovery Librarian looked into the problem and found that when the user tried finding the title, *PMLA, the Publications of the Modern Language Association*, an important title for our English literature students, the appropriate records with the Libraries' holdings attached did not display on the first page of results; they were often buried and ended up around the fifth or sixth page of results.¹⁰ This search surfaces article and issue records from Crossref, all titled "PMLA." The results do not provide enough metadata to the patrons to understand what the articles are about (see Figure 1). These records do link to the journal's homepage and display our coverage. Nevertheless, the Libraries' print holdings would require expert searching to find. This is an example of the problem-solving libraries with outsourced catalogs face: library personnel need to work with multiple third parties to trouble-shoot problems.

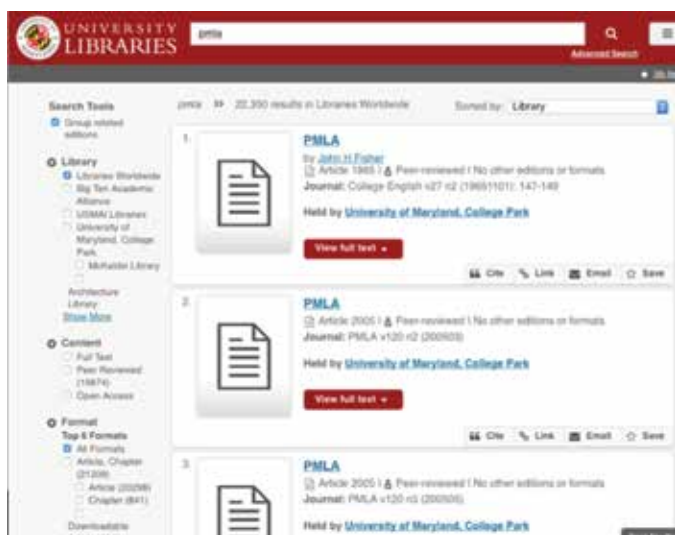


Figure 1

When this issue first arose, the Discovery Librarian worked to find the source of the problem: was this a problem with the system provided by OCLC (WCD), or a problem with the metadata provided by



Figure 2

Crossref? He began by exploring the records in WorldCat Discovery. They were created from the Crossref data, had a minimal amount of metadata, yet included DOIs. The DOIs in the article records redirected to a defunct DOI page on Crossref, but the DOIs in the issue records successfully directed to the correct issues of the journal. He then used Crossref's Metadata Search tool to search for more about these records. The tool provides access to the Crossref metadata in a JSON format, wherein he found that many of the records had "Test accounts" listed as the publisher (see Figure 2).¹¹ Because the records appeared to be test records, the Discovery Librarian contacted Crossref to ask if there was anything that could be done about them. The contact at Crossref explained that these records were a holdover from an older method for managing defunct DOIs, and that there were no current plans to fix those DOIs. The Discovery Librarian posted on the WorldCat Discovery Community Center, a forum hosted by OCLC for librarians using WorldCat Discovery, asking if other librarians had encountered this problem. A few librarians responded about similar situations and shared an enhancement request. Since then, OCLC has shared that they hope to make changes to their algorithm that would help fix these problems. On the other hand, Crossref could clean-up the metadata for these defunct DOIs.

continued on page 19

Opportunity: Automated KBART Feeds¹²

One opportunity libraries can leverage with the automation afforded by the connection of WCD and WorldShare Collection Manager is automated KBART feed services. This service enables publishers to send a library's entitlements to OCLC to automatically activate its holdings in the knowledge base. While the process is automated, initiating the service is not a matter of merely flipping on a switch; it requires manual intervention.

In our case, the KBART feed activates titles in two collections, one for serials titles and one for monograph titles, which cannot have titles already activated in them when the automated feed service begins. The Discovery Librarian worked with the Head, CRDM, to deselect titles in these collections. To prevent loss of access, they worked together to activate these titles elsewhere: The Discovery Librarian created a collection for the eBooks, while the Head, CRDM worked with her staff to ensure that the serials were activated in other collections.

After they were sure that the eBooks and serials were activated in other collections, the Discovery Librarian deselected the eBook and journals collections, and the Head, CRDM retrieved the Libraries' credentials from the provider, **Springer Nature**, and sent the command, via email, to OCLC to start the automated feeds. Subsequently, the eBook collection received 44,489 records from **Springer Nature**, 29 of which were listed as invalid, meaning 29 eBook entitlements were not activated. Upon review of the report for the load, it was found that these titles did not yet exist in the knowledge base collection. The Discovery Librarian reached out to **Springer Nature**, and over time, the missing titles were added.

Overall, automated feeds demonstrate the potential of the Collection Manager and the KBART format. Vendors sending KBART data for your entitlements to manage library e-resources holdings updates is the goal of the KBART automation recommended practice and is something that has been gradually adopted by vendors and system providers. While there is much promise in the automation, the process is not perfect, requiring manual interventions and checks to ensure quality.

Opportunity: ProQuest Dissertations and Theses (PQDT) Global KB project

By leveraging WorldCat data using WCD, libraries are able to provide enhanced discovery and access to materials not previously available in their traditional catalogs. However, the scope of WorldCat is larger than the WCKB. Materials the library has access to may be discoverable in WorldCat because of library-contributed records, but without a corresponding WCKB collection, a library cannot provide access to those materials. At the **University of Maryland**, this situation has resulted in ILL requests for such materials. In particular, the ILL department found that **ProQuest** Dissertations & Theses titles made up 16% of ILL requests received and cancelled because the title was available on the PQDT platform, the largest of any single platform in their study.¹³ So while the discovery layer enables the Libraries' users to find records for resources to which they are entitled, it does not offer access to those resources because of the lack of a knowledge base collection. Furthermore, ILL staff are inconvenienced by having to review and redirect patrons to the PQDT platform. The Discovery Librarian undertook the work of creating a collection for PQDT titles to provide both discovery *and* access in WCD. PQDT is a large collection; as of this writing, **ProQuest** states that the database contains 5 million items, and from some searching in WorldCat, there are an estimated 1 million titles cataloged there.

In order to work on such a large collection, the Discovery Librarian developed a Python script to use the WorldCat Search API to find records for ETDs and to write the data to a file. The script is run from the command line and searches only a single year at a time because the API limits access to the first 100,000 search results, meaning one cannot pull all the records at the same time. Initially the script wrote the results to a file and required manual data cleanup and transformation of the MARC21XML into KBART using MarcEdit. The Discovery Librarian has since refined the script to automate the data cleanup and conversion from MARC21XML to KBART. While searching by year

continued on page 20

IMF PUBLICATIONS

AT NO CHARGE.

FREE ACCESS TO OUR

E-COLLECTION

OF MORE THAN

21,000 TITLES

STARTING JANUARY 1, 2020.

VISIT ELIBRARY.IMF.ORG

BECAUSE KNOWLEDGE
IS A PUBLIC GOOD

INTERNATIONAL MONETARY FUND

has generally operated within the 100,000 search result limit, some years return well over 100,000 results, so logic to initiate additional searches, running through the alphabet to search for items based on the first letter for the title and author, was added. The script starts with the title search, but if too many results are returned it runs the author search as well. In practice this has not been found to be a wholly effective method, but it enables the searches to run.

Once the script has the MARC data, it saves elements needed for the KBART file including title, author, and publication data. When the script reads the 856 (URL) fields, it uses regular expressions to find particular URLs and extract certain elements. Many ProQuest URLs are structured the following way: search.proquest.com/(a unique identifier). Because the URLs are coming from OCLC records, they often have proxies prepended, so the script uses the regular expression, “docview/(\\d+),” to find the unique identifier and rewrite the URL altogether removing proxies or other bad data. The script performs similar cleanup for URLs containing “gateway.proquest.com” or “wwwlib.umi.com.” If a matching URL is not found during this process, the script records the value “not found” for that field. After the URLs are revised, the script writes the data to the text file and then moves on to the next set of search results. The output still requires manual checking, with the assistance of a recently hired Coordinator in Discovery and Metadata Services. ILL personnel now send the Discovery Librarian a monthly report of theses and dissertations that patrons have requested that ILL personnel have canceled and referred; these titles can then be manually added to the collection. After about a year of work, the knowledge base contains 275,000 titles.

Conclusion

The authors have illustrated that within libraries, interdependence across unit and division lines is indisputable in highly automated environments. In the case of *Transportation research record*, within Collection Services, staff in three of the four units played roles in identifying problems and contributing to their resolutions. In this effort, good communication skills have shown to be essential. We have also demonstrated that hand in hand with highly technical skills, institutional memory plays an important role in the process of electronic resources management in libraries.

Working across division lines, e.g., the Collection Services Discovery Librarian’s work with the ILL department, has also been shown to be highly valuable. Beyond collegiality, creativity as shown by the Discovery Librarian’s approach to assisting the ILL department’s PQDT problem is a useful and effective complement to technological “knowhow.”

The methods and tools for creating the PQDT knowledge base collection have supported the creation of many others. In addition to subscription databases, the Discovery Librarian has developed and contributed a number of open access collections (which also includes collections whose titles are in the public domain) to the WorldCat knowledge base. These open access collections include the **University of Nebraska-Lincoln** Zea books, Indiana Authors and Their Books from **Indiana University Libraries**, the Illinois Open Publishing Network, **University of Nebraska-Lincoln** Open Access Journals, **ACRL** Open Access titles, and more. The Discovery Librarian is looking into additional avenues for sharing these KBART files, such



as sharing them on GitHub in addition to sharing them via the WorldShare Collection Manager.

While the automated KBART feed is an incomplete story, this case demonstrated how the use of a discovery product aggregating data from multiple vendors can be complicated for librarians to untangle. As with the case of PMLA, librarians troubleshooting vendor data need to understand where the data comes from and work with stakeholders, including other librarians and vendors, to understand and attempt to resolve problems in library systems. It is also important to understand that when a solution is out of librarians’ hands and dependent on factors such as providers’ development timelines, it is important for providers to communicate about those factors, such as their development schedules, to librarians who must responsibly keep their customers apprised of these situations beyond their control. We feel for our public services librarians, such as our Humanities Librarian, who will undoubtedly encounter the same recurring and new problems when providing services to library patrons seeking discovery and access to the Libraries’ resources. 🐼

Endnotes

1. The **UM Libraries** are members of a seventeen-member library consortium. The Libraries share the ExLibris’ ALEPH ILS with the other consortium members, but have opted out of ExLibris’ SFX services. See “USMAI,” viewed Aug. 29, 2019, <http://www.usmai.org/> and “USMAI (University System of Maryland & Affiliated Institutions); and: Summary of Available and Shared E-Resources, Platforms, and Services, Version 1.0, updated 4/25/2019, viewed August 29, 2019, http://www.usmai.org/sites/public/files/USMAI_Summary_Available_and_Shared_E-Resources_Platforms_Services.pdf.
2. **University of Maryland Libraries**. Annual report, 2013.
3. WorldCat is **OCLC**’s catalog record database.
4. This number excludes copy catalogers activating individual title access in WCKB collections and making minor adjustments, i.e., adding or revising the order of the ISBN fields in the local catalog’s vendor records.
5. See “Apache Groovy,” *Wikipedia*, https://en.wikipedia.org/wiki/Apache_Groovy viewed July 29, 2019.
6. See TRR at the *ISSN Portal*, Print ISSN, 0361-1981 (<https://portal.issn.org/resource/ISSN/0361-1981>); e-ISSN, 2169-4052 (<https://portal.issn.org/resource/ISSN/2169-4052>).
7. The phrase “master record” here and throughout the paper refers to records from **OCLC**’s WorldCat database.
8. For information on the Libraries’ trouble ticketing system, see **Rebecca Kemp Goldfinger** and **Mark Hemhauser**, “Looking for Trouble (Tickets): A Content Analysis of University of Maryland, College Park E-Resource Access Problem Reports,” *Serials Review* 42, no. 2 (2016): 84-87.
9. See “Knowledge base collections,” **OCLC**, https://help.oclc.org/Metadata_Services/WorldShare_Collection_Manager/Choose_your_Collection_Manager_workflow/Knowledge_base_collections, viewed July 30, 2019.
10. See PMLA at the *ISSN Portal*, Print ISSN, 0030-8129 (<https://portal.issn.org/resource/ISSN/0030-8129>); e-ISSN, 1938-1530 (<https://portal.issn.org/resource/ISSN/1938-1530>).
11. JSON example: <https://api.crossref.org/v1/works/10.1632/pmla.2003.118.6.1434d>.
12. See KBART Automation Working Group. *KBART Automation: Automated Retrieval of Customer Electronic Holdings: NISO RP-26-2019*, (Baltimore, MD: National Information Standards Organization, 2019), https://groups.niso.org/apps/group_public/download.php/21896/NISO_RP-26-2019_KBART_Automation.pdf.
13. **Hilary Thompson**, “Find It Fail: What ILL can tell us about Challenges related to Known Item Discovery,” presentation at the UM Libraries’ Library Research & Innovative Practice Forum, June 4, 2015, viewed Aug. 29, 2019, <https://drum.lib.umd.edu/handle/1903/16385>.