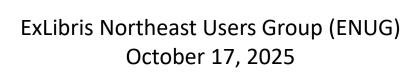# Use of Python script and API calls to batch triage WorldShare Cataloging Partner files

John Myers

Catalog & Metadata Librarian

Schaffer Library, Union College

ExLibris Northeast Users Group (ENUG)
October 17, 2025

UNION
COLLEGE

Key People

- Corinne Chatnik
- Hamza Ghumman
- Lilah Kelly
- Son "Kyrie" Nguyen
- James Gaskell

Slide 2:

The Triage program/script has been an ongoing project of refinement, and I want to acknowledge those who have made it possible:

- Corinne Chatnik, Digital Collections & Preservation Librarian
    - Who has taken the lead in overseeing our students' Python coding efforts
- Hamza Ghumman, Student

    Our original JAVA coder in 2020

- Lilah Kelly, Student

    Our JAVA coder who reworked the program in 2021

- Son "Kyrie" Nguyen, Student
    - Our Python coder who translated the program out of JAVA in 2024
- James Gaskell, Student
    - Our Python coder who made our most recent improvements in 2025

The role of our student coders was crucial, not just for their programming expertise but also in the opportunity to provide experiential learning – a mission commitment for Schaffer Library and Union College.

The Acquisitions/Cataloging Pipeline

An Evolution:
- One-by-one as their individual installment
- In a "box" (that is, a file)
- As a picture

Slide 3:

I've come to conceptualize the evolution of our approach as somewhat of a jig-saw puzzle.

1. Historically, we dealt with the pieces individually, on a one-by-one basis as their own installment – pick up a book, catalog a book, put down a book.

2. About 20 years ago, we transitioned to getting all of the puzzle pieces in a single blank box – the order/discovery records were electronically transferred from the vendor; and the ultimate catalog records were delivered through OCLC's WorldShare Cataloging Partners. But we still had to put the pieces in place after delivery, without the picture to guide us. While the records were collectively present in the catalog, we still considered them on a one-by-one basis.

3. The current script effectively offers the ability to see the entire picture, and effectively assembles most of the puzzle, to highlight the errant and the missing pieces.

WorldShare Cataloging Partners Process

- Library orders book via vendor platform
- Stub bibliographic record with order record delivered to Alma
- Shipment compiled by vendor
- Invoice data sent to OCLC
- OCLC compiles a file of call numbers and delivers to vendor (EBSCO)
- Shipment processed by vendor with call number labels
- Shipment shipped
- OCLC compiles a file of bibliographic records with the POL and other appropriate vendor data and the identified call numbers
- This file is imported into Alma using an import profile that matches on the POL

Slide 4:

Here we have a rough outline of the WCP process.

- We order a book and receive a stub bibliographic record to which the order record is attached
- In parallel, a call number file is generated to support pre-processing and a file of records is compiled for delivery
- The books are shipped and we retrieve the file
- The records in the file overlay the corresponding stub records by matching on the embedded POL numbers

With its implementation, we gained significant efficiency by eliminating the need for catalogers to individually identify, edit, and import records on a title-by-title basis. There remained substantial burdens to assess the quality of delivered records and the presence of call numbers, as we initially still fully processed the books in-hours. To mitigate this task, we further developed documentation so that work-study student assistants could make this preliminary pass. And then there were post-processing quality assurance checks to be performed by full-time staff.

Problem(s) & Solution(s)

?

- Problem 1: Need to assess record quality with a lack of access to shipments during NY Pause

- Problem 2: Efficient and reliable assessment of Call number dynamics during the NY Pause and for a pre-processing contract.

- Solution A: Introduce a "reduced instruction set" to support initial record quality assessment while remote and without access to the physical volume

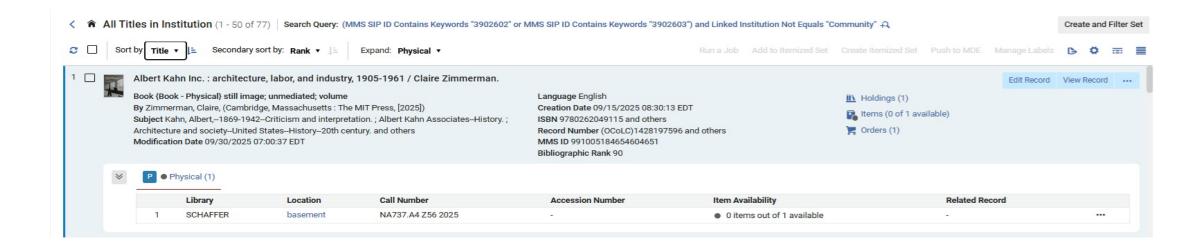- Solution B: Develop the Alpha version of our "Triage Program"

!

Slide 5:

Enter COVID-19 (2020), with two significant impacts. First, we needed to avoid a backlog of items to be cataloged, since we were still ordering but experienced a 4-month inability to access the physical materials due to the NY Pause. This included the identification of low-quality records or mis-delivered records and the need to supply or adjust call numbers due to their absence, or the need to add size prefixes.

We began using a reduced set of quality checks for initial cataloging. This exposed about 10% of records that needed some TLC, and resolved all but 1-5 that persisted with significant issues. These few could later be identified and pulled from the shipments and resolved with the item in hand. As work progressed over the spring, and with the happy conjunction of a student seeking an internship in programming, the instructions for the quality checks were translated into the initial JAVA coding. Over the course of five years, this initial coding has been re-coded as a second JAVA iteration, transitioned to Python, and then had enhancements made to the Python.

The Basic Process

- Obtaining a set of imported records from the import job report and exporting the set content to Excel
- Harvesting the MMS IDs from that Excel sheet
- Returning those MMS IDs to Alma via an API to harvest the entire bibliographic records
- Assessing specific data elements from the records to output to a triage output in Excel
- Act on those assessments in the output Excel

Slide 6:

The process now involves:

- Obtaining a set of imported records from the import job report, and exporting the set content to Excel

- Harvesting the MMS IDs from that Excel sheet

- Returning those MMS IDs to Alma via an API to harvest entire bibliographic records

- Assessing specific data elements from the records, to be output to a subsequent Excel sheet of triage information

- Act on those assessments from the output Excel

Here we have a single record snippet that is part of the Alma set content from an import job. We can see the MMS SIP ID element as the basis for the full logical set.

Assessing & Addressing Holding Records

	A	B	C	D	E
	MMS ID	**OCLC#**	**Floor_Status**	**Size_Status**	**Call_Number_Assessment**
	99100517651370465l	1464589640		No edits: Regular size	HM717 .S38 2025
	991005176513604651	1513516740	Change location to floor 2	No edits: Regular size	BF412 .L49 2025
	991005177012604651	1456362110		No edits: Regular size	N7355.4 .V65 2025
	991005176513104651	1503539065		No edits: Regular size	HF5415.126 .R423 2025
	991005177012304651	1518011513	Change location to floor 2	Change to Oversize q	Manually Review: Duplicate call numbe

Slide 7:

So, what do we check?

First were some checks to drive holding record edits – our main sequence is divided into two locations, with the bulk in the basement, which was templated as the default in our orders; and our main sequence is divided into 4 size segments, with prefixes for quarto, folio, and double folio sizes. The script outputs flags for changing the default basement location to the alternative second floor value based on call number and size. It further flags for the need to insert a size prefix (and by extension the need to relabel). Ancillary to these checks is whether the size is missing from the bibliographic record and whether there are call number issues, including incomplete, multiple, or absent call numbers.

Here we have a snippet highlighting the triage assessments for necessary edits to the holding records.

Assessing & Addressing Bibliographic Records

MMS ID	OCLC#	Call_Number_Assessment	Title	Overall_Condition	Pub_Locn
991005176511204651	1507922209	PS3619.C4848 B87 2025	Bury our bones in the midnight soil / V.E.	Pre-publication Record, Upgrade	
991005176207604651	1512642152	JC337 .G553 2025	A civil society teaching primer : seeing thr	High quality Record, Minimal Review	Missing
991005177009604651	1459000053	HD30.23 .H47 2025	DATA SCIENCE FOR DECISION MAKERS : us	Substandard Record, Overlay	Missing
991005177008904651	1451508837	HD6060 .C35 2025	EARNING POWER.	Substandard Record, Overlay	Missing
991005176511604651	1435631801	DS917.52 .C428 2025	A fractured liberation : Korea under US oc	High quality Record, Minimal Review	
991005176208604651	1473691311	PN1995.9.S5 C43 2025	Screening Sherlock : a cultural history of tl	High quality Record, Minimal Review	
991005177011904651	1449674270	JC573.2.U6 P493 2025	The shadow gospel : how anti-liberal dem	High quality Record, Minimal Review	
991005180755504651	1485358428	Manually Review: No call number	SMART AND FUNCTIONAL COATINGS : nan	Substandard Record, Overlay	Missing
991005179457104651	1438663775	PR4148.I52 H63 2025	WILLIAM BLAKE AND THE SEA MONSTERS (	Substandard Record, Overlay	Missing

Slide 8:

Secondly, there are bibliographic checks – primarily a check on the overall quality of the record. This mirrors an in-house Brief Level rule in Alma that assesses the baseline Encoding Level for the record, source of the record, call number checks, and a check on the presence of subject headings. As mentioned, the checks for holding record edits alert us to issues with the presence of the size and with the call numbers. We also check for a record representing the wrong format (electronic rather than physical), the lack of pagination, and the presence of "xx" coding in the Country element for the record. Most of these checks are redundant with the primary record Encoding Level check but these help to highlight specific problems.

Here we have a snippet highlighting the triage assessments for bibliographic record issues.

Assessing & Addressing Internal Consistencies

MMS ID	OCLC#	Title		Illustration_Status	Bibliography_Status	Index_Status
991005176511204651	1507922209	Bury our bones in the midnight soil / V.E.	-			
991005176207604651	1512642152	A civil society teaching primer : seeing thr	-			Index mismatch
991005177009604651	1459000053	DATA SCIENCE FOR DECISION MAKERS : us	-			
991005177008904651	1451508837	EARNING POWER.	-			
991005176511604651	1435631801	A fractured liberation : Korea under US oc	-			
991005176208604651	1473691311	Screening Sherlock : a cultural history of t	-			
991005177011904651	1449674270	The shadow gospel : how anti-liberal dem	-			
991005180755504651	1485358428	SMART AND FUNCTIONAL COATINGS : nan	-	Illustration mismatch		
991005179457104651	1438663775	WILLIAM BLAKE AND THE SEA MONSTERS (	-			

Slide 9:

Lastly, there are logical checks (T/T, F/F, T/F, F/T) between coding in the 008 control field and the body of the record for recording the presence of illustrations, bibliographical references, and indexes.

These are the latest refinement to the coding. They verify the consistency or lack thereof for these elements. They greatly simplify part of the final quality assurance checks prior to delivery to Access Services.

Completed Triage Efforts

Floor_Status	Size_Status	Call_Number_Assessment	Title	Overall_Condition	Pub_Locn
Change location to floor 2	Change to Oversize q	N7265 .E54 2025	Artists of the Middle East : 1900 to now / S	High quality Record, Minimal Review	
	No edits: Regular size	PN1031 .S385 2025	Breath lines : how poems work and why tl	High quality Record, Minimal Review	
	No edits: Regular size	QD431 .R57 2025	The color of North : the molecular languag	High quality Record, Minimal Review	
	No edits: Regular size	Manually Review: Duplicate call numbe	Like : the button that changed the world /	High quality Record, Minimal Review	
Change location to floor 2	No edits: Regular size	DB71 .B377 2025	Maria Theresa, Empress : the making of th	High quality Record, Minimal Review	
	No edits: Regular size	Manually Review: Duplicate call numbe	More everything forever : AI overlords, sp	High quality Record, Minimal Review	
Change location to floor 2	No edits: Regular size	BF575.A6 S474 2025	The narrowing : a journey through anxiety	High quality Record, Minimal Review	
	No edits: Regular size	Manually Review: No call number	NOBODY MEN : neutrality, loyalties, and f	Substandard Record, Overlay	Missing
	No edits: Regular size	E249.3 .A55 2025	The painter's fire : a forgotten history of tl	High quality Record, Minimal Review	
	Manually Review: Missing size	Manually Review: Incomplete call numl	Uncharted : How Trump Beat Biden, Harris	Substandard Record, Overlay	Missing
Change location to floor 2	No edits: Regular size	DA950.7 .S34 2025	Rot : an imperial history of the Irish famin	High quality Record, Minimal Review	

Red – pull, then route to staff
Yellow – relabel, then Final Review
Green – direct to Final Review
[plain] – direct to Final Review

Slide 10:

As the cataloger works through the triage outputs, if there are anomalies or issues that require the item in hand or require editing in OCLC, those cells are highlighted in Red and the student assistants know to pull and route those items; if labeling or relabeling is involved, those cells are highlighted in Yellow and the student assistants know to print and apply labels; if the anomalies and issues are resolved, those cells are highlighted in Green; and if there are no anomalies or issues involved, those cells remain unhighlighted. Aside from the red highlighted items, the items are then passed to a final quality assurance check by the students to verify the accuracy of key elements such as the title statement, the publication statement, and the presence of illustrations, bibliographical references, and indexes.

Here we have a snippet of the post-cataloged spreadsheet, as shared with the student assistants.

What Do We Get?

- A process that identifies exceptions rather than being driven by them
- Time-labor savings – the triage process for shipments takes about a half-hour or less vs. multiple hours
- Time-labor savings – the final quality assurance checks have fewer and, if present, more obvious anomalies
- Time-delivery savings – the overall cataloging pipeline is day(s) vs. a week plus
- Financial savings – the price differential between the basic services we obtain from the vendor vs. the high-level cataloging services is approximately $7 per title, regardless of whether the high-level vendor cataloging service is warranted, when the vast majority of titles don't require it.

Slide 11:

What do we get:

- A process that identifies exceptions rather than being driven by them

- Time-labor savings – the triage process for shipments takes about a half-hour or less vs. multiple hours

- Time-labor savings – the final quality assurance checks have fewer and, if present, more obvious anomalies

- Time-delivery savings – the overall cataloging pipeline is a day or so vs. a week or more

- Financial savings – the price differential between the basic services we obtain from the vendor vs. the vendor's high-level cataloging services is approximately $7 per title, regardless of whether the high-level vendor cataloging service is warranted, when the vast majority of titles don't require it.

Pitfall and Prospect

The Pitfall:

- The process is predicated on Worldshare Cataloging Partners and sourcing records from OCLC

The Prospect

- Using API's to not just retrieve records data but also to edit it

Slide 12:

A couple details I was reminded of yesterday:

      The pitfall, as we all face budgetary pressures and OCLC subscription prices, is the process' dependence on OCLC. While it could work to identify issues with records from any source, a source of improved records could be a challenge. In-house editing or Z39.50 searching may be a workaround.

      The prospect is deploying APIs not just to harvest the data for assessment but also to roundtrip edits back to Alma in corrected form. Following the catalogers Hippocratic guidance – first, do no harm – we took the conservative approach to engage these edits on a human mediated basis.

Slide 13:

      I will close with a QR code to our Python code, available on GitHub, which will then require an institution-specific API key to tap into one's Alma instance to plug in the coding.

      If you are interested in our process documentation, for running the Triage Program and Process, or for our Final Quality Assurance checks, here is my email.

[A live demo followed]

[Q&A followed]

Resources

Github:
https://github.com/schaffer-library

Email:

myersj@union.edu

Image Credits

Slide 2: Schaffer Library, personal library

Slide 3: Jigsaw puzzle pieces, Nathalia Segato, via Unsplash.com

Slide 4: OCLC logo, OCLC; EBSCO logo, EBSCO

Slide 6: Alma screen grab, personal library

Slide 7: Triage program screen grab, personal library

Slide 8: Triage program screen grab, personal library

Slide 9: Triage program screen grab, personal library

Slide 10: Triage program screen grab, personal library

Questions?

Live Demo?