Streamlining and Enhancing Copy Cataloging Workflows Through APIs and Workflow Modifications

John Myers, Catalog & Metadata Librarian, Union College
(myersj@union.edu)

Good afternoon and good morning. Thank you for joining to hear about evolution and revolution of cataloging workflow at Union College. This is similar to a presentation delivered at SUNYLA in February 2021. I would like to acknowledge the significant coding contributions of a student, Hamza Ghumman, who interned with me during summer 2020. He was instrumental in developing the Java program that is the heart of our revised workflow. He was unavailable to co-present today.

Pre-COVID Workflows

- Use of OCLC's WorldShare Cataloging Partners to deliver files of records corresponding to shipped materials from our primary vendor
- Assessment by student employees, on a one-by-one basis with items in hand, of shipped materials into "ready to process" and "needs cataloger attention" streams
- Resolution of "needs cataloger attention" items by hourly cataloging staff (or sent to Librarian for higher level correction).
- Processing of materials in the respective streams

The process outlined here is the "two touch" method alluded to in the program description, in which items are looked at individually and in a fairly detailed manner at two steps of the process – once at the cataloging phase prior to processing and then once more to be reviewed following that processing. Despite being facilitated by the deployment of OCLC's WorldShare Cataloging Partners, with the delivery of matching OCLC records to overlay our brief vendor records, it is little different from the classic cataloging mode that might be characterized as, "pick up a book, catalog a book, set down the cataloged book, repeat."

After processing is complete, all items then received the "second touch" of going through a Final Quality Assurance Review. As with the initial "touch" this was done on a one-by-one basis

NY PAUSE Workflow and Forecasts

- Ordering continued but with no access to delivered materials
- Remote assessment by hourly staff, on a one-by-one basis working from an Alma set, generated from the import process, of shipped materials into "ready to process," "able to be edited remotely," and "require item in hand" streams
- Recognition that uncertainties for work dynamics in the fall term dictated that we explore pre-processing
- Recognition of a strong need to be able to assess materials "in batch"

In mid-March of 2020 we evacuated campus, when the first on-campus positive case was reported. This was rapidly followed by the issuance of state-wide executive orders that instituted the NY Pause. These saw all non-essential businesses shift to remote operations, including college campuses beyond essential services. The NY Pause was finally lifted for academic institutions at the end of July 2020, although precautionary restrictions persisted. As implemented at our campus, these included de-densifying occupancy by employees, including student employees, and stringent guidance to not engage the onsite community when presenting any COVID-correlated symptoms. Practically up to the point that the Fall term started, it was unclear to what degree the Fall term would be onsite, blended/hybrid, or fully remote.

We had long held that use of student employees to process materials was more economical and offered more control than would be available through a pre-processing contract with a vendor. The unreliability of access to a student labor pool though altered this equation significantly. We had caught up on the processing from the purchases during the NY Pause through the assistance of other campus departments but this would not be available during the school year for future purchases.

A major advantage to preprocessing would be the ability to route things almost directly to the shelves, but there were several details that we desired to address but which could not

be profiled for – including the treatment of oversize items. Examining books one-by-one would lessen the efficiency gained by the preprocessing, especially in light of insights gained during the NY Pause that there were fairly few situations under which an item could not be processed/labeled, and that there were extremely few cataloging anomalies that would warrant re-processing.

We needed an effective means to survey the records for an entire shipment of books without the need to look at them one-by-one

New Workflow

- The Alma set associated with each shipment is exported in Excel format
- The Excel data is input to the new program, which generates API queries back into Alma
- New Excel data is output with evaluative reports
- Hourly staff members make corrections where possible, remotely, and round-trips an updated, color coded version of the Excel data
- Student employees use the new spreadsheet to identify items that need to be (re)labeled or that need to be pulled for Librarian attention

In our new workflow, the first one-by-one "touch" has largely been eliminated. Through use of the internally developed Java program, an entire file or set of records is addressed in aggregate. Specific problems are reported out and then resolved. Student employees can match the reported problem with the corresponding item and pull it from the truck for specialized handling. The balance of a shipment passes through without being examined as part of a "first touch."

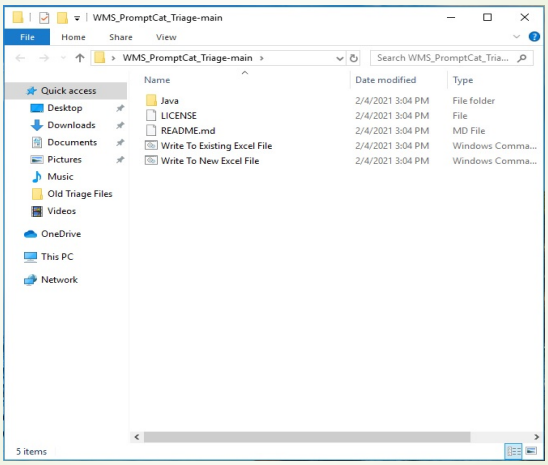
With most of the processing done by the vendor and minor adjustments addressed, materials quickly move to the Final Quality Assurance Review. This is still done on a one-by-one basis as before.

What are we looking for?

- What is the condition of the call number – absent, complete, more than one?
- What is the class letter (for our non-default floor location)?
- What is the size (for prefix and non-default floor location)?
- Is a print (vs. online) format record delivered?
- Are multiple volumes involved (or lack of pages in a single volume)?
- Are there extra spaces in key fields?

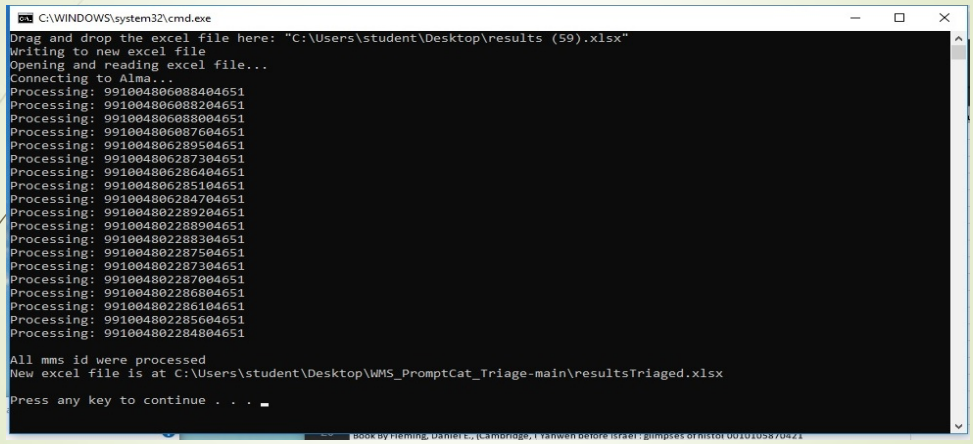
Based on our experience of "cataloging remotely" during the NY Pause, there were several details that we identified as important for facilitating the processing of new materials and addressing major issues with delivered records. As you can see, concerns with call numbers predominated but not exclusively so.

Our Interface



Hamza, my student intern, pulled the program elements into a convenient folder that resides on our workstations. He developed options to overwrite the source file or create a new file, although we have only deployed the “Create a New File” option. Opening this option provides access to a command line interface into which the source file is dragged and then the program is executed.

Program Interface (post processing)



The program processes the MMS IDs, that is, Alma's system control numbers, and returns this output in the interface and a new Excel file in the program's folder.

The Output Excel

Title	Overall record condition	Call number assessment	Floor status	Size status	Format assessment	Bri	Coding problems
Great exodus, great wall, great party / Chessy Normile ; [introc	High quality Record, Minimal Revi	PS3614.O7655 G74 2020		No edits: Regular size		9	
The handbook of China's financial system / edited by Marlene	High quality Record, Minimal Revi	HG3334 .H36 2020		No edits: Regular size		10	
In praise of forgiveness / Massimo Recalcati ; translated by Ali	High quality Record, Minimal Revi	BF637.F67 R43 2020	Change location to floor2	No edits: Regular size		10	
The radiant lives of animals / Linda Hogan.	High quality Record, Minimal Revi	PS3558.O34726 A6 2020		No edits: Regular size		10	
The amateur hour : a history of college teaching in America / J	High quality Record, Minimal Revi	LB2331 .Z56 2020		No edits: Regular size		10	
An introduction to psychological tests and scales / Kate Mirian	High quality Record, Minimal Revi	BF176 .L64 2021	Change location to floor2	No edits: Regular size		9	
Mercy : a memoir of medical trauma & true crime obsession / I	High quality Record, Minimal Revi	RC552.P67 T73 2020		No edits: Regular size		9	
Social and emotional learning in the Mediterranean : cross cul	High quality Record, Minimal Revi	LB1072 .S67 2020		No edits: Regular size		10	
The wind traveler : a novel / Alonso Cueto ; translated by Fran	High quality Record, Minimal Revi	PQ8498.13.U34 V5313 2020		No edits: Regular size		10	
Aramaean borders : defining Aramaean territories in the 10th-	High quality Record, Minimal Revi	DS59.A7 A73 2019	Change location to floor2	No edits: Regular size		10	
Calvin's political theology and the public engagement of the c	High quality Record, Minimal Revi	BX9418 .T85 2017	Change location to floor2	No edits: Regular size		9	
Cynicism and Christianity in antiquity / Marie-Odile Goulet-Ca	High quality Record, Minimal Revi	B508 .G68213 2019	Change location to floor2	No edits: Regular size		10	
Handbook of conspiracy theory and contemporary religion / e	High quality Record, Minimal Revi	BP603 .H355 2019	Change location to floor2	No edits: Regular size		10	
Hidden Divinity and Religious Belief : New Perspectives / ed	Substandard Record, Overlay	BL200 .H53 2018	Change location to floor2	No edits: Regular size	Multivolume title; create items	3	
Inward baptism : the theological origins of evangelicalism / Ba	High quality Record, Minimal Revi	BT21.3 .T56 2020	Change location to floor2	No edits: Regular size		10	
Jonathan Z. Smith on religion / Christopher I. Lehrich.	Pre-publication Record, Upgrade	BL43.S645 L44 2021	Change location to floor2	Manually Review: Missing siz		5	
New directions in the anthropology of dreaming / edited by Je	High quality Record, Minimal Revi	BF1078 .N4525 2021	Change location to floor2	No edits: Regular size		10	
The science of religion : a defence: essays by Donald Wiebe / e	High quality Record, Minimal Revi	BL41 .W54 2019	Change location to floor2	No edits: Regular size		10	
Yahweh before Israel : glimpses of history in a divine name / D	High quality Record, Minimal Revi	BM610 .F55 2021	Change location to floor2	No edits: Regular size		10	

And here we have the results.

Off screen, on the left side of the Excel output in this screen capture is an MMS ID column, to facilitate searching in Alma

We included overall record condition and also a report of our local Brief Level Rule as a back up for assessing the records, but the other columns are effective on their own.

As previously noted

- We test for call number issues – if only a single LC call number is present, that call number is the output; if there are less than 1 or more than 1 then there are alerts for those conditions. This set has no call number issues.
- We test for location determinations – most of our books go into our basement stacks and this is our default. The LC range A-D, as well as oversize books, are all shelved on floor 2
- We test for size determinations – if oversize, then we also add a size prefix. Most of those here are normal sized, with one record a CIP record that will require physical examination (or overlay if the record has been improved in OCLC in the interim between the vendor working on the item and our work on the item)
- We test for format issues – chiefly concerned about delivery of electronic resource

records for our tangible resources, but also the potential need for resolving multivolume sets. The sieve is crude for that, so we are typically also catching incomplete statements of extent (that is, 1 v.)
- Coding problems catches a few other issues, such as the extra spacing within a field.

This is a somewhat anomalous report, with the large cluster of B-classed titles, requiring edits to the holding location; we don't normally see so many location edits.

The Edited Excel

Title	Overall record condition	Call number assessment	Floor status	Size status	Format assessment	Bri	Coding problems
Great exodus, great wall, great party / Chessy Normile ; [intro	High quality Record, Minimal Revie	PS3614.O7655 G74 2020		No edits: Regular size		9	
The handbook of China's financial system / edited by Marlene	High quality Record, Minimal Revie	HG3334 .H36 2020		No edits: Regular size		10	
In praise of forgiveness / Massimo Recalcati ; translated by Ali	High quality Record, Minimal Revie	BF637.F67 R43 2020	Change location to floor2	No edits: Regular size		10	
The radiant lives of animals / Linda Hogan.	High quality Record, Minimal Revie	PS3558.O34726 A6 2020		No edits: Regular size		10	
The amateur hour : a history of college teaching in America / J	High quality Record, Minimal Revie	LB2331 .Z56 2020		No edits: Regular size		10	
An introduction to psychological tests and scales / Kate Miriam	High quality Record, Minimal Revie	BF176 .L64 2021	Change location to floor2	No edits: Regular size		9	
Mercy : a memoir of medical trauma & true crime obsession / I	High quality Record, Minimal Revie	RC552.P67 T73 2020		No edits: Regular size		9	
Social and emotional learning in the Mediterranean : cross cul	High quality Record, Minimal Revie	LB1072 .S67 2020		No edits: Regular size		10	
The wind traveler : a novel / Alonso Cueto ; translated by Fran	High quality Record, Minimal Revie	PQ8498.13.U34 V5313 2020		No edits: Regular size		10	
Aramaean borders : defining Aramaean territories in the 10th-	High quality Record, Minimal Revie	DS59.A7 A73 2019	Change location to floor2	No edits: Regular size		10	
Calvin's political theology and the public engagement of the c	High quality Record, Minimal Revie	BX9418 .T85 2017	Change location to floor2	No edits: Regular size		9	
Cynicism and Christianity in antiquity / Marie-Odile Goulet-Ca	High quality Record, Minimal Revie	B508 .G68213 2019	Change location to floor2	No edits: Regular size		10	
Handbook of conspiracy theory and contemporary religion / ec	High quality Record, Minimal Revie	BP603 .H355 2019	Change location to floor2	No edits: Regular size		10	
Hidden Divinity and Religious Belief : New Perspectives / edit	Substandard Record, Overlay	BL200 .H53 2018	Change location to floor2	No edits: Regular size	Multivolume title; create items	3	
Inward baptism : the theological origins of evangelicalism / Ba	High quality Record, Minimal Revie	BT21.3 .T56 2020	Change location to floor2	No edits: Regular size		10	
Jonathan Z. Smith on religion / Christopher I. Lehrich.	Pre-publication Record, Upgrade	BL43.S645 L44 2021	Change location to floor2	Manually Review; Missing siz		5	
New directions in the anthropology of dreaming / edited by Je	High quality Record, Minimal Revie	BF1078 .N4525 2021	Change location to floor2	No edits: Regular size		10	
The science of religion : a defence: essays by Donald Wiebe / e	High quality Record, Minimal Revie	BL41 .W54 2019	Change location to floor2	No edits: Regular size		10	
Yahweh before Israel : glimpses of history in a divine name / D	High quality Record, Minimal Revie	BM610 .F55 2021	Change location to floor2	No edits: Regular size		10	

Once edits are made by the Copy Cataloger, the cells previously flagged by text are color filled to indicate the action taken and results:

- Green filled cells – indicate a condition has been addressed and there is no impact on the existing label
- Yellow filled cells – indicate a condition has been addressed and a label needs to be printed and applied to the item (either correcting an existing label or addressing the lack of a label)
- Red filled cells – indicate a condition has not been addressed and requires physical examination to resolve or Librarian attention

This color coding guides the students in triaging the shipment into items that need to be pulled for further attention; items that need labelling or relabeling; and items that are ready for the Final Quality Assurance Review.

Current Issues and Future Possibilities

- Keeping files in sync with shipments
- Small shipments
- Could retool the coding to work from MMS ID outputs in the original Acquisitions "All Titles" set
- Reporting out the presence of extraneous web content
- Reporting the mismatch between variable and fixed field coding
- Give the program "Write" access to Alma to automate some of the edits?

Looking ahead there are things that we might further adjust or do differently:

- The expectation was that we would get one decent sized shipment of materials a week, and that these shipments would neatly surge through the department. This has not universally been the case, so it requires some care to make sure that the program outputs are properly matched to the corresponding truck(s)

- Additionally, we have recently been seeing small "mid-week" shipments or even weeks with a "short" shipment. The Return on Investment of running these smaller shipments and files through the program is questionable.

- During the NY Pause, when conducting one-by-one remote review of shipment entries, we fiddled with the Alma set output from the import job report. This mentality carried forward during the development phase for the program. The program is currently dependent on the column assignment for the MMS ID data as output by Alma from Alma's "Physical Item" set. With ongoing use of the program however, we have realized that the baseline source set may be sufficient. The program could be adjusted to work directly from the "All Titles" set scope that corresponds to the baseline set output from the import job report.

- As a policy, we generally remove 856 fields for links to supplemental web content. This isn't uniformly addressed in editing though and it would be helpful to highlight when they are present.

- A major cluster of Final Review issues is a mismatch between what is recorded in variable fields and their corresponding fixed fields, such as the presence of illustrations, bibliographical references, and indexes. The program can't help if the record is internally consistent but wrong, but it could report out those instances of internal inconsistency.

- At present, in an abundance of caution, our API key only supports "Read" access to Alma. There may be value in exploring if some edits through the API could be supported if it had "Write" access – such as modifying the location code value, adding shelf location prefixes, or removing extraneous spaces.

Program Availability

GitHub:

https://github.com/Hamza-Ghumman/WMS_PromptCat_Triage

Here is the GitHub link for the raw program. While configured for Alma, it is hoped that any system that supports both import reports and API interaction can leverage this framework for aggregate assessment of files of catalog records.

Thank you for your attention today.

Selected Q&A comments (with expansions from extemporaneous replies during the presentation)

- Please put the GitHub link in the chat
- How was remote cataloging accomplished?
- What was the nature of interactions with the student who did the coding?
- Did you have anybody that did not have remote access from home?

- GitHub link (in hopes that it is adequately captured in the PDF of the notes, if not accessible in the source slide)
    - https://github.com/Hamza-Ghumman/WMS_PromptCat_Triage

- How was remote cataloging accomplished?
    - For our regular purchases, we have OCLC records delivered and matched using OCLC's WorldShare Cataloging Partners program (formerly known as PromptCat). During the NY Pause, for the most part, we had to go on faith that the records that "looked good" were good. These included LC records but also full-level records contributed by any OCLC member institution. The major caution was looking out for M-level vendor-sourced records, for which there were 3 outcomes: we found such records had been improved; we could reliably match to a better record; or we just had to wait until we could verify data and record selection with item in hand. Outside of record overlays for improved versions of the same record or alternate records for the same content, we were primarily "cataloging" to adjust for size and location, to support rapid processing once we returned.
    - For pair of large few gifts, that were addressed while I and the students were onsite part-time but my copy cataloger remained off-site, the bibliographic

information for items were entered into Google Sheets. Usually just ISBN and publication year when the ISBN was available, otherwise more complete data of author, title, publisher, and date. The copy cataloger accessed this data and made best possible matches in OCLC from it. Any problems with a loaded record or inability to select a record were annotated and then I addressed those when onsite. The copy cataloger's selections were then verified as part of the final QA review.

- What was the nature of interactions with the programming intern?
    - Based on experiences from the "remote cataloging" during the NY Pause, I had a clear idea of what I wanted to see happen to maximally support processing of items with minimal cataloger intervention. These are outlined on one of the earlier slides in this presentation. I discussed these with the intern (over Zoom), who then went away to pound on the coding to generate outputs. He presented what he had and then we had a few cycles of refinements, including the examination of extra spaces for example. In an early version, they caused the program to crash. He refined his coding but then mentioned, "I could have the program point them out so they can be fixed in the record," to which I quickly acceded. It was a very conversational process. I think he was very pleased to explore the possibilities of addressing a "real world" problem on his own.

- Did you have anybody that did not have remote access from home?
    - Our ILS is Cloud based and general campus communications are through Google Apps for Education, also cloud based. We also implemented Trello.
    - All staff and librarians had home-based internet access
    - Depending on circumstances, individuals used personal PCs/Macs, took home their campus-issued workstation, or obtained a campus-issued loaner laptop.