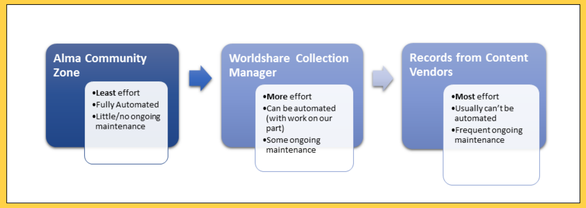
Evaluating and Loading Ebook Metadata from OCLC WorldShare Collection Manager

Kelly Thompson & Stacie Traill
University of Minnesota Libraries

Why WorldShare Collection Manager?

- For each ebook package, we select a metadata source by attempting to balance metadata **coverage** and **completeness** vs. effort required.



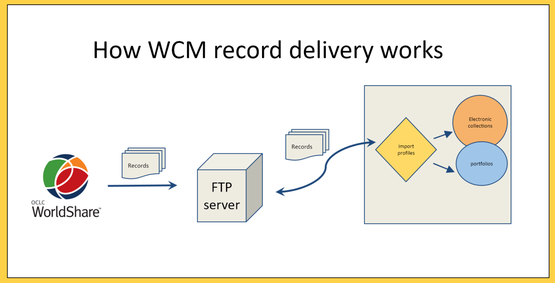
- Record completeness is evaluated by a Python script, which assigns a completeness score to each record based on presence of elements that can affect resource discovery and identification.

	mean	stdev	total_re	isbn_co	cat_lan	rsrc_la	rda	classif	author	alt
Mean Record Score	25.9412									
Standard Deviation		5.35573								
ocn810314027			31	4	1	1	0	1	0	
ocn847231436			21	1	1	1	0	1	0	
ocn847232913			23	1	1	1	1	1	1	
ocn847514547			24	1	1	1	1	1	0	
ocn847517654			18	1	1	1	1	1	0	
ocn857067301			21	4	1	1	0	1	1	
ocn873761834			34	4	1	1	0	1	0	
ocn886887634			28	6	1	1	0	1	0	
ocn893628027			38	8	1	1	0	1	0	
ocn904283678			25	3	1	1	1	1	0	
ocn904377446			30	4	1	1	1	1	0	
ocn913335633			25	4	1	1	1	1	0	
ocn914274252			28	6	1	1	0	1	1	
ocn919501347			22	2	1	1	0	1	0	
ocn921309520			22	4	1	1	1	1	0	
ocn956739626			30	4	1	1	1	1	0	

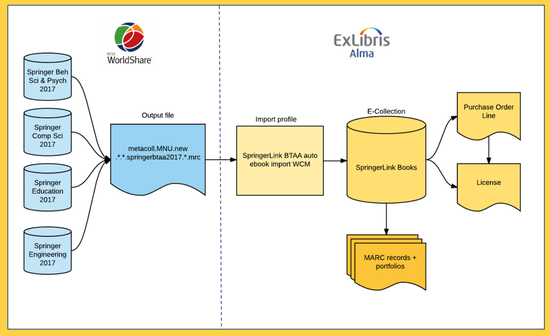
- For collections available from multiple sources, WCM usually offers more complete records.

Automating WorldShare/Alma Communication

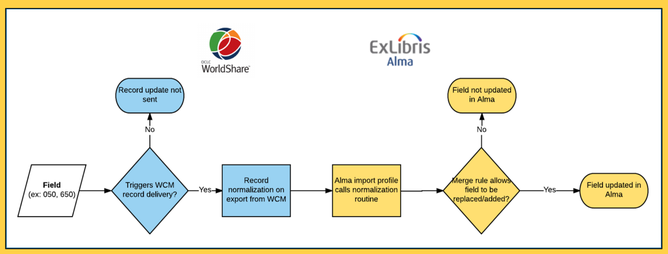
- Records are automatically delivered via FTP



- Records are imported using import profiles that fetch files from the FTP on a schedule.

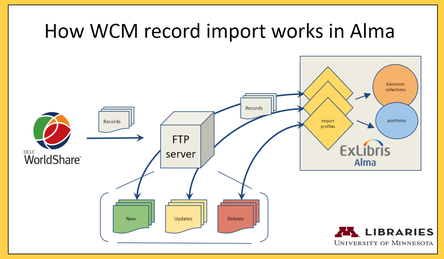


- Record normalization, delivery schedule, and additions are customized in WorldShare.



Automating Record Updates

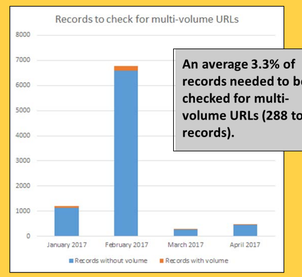
- Records are supplied in 4 groups: new, updates, merges, and deletes.



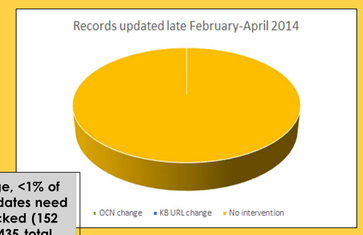
- Scheduled Python scripts fetch new and updated files from the server and analyze record quality indicators.

New records	Update records
▸ Import, then manually clean-up ▸ "Provide access first, then worry about data quality" ▸ Don't want to delay access because the description isn't perfect ▸ Looking for: ▸ Multi-volume records ▸ Language of cataloging not English ▸ Records for print resources ▸ Output: reports	▸ Filter with a script before import ▸ "Access exists, do no harm" ▸ Can afford to wait to do analysis before updating the description ▸ Looking for: ▸ URL changes handled manually ▸ OCN changes checked manually after automatic process ▸ Everything else automatically processed ▸ Output: Record batches

- Scripts generate pre-sorted record files or reports; staff spend targeted time on records with highest-impact problems on discovery & access. Others are handled automatically.



An average 3.3% of records needed to be checked for multi-volume URLs (288 total records).



On average, <1% of record updates need to be checked (152 out of 354,435 total records).

LIBRARIES
UNIVERSITY OF MINNESOTA