Integrating Alma and OCLC Worldshare Collection Manager for Ebook Management

Stacie Traill & Kelly Thompson
Metadata Analysts
University of Minnesota - Twin Cities

ELUNA
May 10, 2017

LIBRARIES
UNIVERSITY OF MINNESOTA

Agenda

- Background and Initial Decisions
  - When to use WCM?
  - Evaluating WCM vs. Alma community zone
  - How WCM record delivery works
- Getting set up
  - Worldshare customization and Alma Import profiles for new records
  - To delete or not to delete? Service issues, part 1
- Handling Updates
- Automating processing and assessment
  - Record quality
  - Service issues, part 2
  - Feeding Alma pre-processed records
  - Contributing to record improvement

University of Minnesota Background

- Alma early adopter (live in December 2013).
- Heavy users of OCLC's Worldcat Collection Sets service for ebook records.
- Required to migrate Collection Sets to Worldshare Collection Manager in Fall 2014.
- Limited staff time to devote to frequent, ongoing manual batch edits and imports.
- 141 Worldshare collections currently configured for MARC record delivery, used to populate 32 Alma ebook collections.

LIBRARIES
UNIVERSITY OF MINNESOTA

When does it make sense to use WCM?

- No appropriate Community Zone collection available (but WCM offers).
- Full local control over bibliographic records is desired.
- Library has the resources (i.e. personnel and time) to manage coordination of WCM with Alma, and ongoing import/update processes.
- It's the only option (EBSCOhost ebooks).
- Can help to align metadata/portfolio management with licensing (when multiple CZ activations would be needed).

Worldshare Collection Manager or Alma Community Zone?

- Coverage
  - Is total KB title count accurate (compared to vendor lists)?
  - For WCM, is a Worldcat record available for every KB entry in a collection?
- Record Completeness
  - Are metadata elements that can affect discovery and identification of the resource present in most records?
- Accuracy
  - Does the record match the resource?
  - Are specific values in each record accurate?

LIBRARIES
UNIVERSITY OF MINNESOTA

Collection Name	WCM	CZ
Sage Knowledge Reference 2016	28/29	26/26
ENVIROnetBASE	1068/1096	1266/1266
Gale Dictionary of Literary Biography	401/414	425/425
IEEE-Wiley Ebooks2010 & Prior	507/509	509/509
National Academies Press Monographs	6307/6317	2258/2258
Oxford Reference Premium*	321/330	366/366
Wageningen Academic ebooks 2002-2017*	266/279	333/333

Coverage

Records vs. KB entries

- Download WCM KBART file
- Filter on oclc_number

**Not always a valid cross-KB comparison. Same- or similarly-named collections may not represent the same content.*

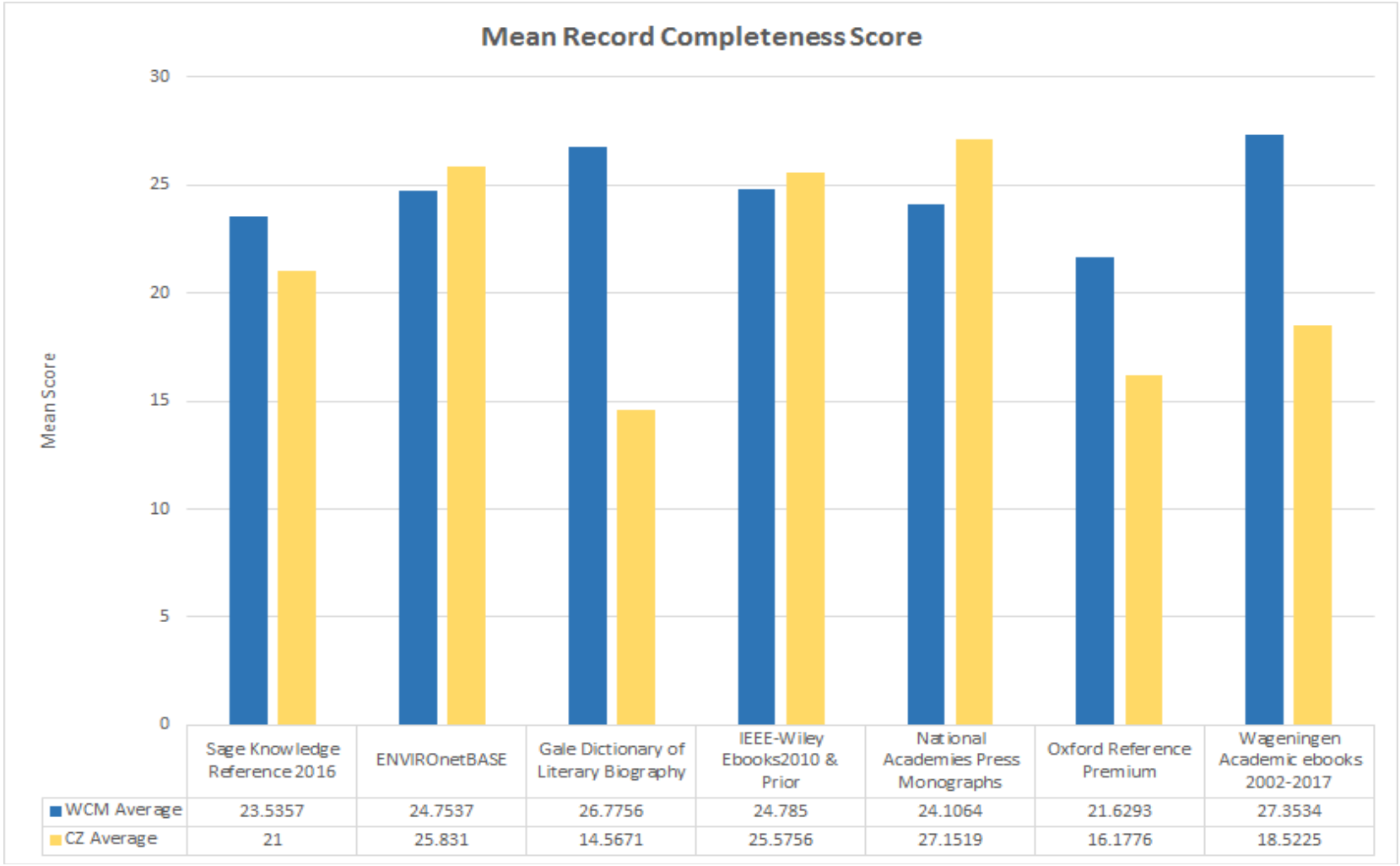
LIBRARIES
UNIVERSITY OF MINNESOTA

Completeness

- Past practice: impressionistic, eyes-on.
- Evolving practice: assign completeness score to every record in a file via script that counts elements affecting discovery or identification of the resource.
- Work in progress (currently a very rough measure).

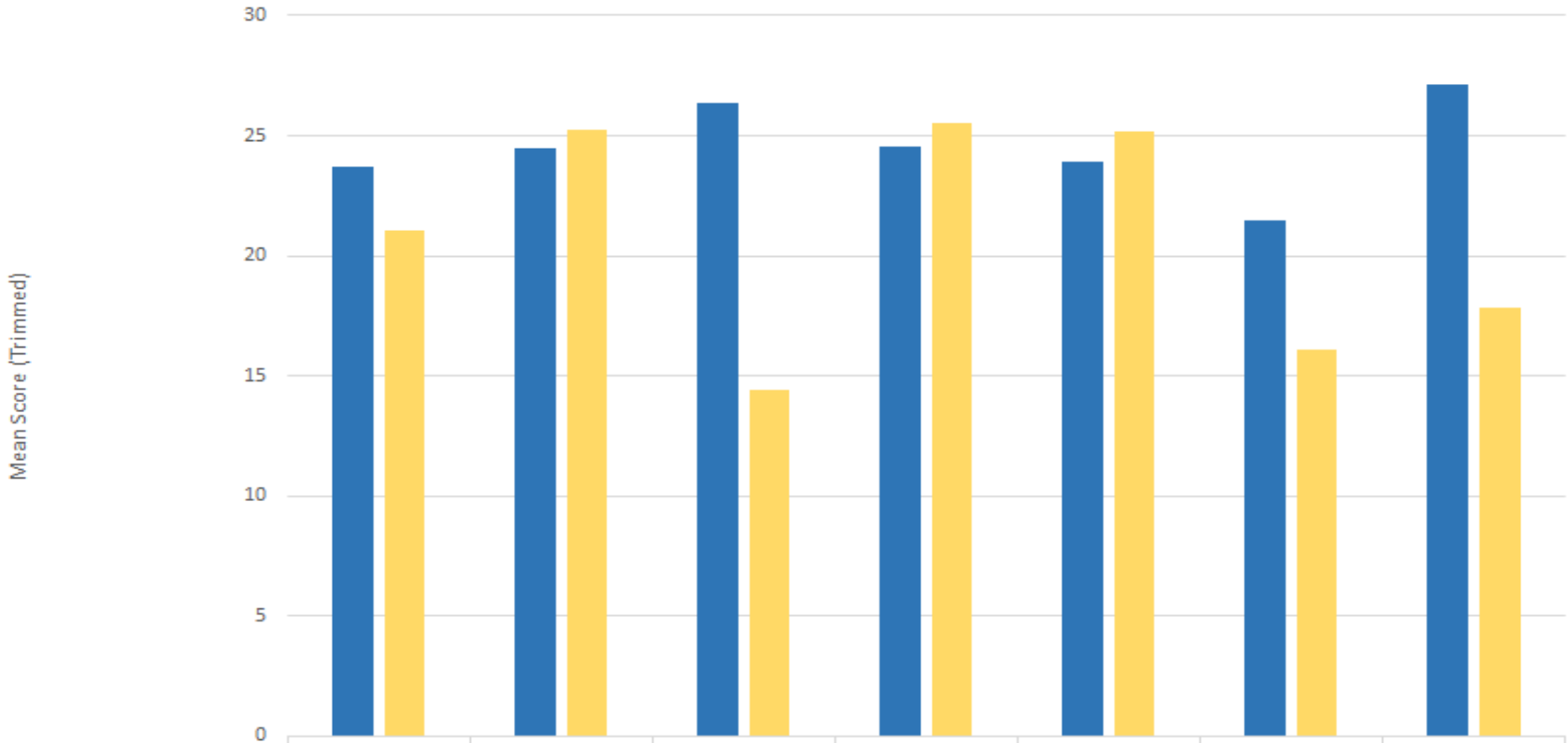
```
        try:
            date_008_int = int(date_008[0])
        except ValueError:
            date_008_int = 1111

        if date_008_int == int(date_imprint[0]):
            rec_qualities['dates_matched'] = 1
        else:
            rec_qualities['dates_matched'] = 0

        lcsh_count = 0
        lcsh_re = re.compile('^.{7}0.*$')
        for subj in subjects:
            if lcsh_re.match(str(subj)):
                lcsh_count += 1
                rec_qualities['subjects_lcsh'] = lcsh_count

        online_rsrc_re = re.compile('^.*online.resource.*$')
        for field in descr:
            subf_a_300 = field.get_subfields('a')
            if online_rsrc_re.match(str(subf_a_300)):
                rec_qualities['descr'] = 1
            else:
                rec_qualities['descr'] = 0

        cat_lang_re = re.compile('^.*eng.*$')
        for field in rec_source:
            subf_b_040 = field.get_subfields('b')
            if subf_b_040 and not cat_lang_re.match(str(subf_b_040)):
                rec_qualities['cat_lang'] = 0
            else:
                rec_qualities['cat_lang'] = 1

        rec_score = 0
        for k, v in rec_qualities.items():
            rec_score = rec_score + v
        rec_qualities['total_rec_score'] = rec_score

        record_dict[record_id] = rec_qualities

    return record_dict
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		isbn_count	authors	alt_titles	edition	toc	abstract	subjects	contribs	series	date_008	date_26X	dates_matched	subjects_lcsh	descr	cat_lang	total_rec_score	
2	906009220	4	1	0	0	0	1	2	1	0	1	1	1	1	1	1	15	
3	907238072	8	0	1	0	1	1	9	1	0	1	1	1	3	1	1	29	
4	908175724	4	0	1	0	1	1	4	3	0	1	1	1	2	1	1	21	
5	908514078	6	0	0	0	1	1	2	1	0	1	1	1	1	1	1	17	
6	908769188	8	0	1	0	1	1	7	2	0	1	1	1	2	1	1	27	
7	910932116	4	0	2	0	1	1	3	1	0	1	1	1	1	1	1	18	
8	911616918	6	0	1	0	1	1	7	1	0	1	1	1	1	1	1	23	
9	912323058	3	0	0	0	1	1	12	4	0	1	1	1	2	1	1	28	
10	913660907	7	0	2	0	1	1	4	1	0	1	1	1	1	1	1	22	
11	913660908	8	0	0	0	1	1	6	1	1	1	1	1	2	1	1	25	
12	913660909	6	0	1	0	0	1	5	1	0	1	1	1	1	1	1	20	
13	919002446	8	0	1	0	1	1	5	3	0	1	1	1	2	1	1	26	
14	919431798	4	0	1	0	0	1	17	1	0	1	1	1	3	1	1	32	
15	919573714	6	0	1	1	0	1	8	2	0	1	1	1	1	1	1	25	
16	919582408	8	0	0	1	1	1	6	1	0	1	1	1	2	1	1	25	
17	920521078	6	0	1	1	1	1	13	1	0	1	1	1	2	1	1	31	
18	927140606	6	0	2	0	0	1	10	2	0	1	1	1	2	1	1	28	
19	936331814	6	0	1	0	1	1	8	2	0	1	1	1	3	1	1	27	
20	936331906	6	0	0	1	0	1	6	1	0	1	1	1	2	1	1	22	
21	945736639	6	0	0	0	1	1	3	1	0	1	1	1	1	1	1	18	
22	950971839	8	0	0	0	1	1	9	1	0	1	1	1	3	1	1	28	
23	951221119	6	0	1	0	1	1	8	1	0	1	1	1	3	1	1	26	
24	951223413	7	0	0	0	0	1	10	1	0	1	1	1	4	1	1	28	
25	951226613	6	0	2	1	0	1	7	1	0	1	1	1	3	1	1	26	
26	954270412	6	0	1	0	0	1	5	2	0	1	1	1	2	1	1	22	
27	960909445	4	0	1	0	1	1	16	2	0	1	1	1	4	1	1	34	
28	965975062	2	0	0	0	0	0	0	1	0	1	1	1	0	1	1	8	
29	965982047	2	0	0	0	0	0	0	1	0	1	1	1	0	1	1	8	
30																	23.53571429	Mean
31																	23.73076923	Trimmed mean



	Sage Knowledge Reference 2016	ENVIROnetBASE	Gale Dictionary of Literary Biography	IEEE-Wiley Ebooks2010 & Prior	National Academies Press Monographs	Oxford Reference Premium	Wageningen Academic ebooks 2002-2017
WCM Average	23.5357	24.7537	26.7756	24.785	24.1064	21.6293	27.3534
CZ Average	21	25.831	14.5671	25.5756	27.1519	16.1776	18.5225

Trimmed Mean Record Completeness Scores (middle 80%)



	Sage Knowledge Reference 2016	ENVIROnetBASE	Gale Dictionary of Literary Biography	IEEE-Wiley Ebooks2010 & Prior	National Academies Press Monographs	Oxford Reference Premium	Wageningen Academic ebooks 2002-2017
WCM 80% Trimmed Mean	23.7308	24.501	26.374	24.5536	23.9523	21.4567	27.1542
CZ 80% Trimmed Mean	21.0417	25.2351	14.4308	25.5098	25.1519	16.0636	17.8704

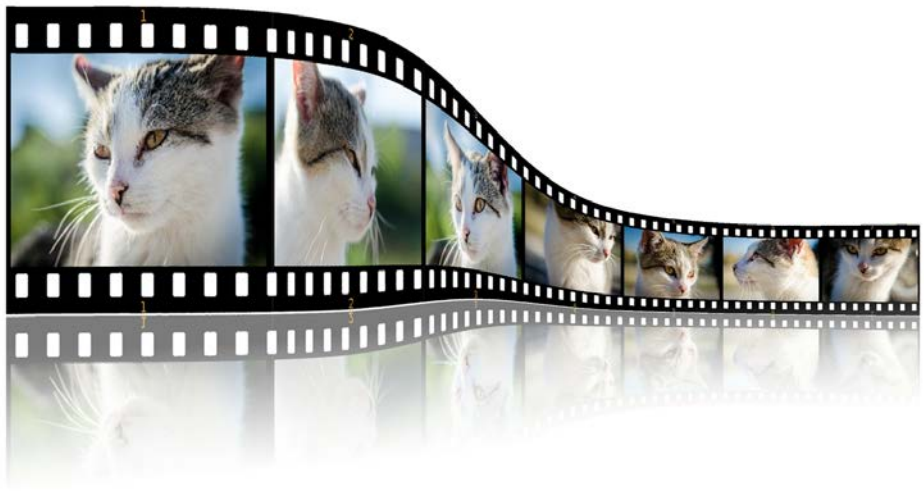
Accuracy

- Automated evaluation is hard.
- Full eyes-on evaluation is not realistic.
- Spot-checking has some value.
- Work in progress.

Disclaimer

All measures are snapshots.

Title counts, record availability, completeness, accuracy -- all are moving targets.



LIBRARIES
UNIVERSITY OF MINNESOTA

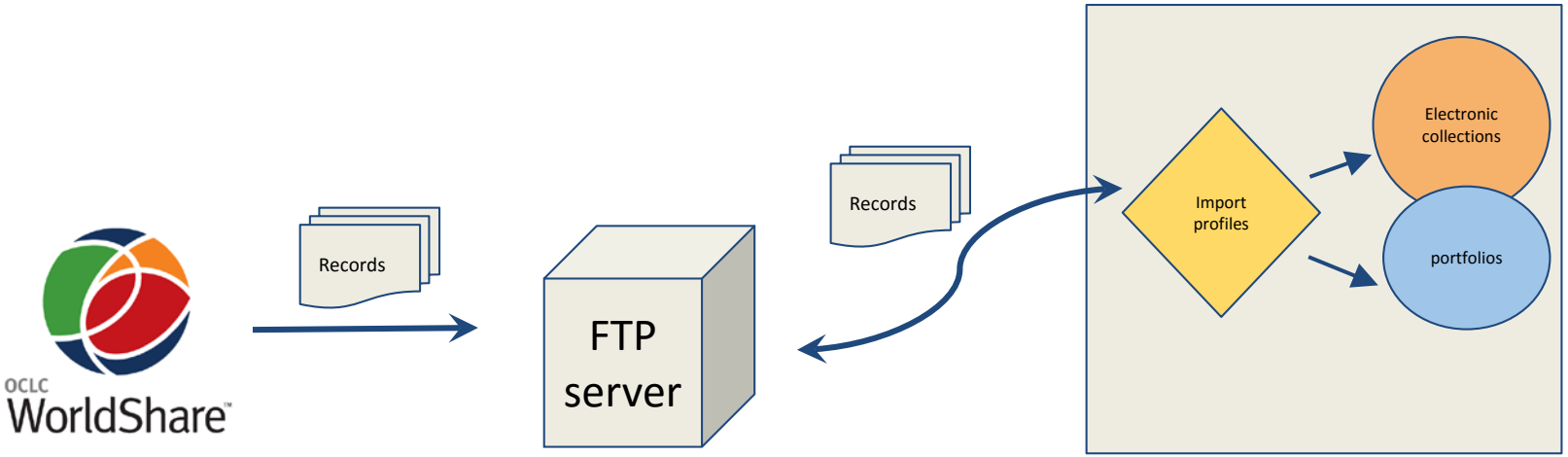
Community Zone metadata enrichment

Alma April release notes:

- Community Zone bibliographic records were enriched from different providers as follows:
  - ProQuest: 1.2 million records
  - IEEE: 320 records
  - Cambridge University Press: 31250 records
  - Credo Reference: 700 records
  - Springer: 195000 records
  - Oxford University Press: 25000 records

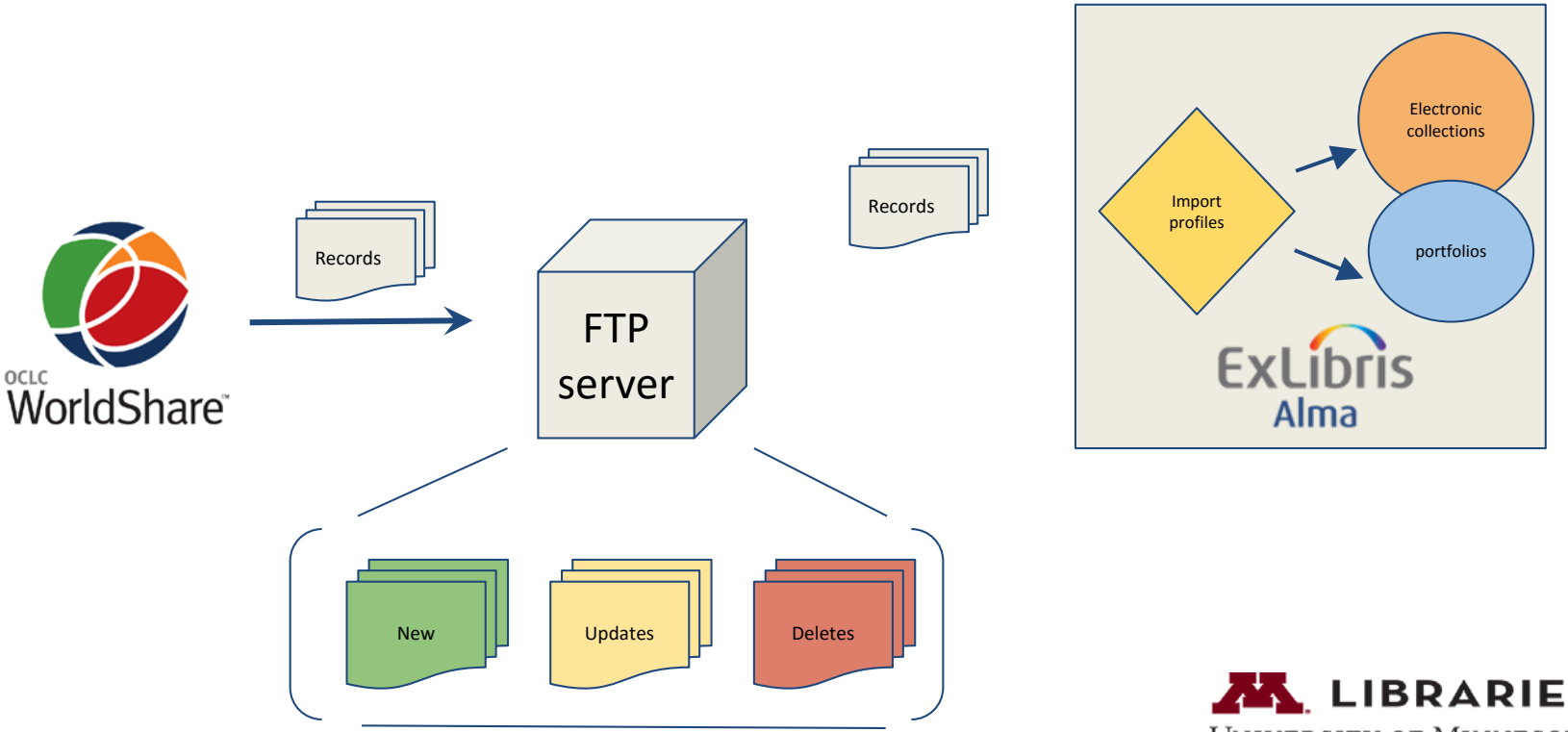
LIBRARIES
UNIVERSITY OF MINNESOTA

How WCM record delivery works



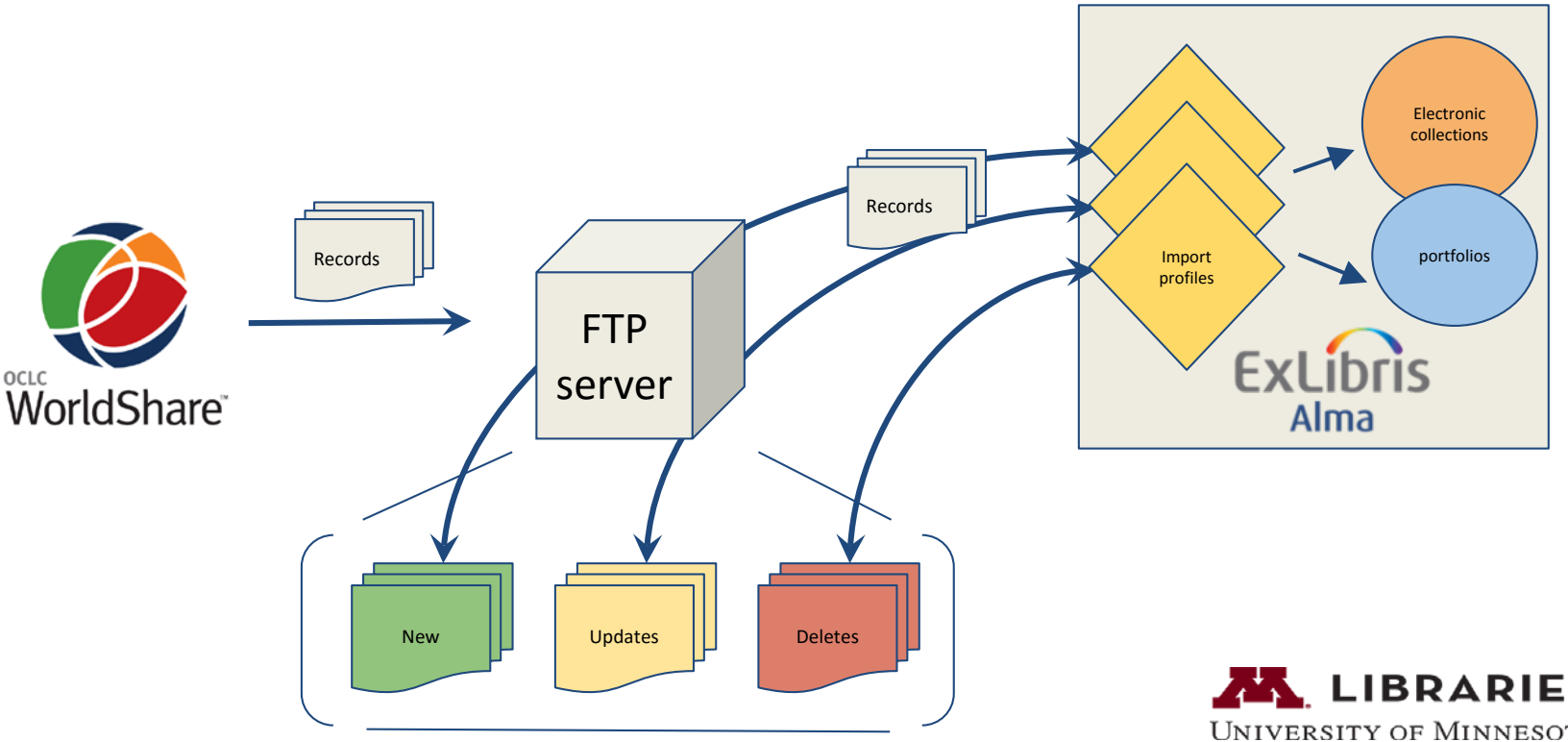
LIBRARIES
UNIVERSITY OF MINNESOTA

How WCM record delivery works

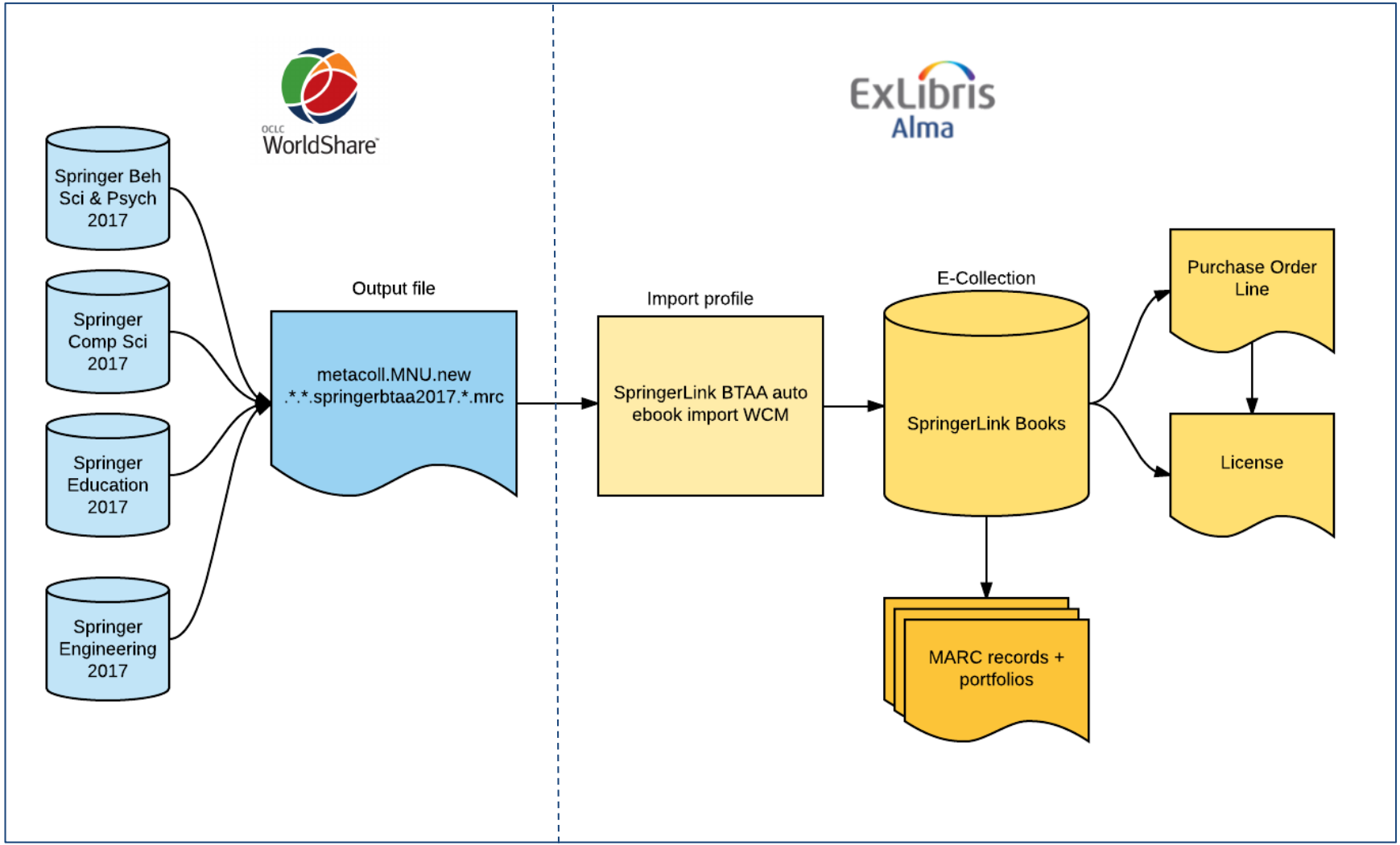


LIBRARIES
UNIVERSITY OF MINNESOTA

How WCM record import works in Alma



LIBRARIES
UNIVERSITY OF MINNESOTA


OCLC WorldShare
ExLibris Alma
Springer Beh Sci & Psych 2017
Springer Comp Sci 2017
Springer Education 2017
Springer Engineering 2017
Output file
metacoll.MNU.new .*.*.springerbtaa2017.*.mrc
Import profile
SpringerLink BTAA auto ebook import WCM
E-Collection
SpringerLink Books
Purchase Order Line
License
MARC records + portfolios

Getting set up

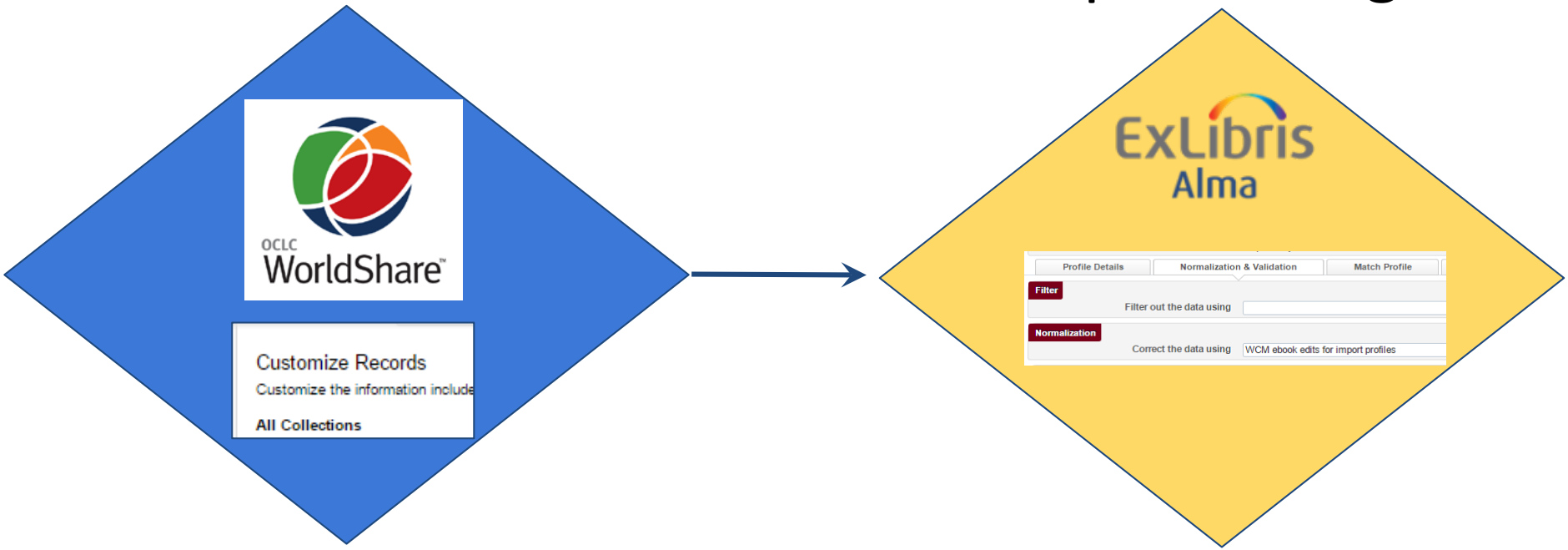
Worldshare Collection Manager:

- Customizes MARC records based on institution-wide and collection-specific settings.
- Produces files of new, updated, and deleted records on a schedule of the library's choosing.

Alma import profiles:

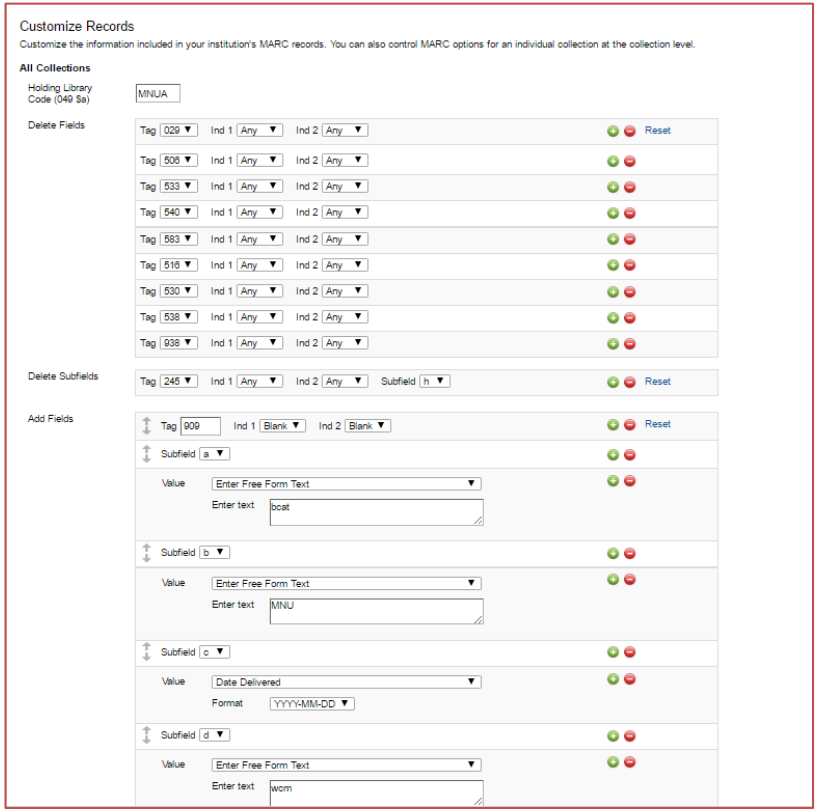
- Fetch files on schedule from an FTP server based on filename pattern matches.
- Customize records via normalization process.
- Apply merge rules so the appropriate fields are added or replaced upon match.
- Create and update (or delete) inventory.

WCM to Alma: New bib record processing



LIBRARIES
UNIVERSITY OF MINNESOTA

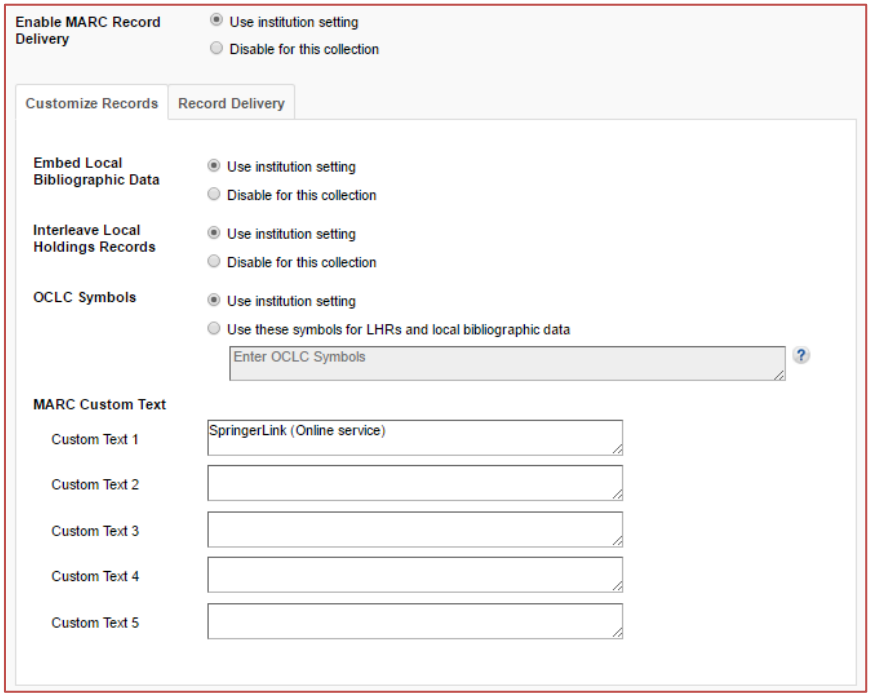
WCM MARC record customization settings

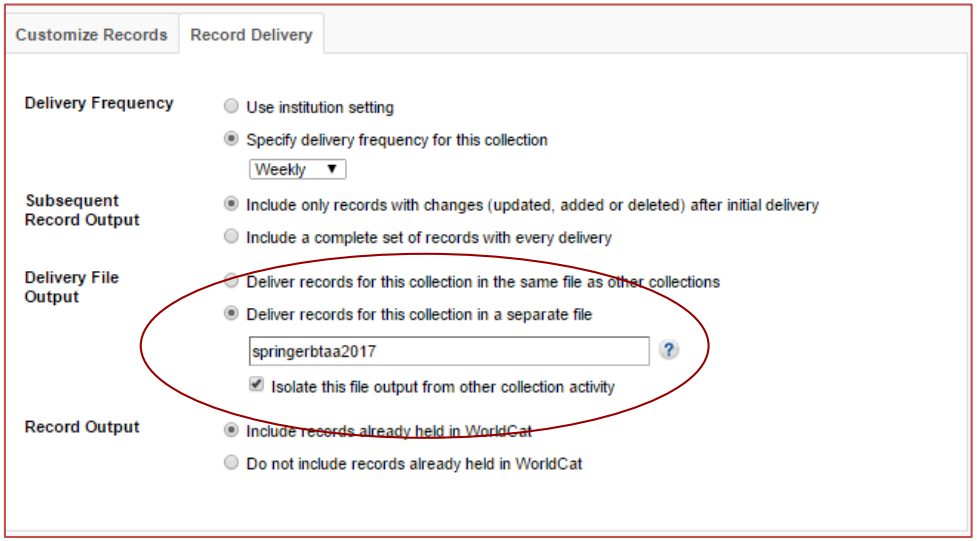


- Work in tandem with Alma import profile normalization process.
- Can do some things Alma normalization can't (e.g., dynamically generate the current date).
- Other edits better handled by Alma normalization, which offers more actions and more sophisticated conditionals.

LIBRARIES
UNIVERSITY OF MINNESOTA

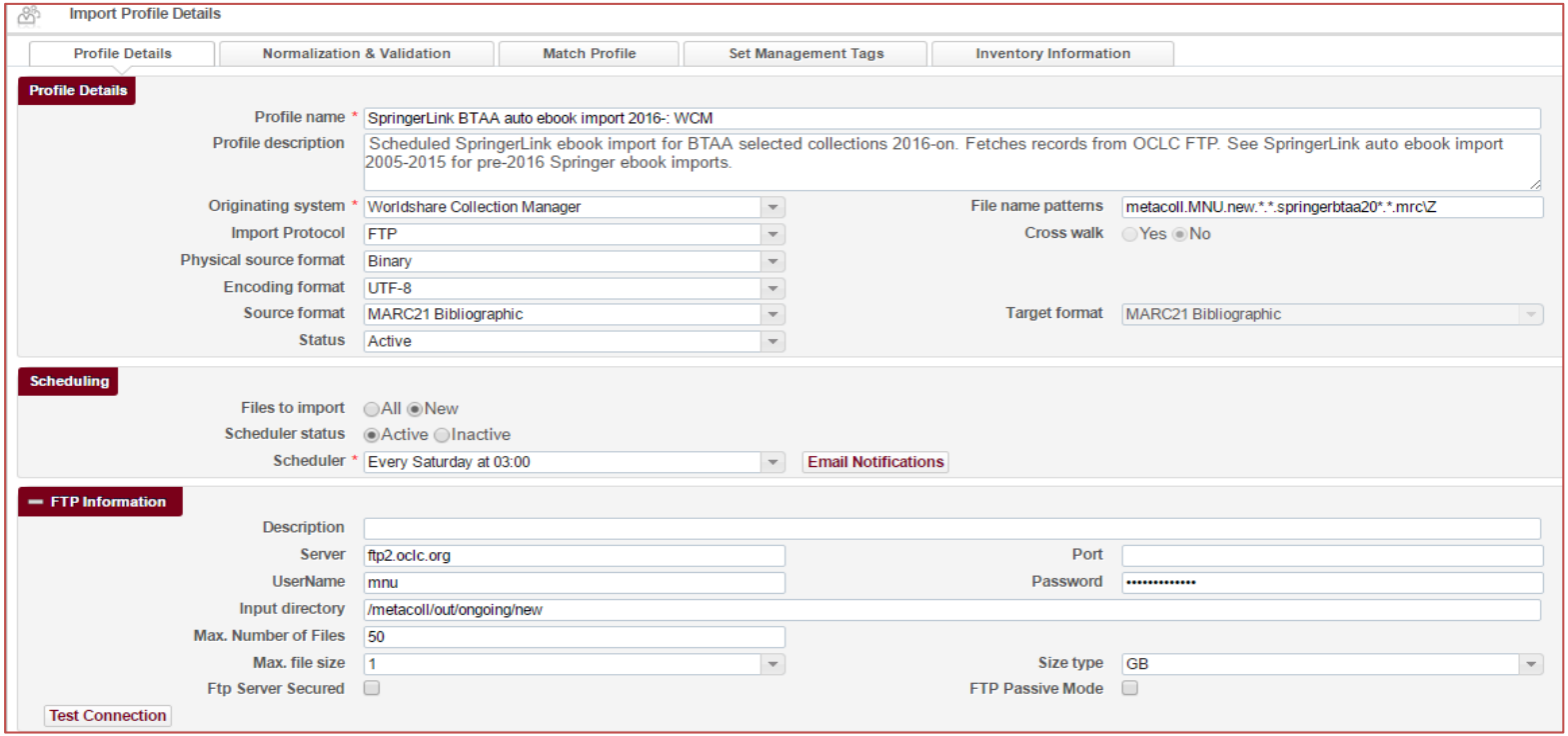
Collection-level MARC delivery configuration





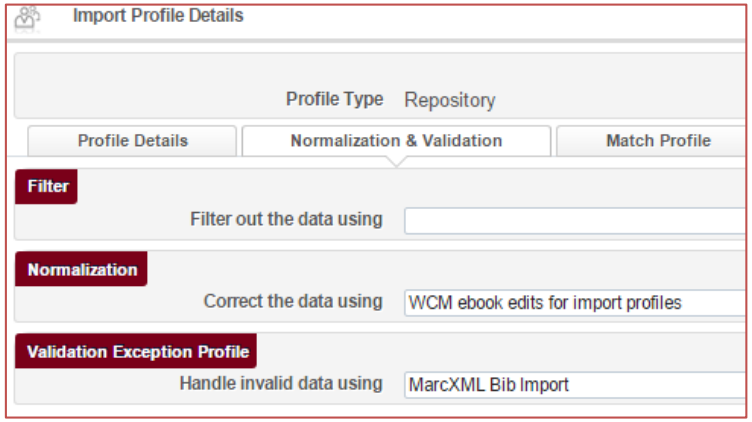
LIBRARIES
UNIVERSITY OF MINNESOTA

Import profiles: New records

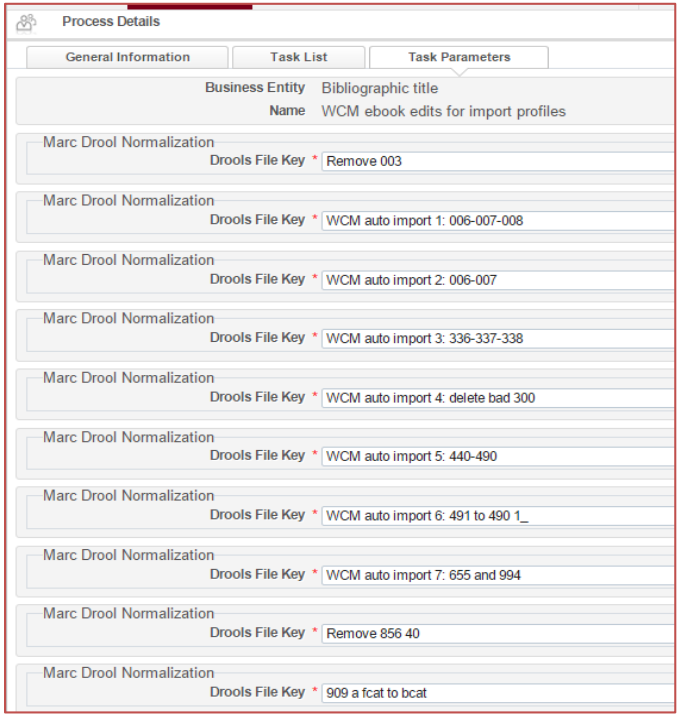


LIBRARIES
UNIVERSITY OF MINNESOTA

Import profiles: New records




```
rule "Ebook import merge"
#existing record is preferred record
when
merge
then
add MARC."909"
add MARC."930"
add MARC."931"
end
```



LIBRARIES
UNIVERSITY OF MINNESOTA

To delete or not to delete...

- UMN decision: tread carefully with delete files, especially for collections with perpetual access.
- WCM sometimes supplies delete records for valid content, not just for content removals (e.g., if linked record in WCM has changed).
- WCM has a problematic track record with delete/replace timing (recently improved).[1]
- Frequent delete/replace actions break Primo permalinks.
- Trade-off: we miss some record updates.

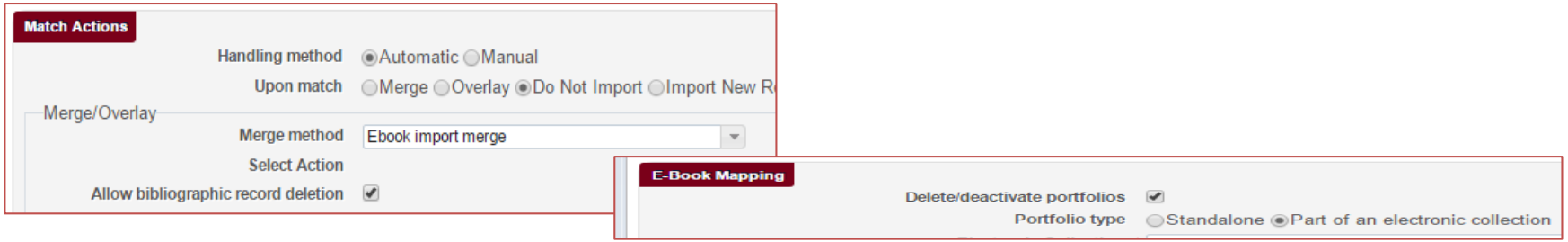
...but for collections where content frequently comes and goes, delete files are **not** optional.

[1]Kristin E. Martin and Christie Thomas, "OCLC Collection Manager for Electronic Resource Record Delivery and Holdings Updates: The Chicago Experience." ALCTS CRS Committee on Holdings Information Forum, ALA Midwinter 2017.

LIBRARIES
UNIVERSITY OF MINNESOTA

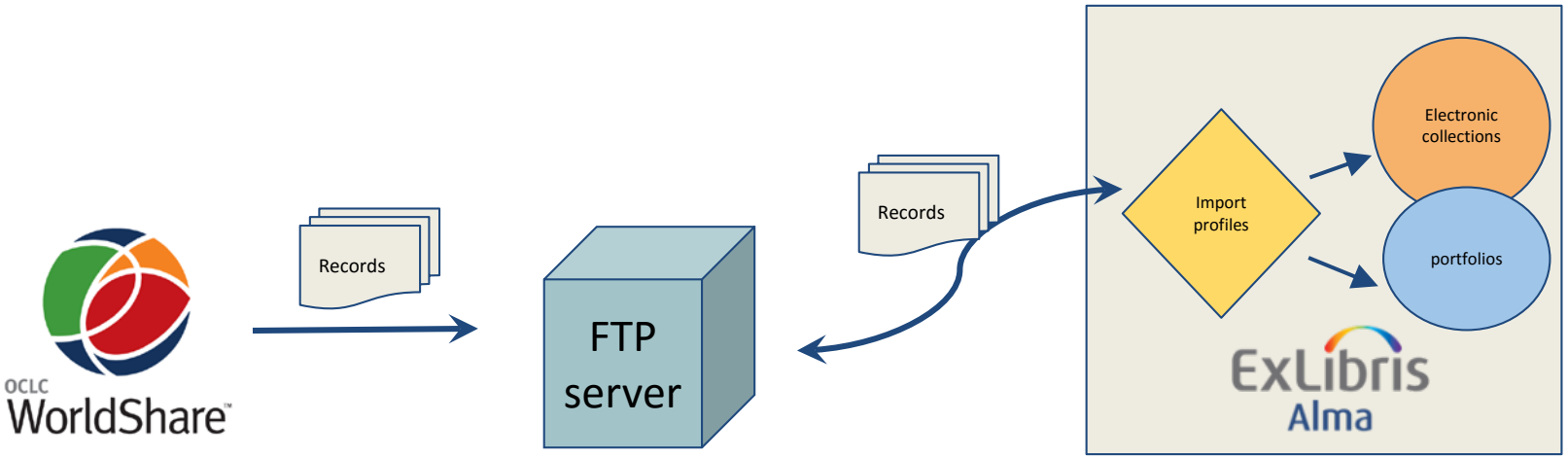
Import profiles: Delete records

- Scheduled to run less frequently than new/update profiles (generally monthly).
- FTP: finds “deletes” filenames in /deletes directory.
- Matches on OCLC ID
- If match is found and inventory exists in linked collection, deletes inventory.
- If no other inventory is linked to bib record, record is also deleted.



LIBRARIES
UNIVERSITY OF MINNESOTA

Automating record analysis



LIBRARIES
UNIVERSITY OF MINNESOTA

Automating record analysis

New 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

Update records

- Filter with a script before import
  - URL changes handled manually
  - OCN changes checked manually after automatic process
  - Everything else automatically processed
- "Access exists, do no harm"
  - Can afford to wait to do analysis before updating the description

New records

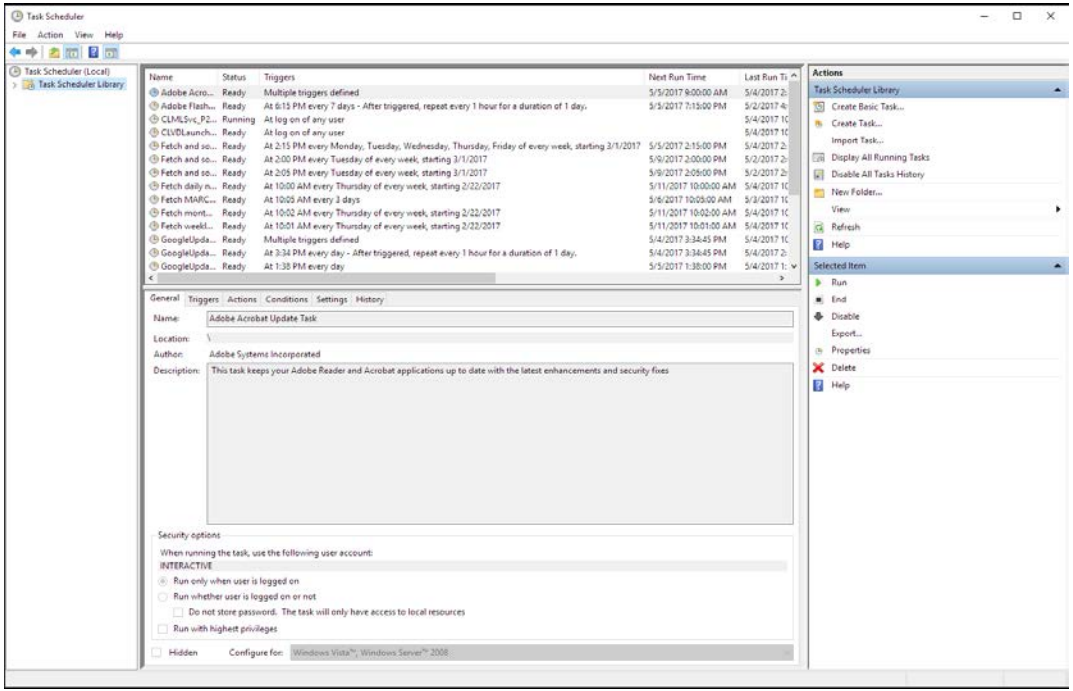
Post-ingest processing

Output: reports

Automating record analysis

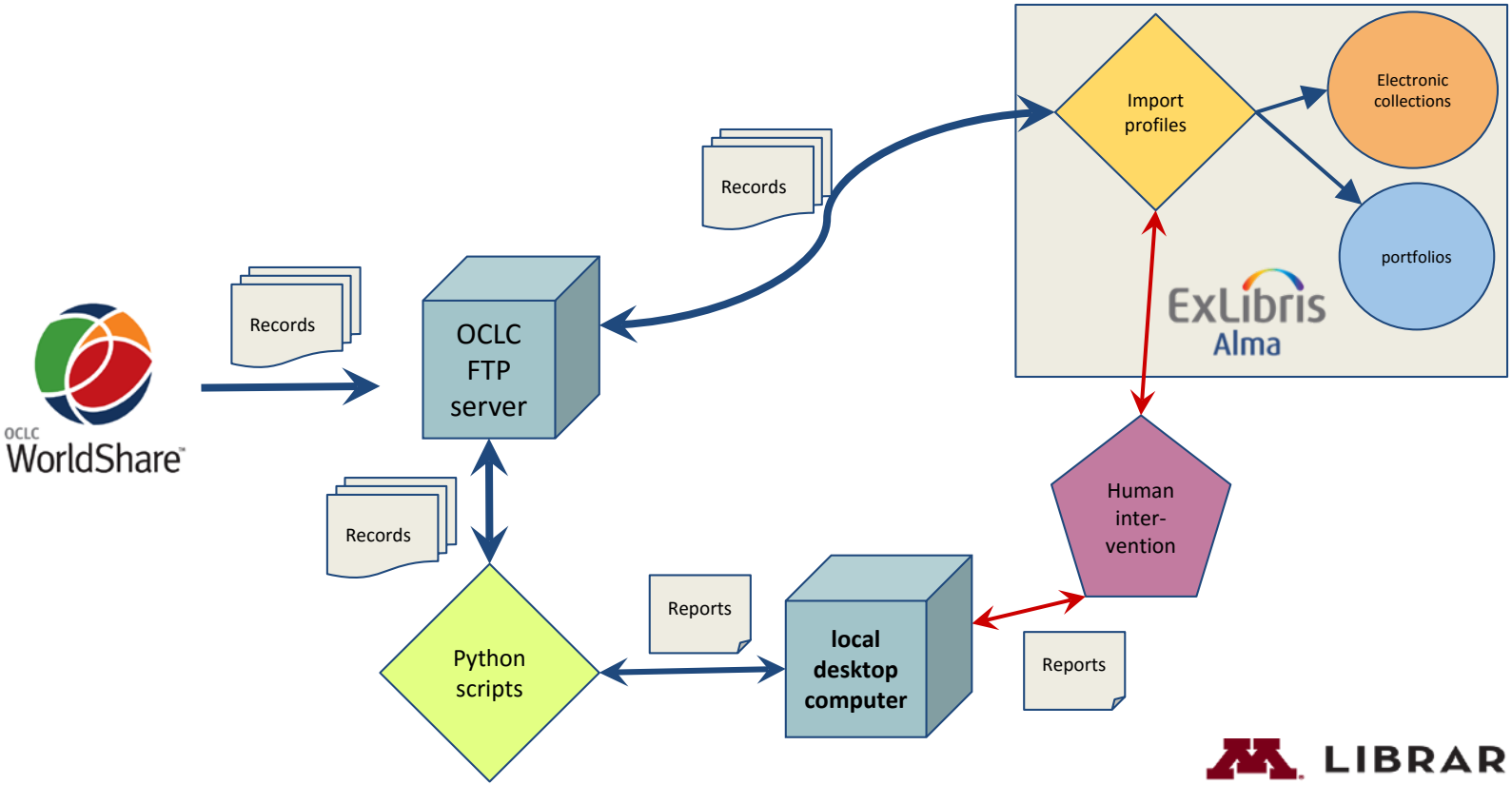
- Use Windows Task Scheduler to automatically run your scripts on a daily or weekly schedule.

- The script fetches the files from the FTP server and evaluates them.



LIBRARIES
UNIVERSITY OF MINNESOTA

Automating new record flagging



LIBRARIES
UNIVERSITY OF MINNESOTA

What are the Python scripts evaluating?

New record evaluation happens **post-ingest**.

Quality indicators:

- Potential multi-volume sets
- Language of cataloging
- Does the record describe an electronic or physical resource
- RDA
- Subject headings and call numbers

“Markers” or “flags”

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**OCN**	**ELvl**	**040b**	**006pos0**	**006pos6**	**006pos9**	**007**	**008**	**300**	**337**	**338**	**050**	**650 or MeSH**
2	ocn975046185	M	eng	m	\	d	cr	o	1 online resource	computer	online resource	050	LCSH
3	ocn976433932	I	eng	m	o	d	cr	o	1 online resource	computer	online resource	050	LCSH
4	ocn976433933	I	eng	m	o	d	cr	o	1 online resource	computer	online resource	050	LCSH
5	ocn978248508	I	eng	m	o	d	cr	o	1 online resource	computer	online resource	050	LCSH
6	ocn978248608	I	eng	m	o	d	cr	o	1 online resource	computer	online resource	050	LCSH
7	ocn978275325	I	eng	m	o	d	cr	o	1 online resource	computer	online resource	050	LCSH

Aggregate information on the markers

	A	B	C	D
1	**Number of records**	**118**		
2	Number of duplicate OCNs	0		
3	Number of unique OCNs	118		
4	Number of records with volume in description	13		
5				
6	**001 OCN**	**count**	**is duplicate**	**has volume**
7	47009084	1		
8	48139193	1		
9	52752095	1		
10	53840958	1		
11	54009499	1		
12	54009538	1		
13	56190233	1		volume
14	60014621	1		volume
15	60325512	1		volume
16	62081006	1		
17	62269072	1		
18	62893009	1		volume
19	71842473	1		
20	75189194	1		
21	75192765	1		
22	75197361	1		

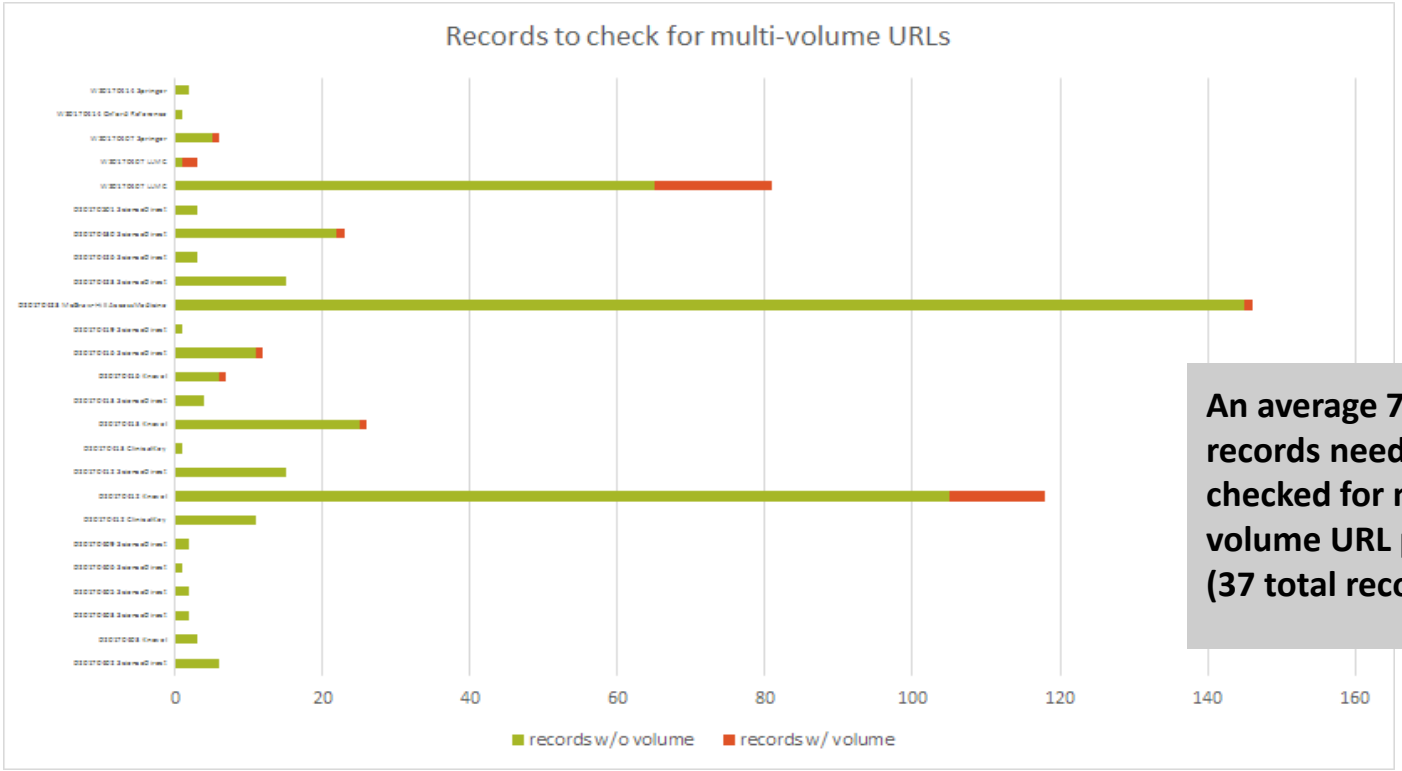
LIBRARIES
UNIVERSITY OF MINNESOTA

Aggregate information on the markers

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	OCN	EnLvl	cat_lang	comp_file_elec_res	online_006	document	categ_e_res_remote	online_008	item_desc	rda_media_type	rda_carrier_type	LC_call_no	LCSH or MeSH	total	percent	non-eng lang?	print record?	rda record?	has class?
2	61482004	I	1	1	1	1	1	1	0	1	1	1	1	10	90.9%				
3	70661167	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
4	82144410	4	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
5	85841308	I	1	1	1	1	1	1	0	1	1	1	1	10	90.9%				
6	ocn190760697	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
7	ocn313470822	K	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
8	ocn314115049	4	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
9	ocn642461781	M	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
10	ocn649473428	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
11	ocn681348027	K	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
12	ocn682112840	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
13	ocn701908647	K	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
14	ocn828724453	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
15	ocn833851240	M	1	1	1	1	1	1	1	1	1	0	1	10	90.9%				
16	ocn864991702	M	0	1	1	1	1	1	0	0	0	1	1	7	63.6%	non-eng lang?		rda record?	
17	ocn867781717	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
18	ocn867789579	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
19	ocn868020090	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
20	ocn868027580	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
21	ocn868155805	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
22	ocn868227607	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				
23	ocn868698338	I	1	1	1	1	1	1	1	1	1	1	1	11	100.0%				

LIBRARIES
UNIVERSITY OF MINNESOTA

April 2017 new records

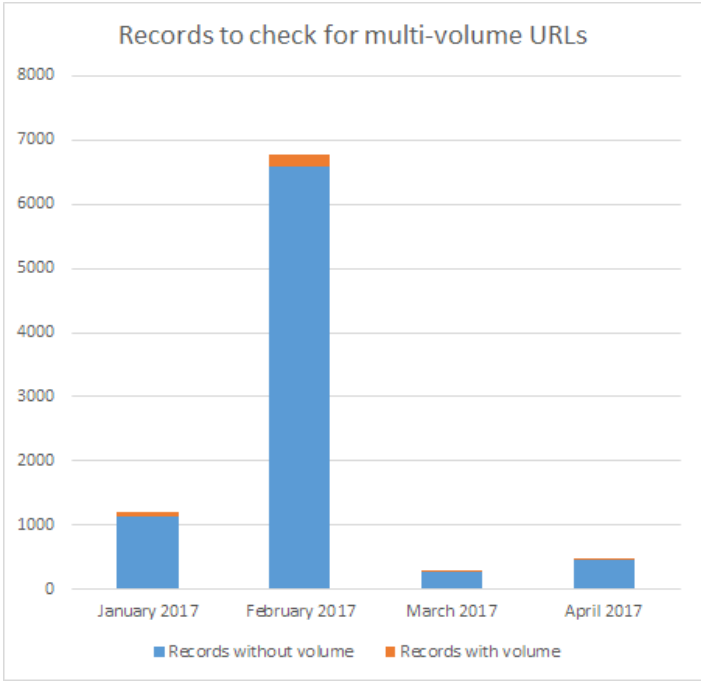


An average 7.5% of records needed to be checked for multi-volume URL problems (37 total records).

LIBRARIES
UNIVERSITY OF MINNESOTA

January 2017 - April 2017 new records

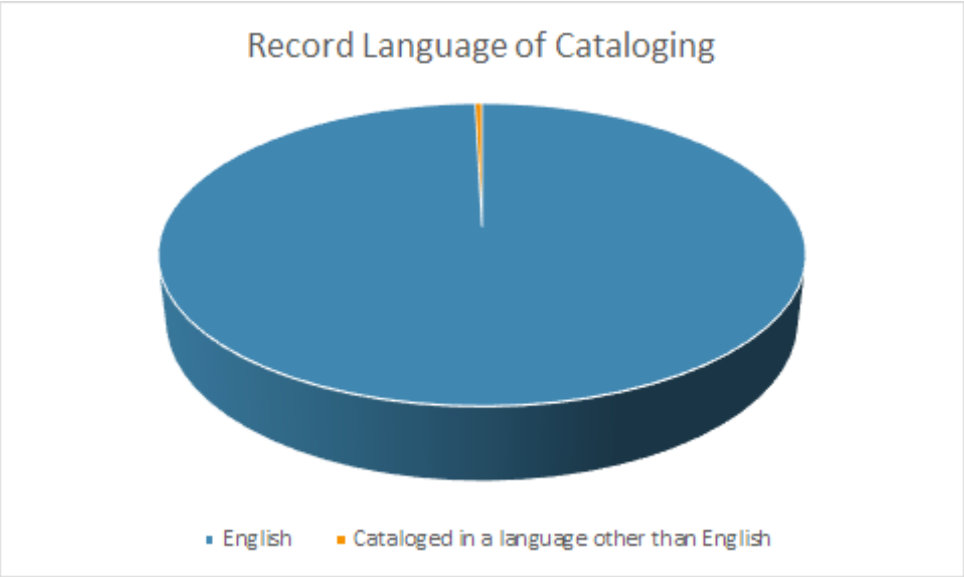
Month	Sum of record count	Sum of records w/o volume	Sum of records w/ volume	Sum of % to be checked for vol	Sum of non-eng language of cataloging	Sum of print record count
January 2017	1202	1142	60	5.0%	1	2
February 2017	6764	6597	167	2.5%	2	3
March 2017	289	265	24	8.3%	3	3
April 2017	491	454	37	7.5%	2	0
Grand Total	**8746**	**8458**	**288**	**3.3%**	**8**	**8**



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

LIBRARIES
UNIVERSITY OF MINNESOTA

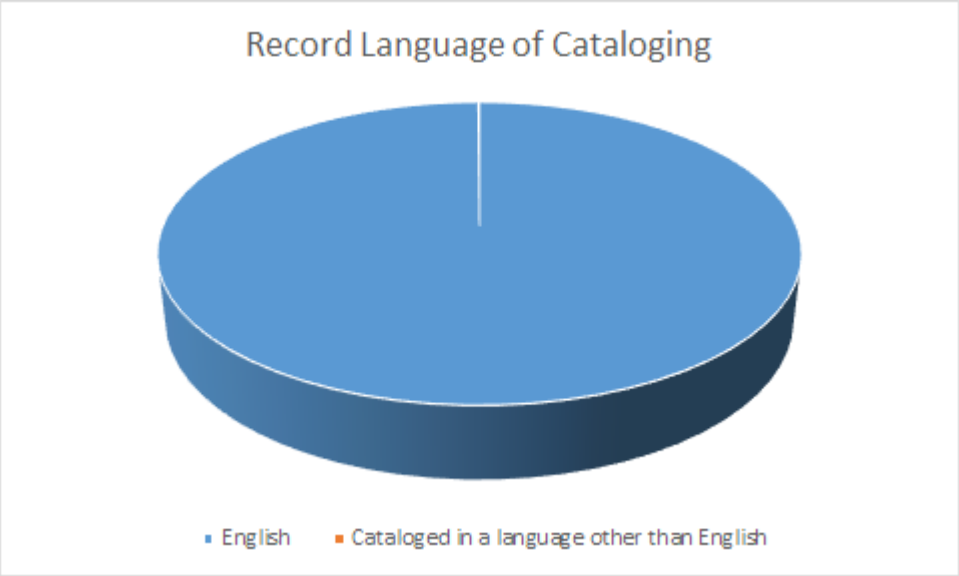
April 2017 new records



2 out of 494 records needed to be examined for language of cataloging

LIBRARIES
UNIVERSITY OF MINNESOTA

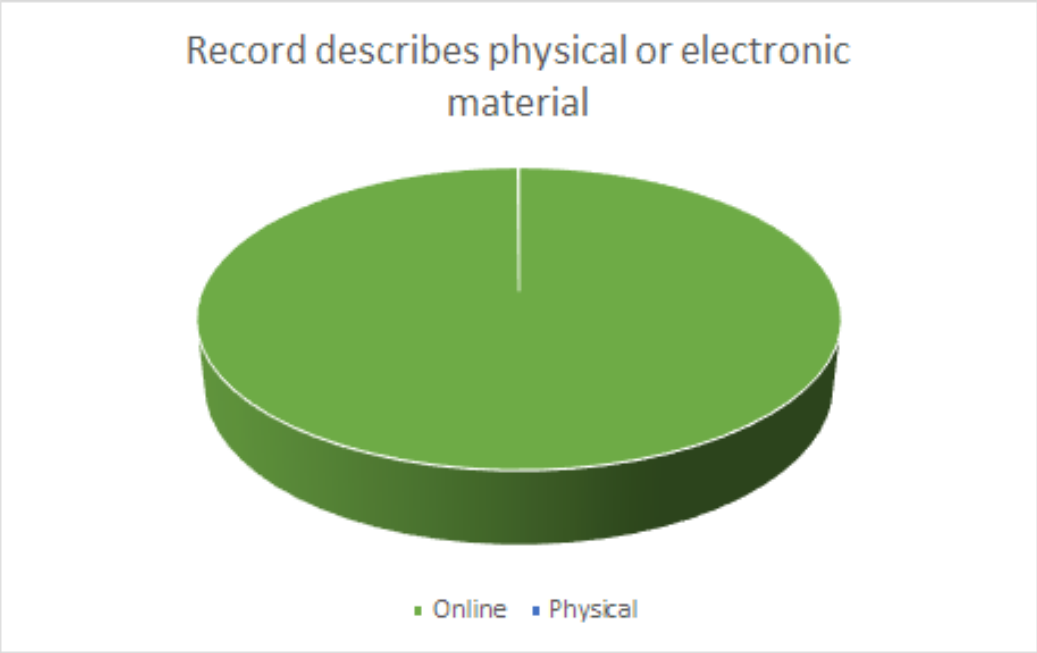
Jan-April 2017 new records



8 out of 8746 records needed to be examined for language of cataloging (~0.09%)

LIBRARIES
UNIVERSITY OF MINNESOTA

Jan-April 2017 new records



8 out of 8746 records needed to be examined to ensure they described electronic, not physical resources (~0.09%)

(There were zero records in April 2017 requiring this check.)

LIBRARIES
UNIVERSITY OF MINNESOTA

Contributing to record improvement

- Records from WCM are only added to Alma if they have an OCN
- This creates gaps in coverage
- Libraries can add or replace OCNs in WCM
  - Time consuming
  - Opportunity to link to best quality record
  - Ensures that if you update a master record in OCLC, your work will be reflected in your local system

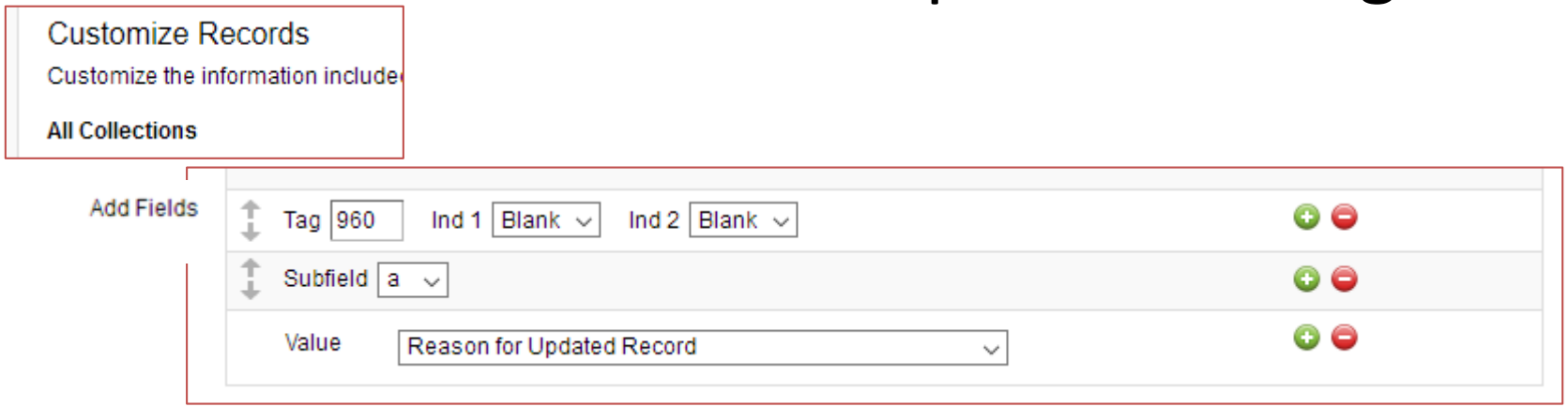
Updated records

Pre-ingest processing

Output: MARC files

LIBRARIES
UNIVERSITY OF MINNESOTA

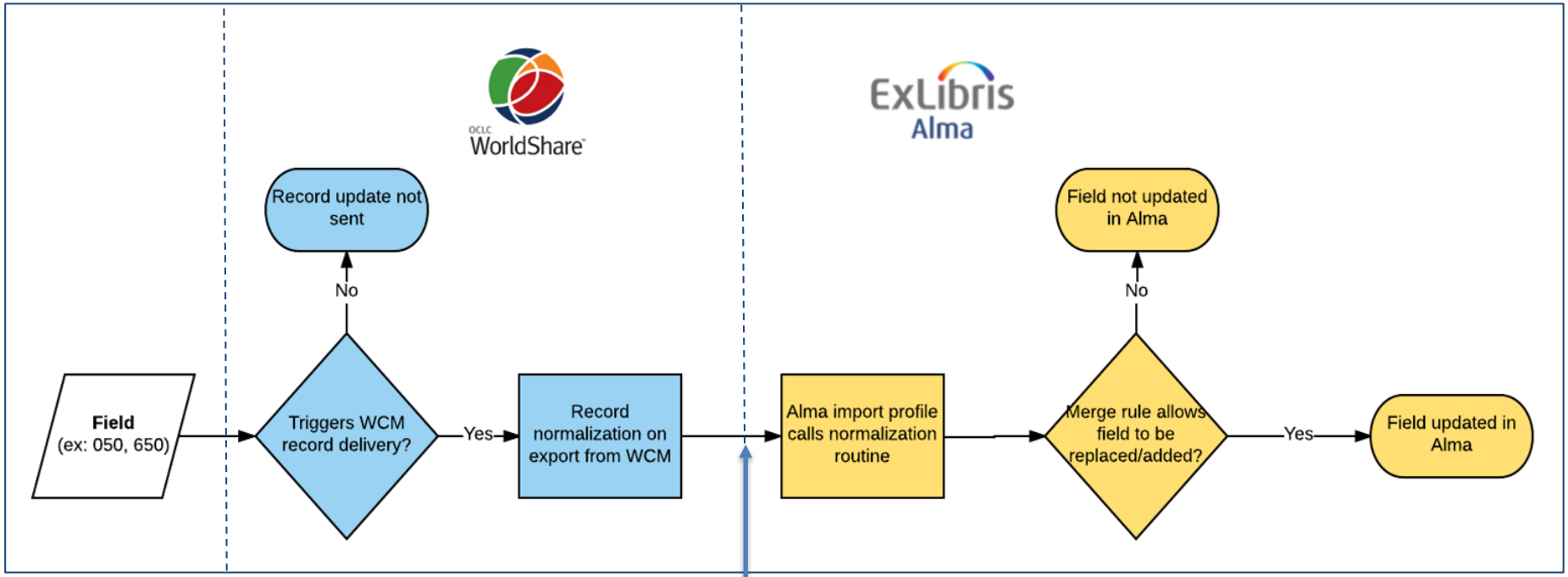
WCM MARC record updates settings



- Can configure which record changes will trigger delivery of an updated record
- Can configure output to include a field with the reason for the update (we use the **960 field)**
- Use this 960 field to sort files for processing

LIBRARIES
UNIVERSITY OF MINNESOTA

Update delivery configuration



python™

LIBRARIES
UNIVERSITY OF MINNESOTA

What are the Python scripts evaluating?

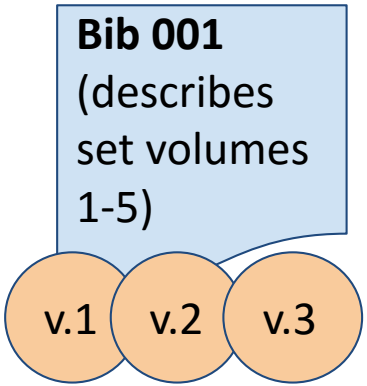
Updated record ingest filtering happens **pre-ingest**:

- Reason for update (field 960)
  - **OCLC number change**
    - Processed automatically but checked manually
    - Script output is a file of MARC records
  - **KB URL change**
    - Handled manually
    - Script output is a file of MARC records
  - **[Everything else - descriptive changes]**
    - Handled automatically
    - Script output is a file of MARC records to place on an FTP server for Alma to pickup

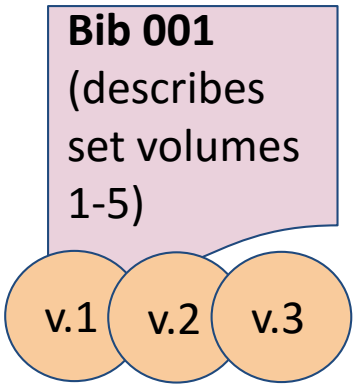
Service quality: Multi-volume records

- For multi-volume records, WCM often mis-manages these URLs.
- For example:

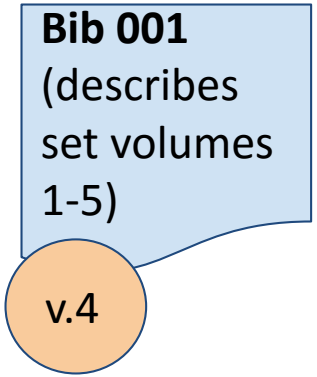
Initial record load



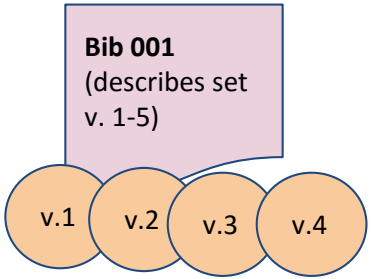
In Alma:



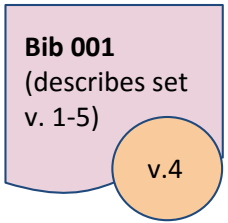
Record update (URL change)



EXPECTED

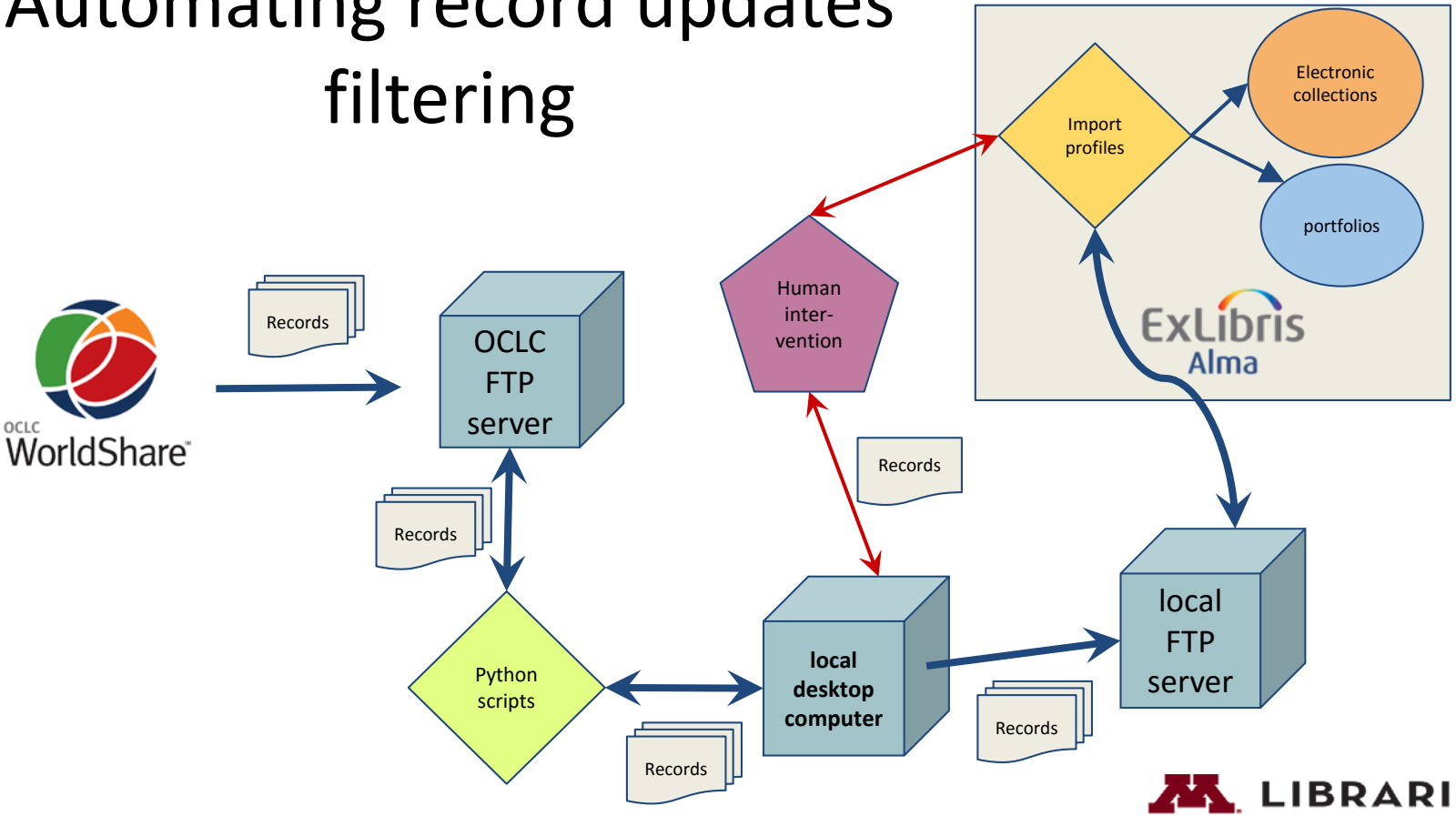


ACTUAL



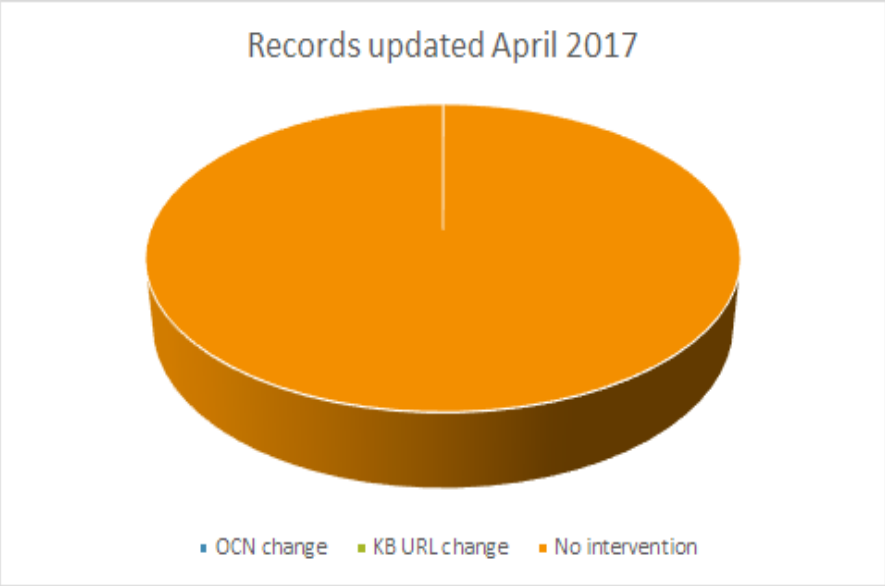
LIBRARIES
UNIVERSITY OF MINNESOTA

Automating record updates filtering



LIBRARIES
UNIVERSITY OF MINNESOTA

April 2017 record updates

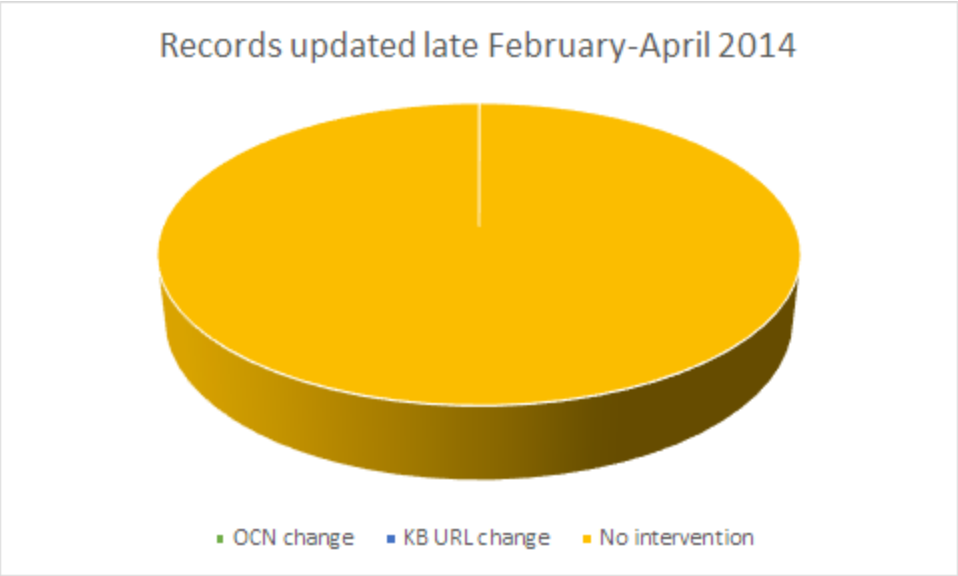


In one month, there were only:

- 14 records with OCN changes to check (0.01%)
- 33 records with URL changes to manually handle (0.02%)
- 144914 record updates needing no intervention (99.98%)

LIBRARIES
UNIVERSITY OF MINNESOTA

Late February-April 2017 record updates



In about two months, there were:

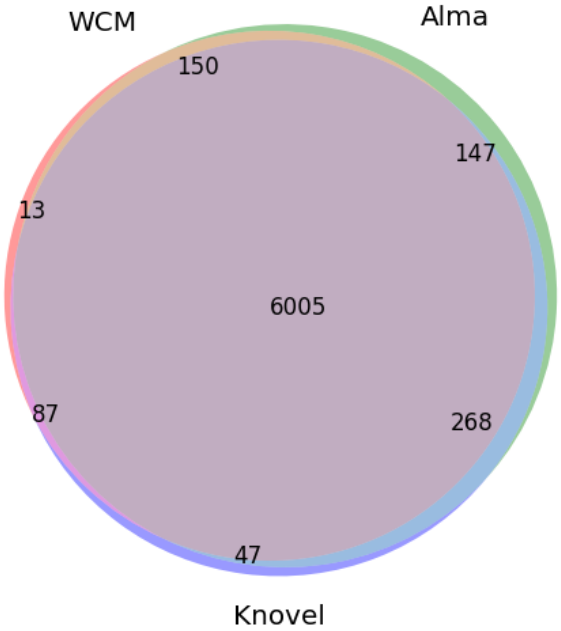
- 28 records with OCN changes to check (0.008%)
- 124 records with URL changes to manually handle (0.035%)
- 354435 record updates needing no intervention (99.957%)

LIBRARIES
UNIVERSITY OF MINNESOTA

Next step: Deletes!

Downsides to not running deletes...

Things get out of sync



LIBRARIES
UNIVERSITY OF MINNESOTA

Scheduling: synchronize your watches

Both WCM record delivery and Alma import profiles are **scheduled**.

You need to make sure these schedules coordinate with each other.

LIBRARIES
UNIVERSITY OF MINNESOTA

Conclusion

- We don’t have all the answers.
- Journey to manageable semi-automated processes has been long (and isn’t over).
- Complexity increases as more collections are managed this way.
- Possibly less worthwhile over time as bibliographic metadata in the Community Zone improves...but we’re not there yet.

Thank You!

Stacie Traill
trail001@umn.edu
Kelly Thompson
kjthomps@umn.edu

LIBRARIES
UNIVERSITY OF MINNESOTA