

Pre-Processing E-Resource MARC Data for Batch Load:

Challenges and Solutions

Kristina Spurgin

Library Data Strategist

University of North Carolina at Chapel Hill

kspurgin@email.unc.edu - @kspurgin

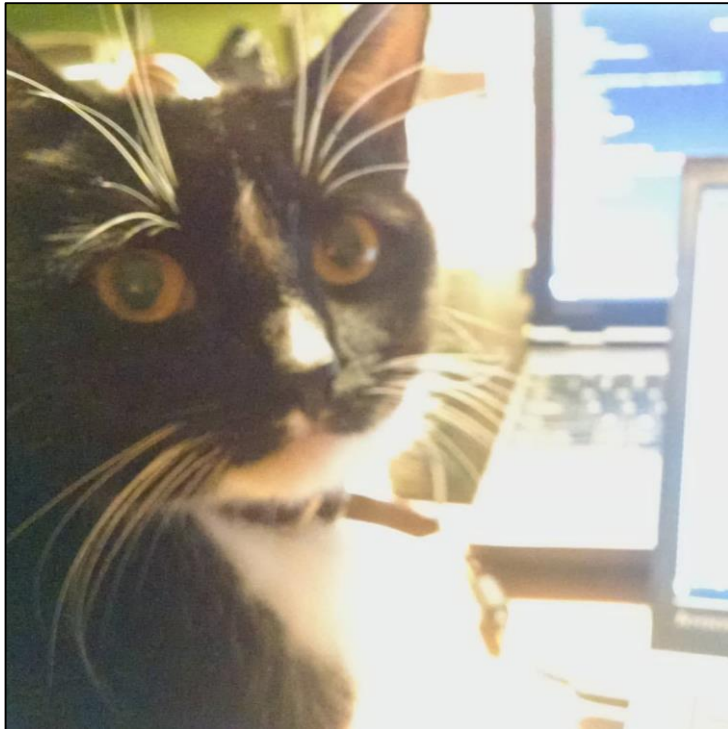


A map of this presentation

A bit of context – my institution and my role in it

OCLC WorldShare Collection Manager (WCM) MARC: brief overview

The WCM MARC-related mess we got into



Another twist: (hint: authority control vendor-related)

My solution: script + new load profile + workflow

Limitations/future work

Questions/discussion

A bit of context

my institution and my role in it



“It’s complicated...”



- University Library, UNC Chapel Hill
- Large institution, ARL member
- 6,877,981 bibs in OPAC
- 3 administrative/accounting units
- +/- 30 branches and specialized collection locations
- 1038 location codes
- Part of Triangle Research Libraries Network, sharing:
 - Index/online catalog
 - Physical storage space
 - Some MARC records
 - Some acquisitions
- 2 staff members with III load profile training
- 2017-03-14: Migrated from Millennium to Sierra

My job: (LIBRARY DATA) Strategist

- Maintain overview of library data...
 - Formats
 - Sources
 - Systems
- ...where we...
 - Create/maintain the data
 - Contribute data
 - Use the data (or could use the data)

My job: Library (DATA STRATEGIST)

- Understand connections—existing and potential
- Lead a team* with domain expertise & tech skills to:
 - Leverage library data **at scale** from different sources...
 - To answer questions and solve problems (usually for our colleagues) across the Libraries** ...
 - AND improve resource discoverability and other user services.
- Consult on projects across the Libraries to increase chances of:
 - Effective use/reuse of currently available data
 - Making data/metadata/system decisions that support future use/reuse of data

* Library Data Strategy & Services (LDSS), and, psst, we have an open Metadata Librarian position right now

** University Libraries (my home), Health Science Library, and Law library

Still my job: E-resources Cataloger

- My section* handles cataloging for e-resource titles representing over half our annual monographs budget
 - 1028 collections with unsuppressed MARC records in the catalog
 - 1,263,251 ebook records
 - 43,162 streaming video records
 - 25,869 streaming audio records
 - 17,575 serial/integrating resource records
 - 66 MARC sources+workflows
- This work still makes sense in the Library Data Strategy & Services section

* Currently 3 FTE total (+1 soon)



OCLC WorldShare Collection Manager (WCM) MARC

brief overview

WCM is a knowledge base... Select (track) your collections...

Metadata Admin

Record Manager

Collection Manager

Collection

My Selected Collections

Close All Tabs

Search Collections with "", My Selected Collections

Activity History

Approve Changes to Global Collections

View Recent Collections

Create a Collection

Settings

My Files

Search Collections with "", My Selected Collections

Collection Actions

Filter by

Results 1 - 100 of 514 Rows

<input type="checkbox"/>	Collection	Type / Attributes	Provider	Updated	Selections	Sharing
<input type="checkbox"/>	2012 Cambridge Journals Full Package Standard UK Collection ID: CUP.cjo	Knowledge base collection • Customizable	Cambridge University Press	03/24/2017 03:57 AM	<input checked="" type="checkbox"/> Selected by my library Titles: 129/318 Deselect	
<input type="checkbox"/>	Academic Electronic Press Collection ID: openly.jsCate.aepress	Knowledge base collection • Customizable	Academic Electronic Press	04/01/2017 04:24 PM	<input checked="" type="checkbox"/> Selected by my library Titles: 0/16 Deselect	
<input type="checkbox"/>	ACM Conference Series Collection ID: ACM.seriesPub	Knowledge base collection • Customizable	Association for Computing Machinery	03/14/2017 12:36 PM	<input checked="" type="checkbox"/> Selected by my library Titles: 0/999 Deselect	
<input type="checkbox"/>	ACS Symposium Series eBooks - 2011 Front Files Collection ID: ACS.symp2011	Knowledge base collection • Customizable	American Chemical Society	03/17/2017 10:52 AM	<input checked="" type="checkbox"/> Selected by my library Titles: 30/30 Deselect	
<input type="checkbox"/>	ACS Symposium Series eBooks - 2012 Front Files Collection ID: ACS.symp2012	Knowledge base collection • Customizable	American Chemical Society	03/17/2017 10:52 AM	<input checked="" type="checkbox"/> Selected by my library Titles: 30/30 Deselect	
<input type="checkbox"/>	ACS Symposium Series eBooks - 2013 Front Files Collection ID: ACS.symp2013	Knowledge base collection • Customizable	American Chemical Society	03/17/2017 10:52 AM	<input checked="" type="checkbox"/> Selected by my library Titles: 35/35 Deselect	
<input type="checkbox"/>	ACS Symposium Series eBooks - 2014 Front Files Collection ID: ACS.symp2014	Knowledge base collection • Customizable	American Chemical Society	03/17/2017 10:52 AM	<input checked="" type="checkbox"/> Selected by my library Titles: 30/30 Deselect	

A collection lists the titles in that collection...

Record Manager

Collection Manager

Collection

My Selected Collections

Search

Close All Tabs

Search Collections with "", My Selected Collections

Collection: ACS Symposium Series eBooks - 2011 Front...

My Files

Collection: ACS Symposium Series eBooks - 2011 Front Files

Save **Contribute Changes to Global Collection**

- Properties
- Linking
- Holdings and MARC Records
- Sharing
- Titles**

Search within

My Selected Titles All Titles

- Titles** Titles available individually Titles with current issues
- Format** E-book/Proceeding Audio
- Image Video Other
- Journal Coverage** Full Text Print Selected Full Text Abstracts Indexed
- [Reset](#)

Results 1 - 30 of 30 Rows

Title	Content Availability	Identifiers	Selections
100+ Years of Plastics: Leo Baekeland and Beyond	E-book/Proceeding: 2011-01-01	OCN: 763178021 ISBN: 9780841226111 Website	<input checked="" type="checkbox"/> Held by my library
Advances in Cereal Science: Implications to Food Processing and Health Promotion	E-book/Proceeding: 2011-01-01	OCN: 768826883 ISBN: 9780841226364 Website	<input checked="" type="checkbox"/> Held by my library
Amphiphiles: Molecular Assembly and Applications	E-book/Proceeding: 2011-01-01	OCN: 760176585 ISBN: 9780841226500 Website	<input checked="" type="checkbox"/> Held by my library
Aquatic Redox Chemistry	E-book/Proceeding: 2011-01-01	OCN: 760176586 ISBN: 9780841226524 Website	<input checked="" type="checkbox"/> Held by my library
Biotechnology and Nanotechnology Risk Assessment: Minding and Managing the Potential Threats around Us	E-book/Proceeding: 2011-01-01	OCN: 763176276 ISBN: 9780841226609 Website	<input checked="" type="checkbox"/> Held by my library

There are a lot of settings at the collection level...

WorldShare[®] Need Help? [User support info](#) | Kristina Spurgin

Metadata Admin

Collection: ACS Symposium Series eBooks - 2011 Front Files

Save Contribute Changes to Global Collection Deselect Collection

- Properties
- Linking
- Holdings and MARC Records

Maintain WorldCat Holdings

- Use institution setting
- Disable for this collection

Data Sources

- All approved data sources
- Selected data sources only
- Manual upload only

ILL allowed

- Yes
- No

Instructions Choose...

Export to Google Scholar

- Use institution setting
- Disable for this collection

Enable MARC Record Delivery

- Use institution setting
- Disable for this collection

Institution settings: MARC: Customization

Settings

Save

WorldCat Discovery

MARC Records

Enable MARC Records Delivery Yes Your symbol is [NOCC](#). No

Customize Records WorldCat Updates Record Delivery Cataloging Partner Options

Customize Records

Customize the information included in your institution's MARC records. You can also control MARC options for an individual collection at the collection level.

All Collections

Holding Library Code (049 \$a)

Delete Fields

Tag	<input type="text" value="029"/>	Ind 1	<input type="text" value="Any"/>	Ind 2	<input type="text" value="Any"/>	<input type="button" value="+"/>	<input type="button" value="-"/>	Reset
Tag	<input type="text" value="938"/>	Ind 1	<input type="text" value="Any"/>	Ind 2	<input type="text" value="Any"/>	<input type="button" value="+"/>	<input type="button" value="-"/>	

Delete Subfields

Tag	<input type="text"/>	Ind 1	<input type="text" value="Any"/>	Ind 2	<input type="text" value="Any"/>	Subfield	<input type="text"/>	<input type="button" value="+"/>	<input type="button" value="-"/>	Reset
-----	----------------------	-------	----------------------------------	-------	----------------------------------	----------	----------------------	----------------------------------	----------------------------------	-----------------------

Add Fields

↑↓ Tag	<input type="text" value="773"/>	Ind 1	<input type="text" value="0"/>	Ind 2	<input type="text" value="Blank"/>	<input type="button" value="+"/>	<input type="button" value="-"/>	Reset		
↑↓ Subfield	<input type="text" value="t"/>	<input type="button" value="+"/>	<input type="button" value="-"/>							
Value	<input type="text" value="Collection Custom Text"/>							<input type="button" value="+"/>	<input type="button" value="-"/>	
Custom Text Number Field	<input type="text" value="1"/>									
↑↓ Tag	<input type="text" value="598"/>	Ind 1	<input type="text" value="Blank"/>	Ind 2	<input type="text" value="Blank"/>	<input type="button" value="+"/>	<input type="button" value="-"/>			
↑↓ Subfield	<input type="text" value="a"/>	<input type="button" value="+"/>	<input type="button" value="-"/>							
Value	<input type="text" value="Enter Free Form Text"/>							<input type="button" value="+"/>	<input type="button" value="-"/>	
Enter text	<input type="text" value="Reason for updated record:"/>									

Institution settings: MARC: What constitutes an update?

Save

Enable MARC Records Delivery

- Yes Your symbol is [NOC](#).
 No

Customize Records

WorldCat Updates

Record Delivery

Cataloging Partner Options

Updates

Select the option for WorldCat Updates.

- None
 Based on changes to WorldCat master records for my knowledge base and query collections
 Based on changes to WorldCat master records for my knowledge base, query, data sync collections and WorldCat holdings

Update Conditions

















Define criteria to determine when you would like to receive updated records based on changes to the WorldCat master record.

- Any change to the WorldCat master record
 Select options for specific changes to the WorldCat master record

Specify **Including** Conditions

Variable Fields

Deliver records if the following fields are added, updated, or deleted in the master record:

Tag	1	X	X	Ind 1	Any	Ind 2	Any	  Reset
Tag	2	4	X	Ind 1	Any	Ind 2	Any	 
Tag	4	9	0	Ind 1	Any	Ind 2	Any	 
Tag	6	X	X	Ind 1	Any	Ind 2	0	 
Tag	6	X	X	Ind 1	Any	Ind 2	2	 
Tag	7	0	X	Ind 1	Any	Ind 2	Any	 
Tag	7	1	X	Ind 1	Any	Ind 2	Any	 
Tag	7	3	X	Ind 1	Any	Ind 2	Any	 

Institution settings: MARC: Delivery

Settings

Save

Provider Settings

- ▶ Sharing
- ▶ WorldCat Discovery
- ▼ MARC Records

Enable MARC Records Delivery Yes Your symbol is [NOC](#). No

Customize Records WorldCat Updates **Record Delivery** Cataloging Partner Options

Record Delivery

Customize how often and under what conditions your records will be delivered.

Standard

Delivery Frequency ?

File Output

- Deliver my serials in a separate file (.ser)
- Deliver records for WorldCat updates in a separate file (.wcholdings)

Back at the COLLECTION level...

Collection: ACS Symposium Series eBooks - 2011 Front Files

Save

Contribute Changes to Global Collection ▼

Deselect Collection

Enable MARC Record Delivery

- Use institution setting
 Disable for this collection

Customize Records

Record Delivery

Embed Local Bibliographic Data

- Use institution setting
 Disable for this collection

Interleave Local Holdings Records

- Use institution setting
 Disable for this collection

OCLC Symbols

- Use institution setting
 Use these symbols for LHRs and local bibliographic data

Enter OCLC Symbols



MARC Custom Text

Custom Text 1

ACS symposium series (online collection). 2011

Custom Text 2

There are some MARC delivery settings for each collection...

Collection: ACS Symposium Series eBooks - 2011 Front Files

[Save](#) [Contribute Changes to Global Collection ▼](#) [Deselect Collection](#)

Scholar Use institution setting
 Disable for this collection

Enable MARC Record Delivery Use institution setting
 Disable for this collection

Customize Records **Record Delivery**

Delivery Frequency Use institution setting
 Specify delivery frequency for this collection
Daily ▼

Subsequent Record Output Include only records with changes (updated, added or deleted) after initial delivery
 Include a complete set of records with every delivery

Delivery File Output Deliver records for this collection in the same file as other collections
 Deliver records for this collection in a separate file
ACS ?
 Isolate this file output from other collection activity

Record Output Include records already held in WorldCat
 Do not include records already held in WorldCat

My Files > Downloads

Downloads

Filter by

Select

[Reset](#)

Results 1 - 10 of 1674

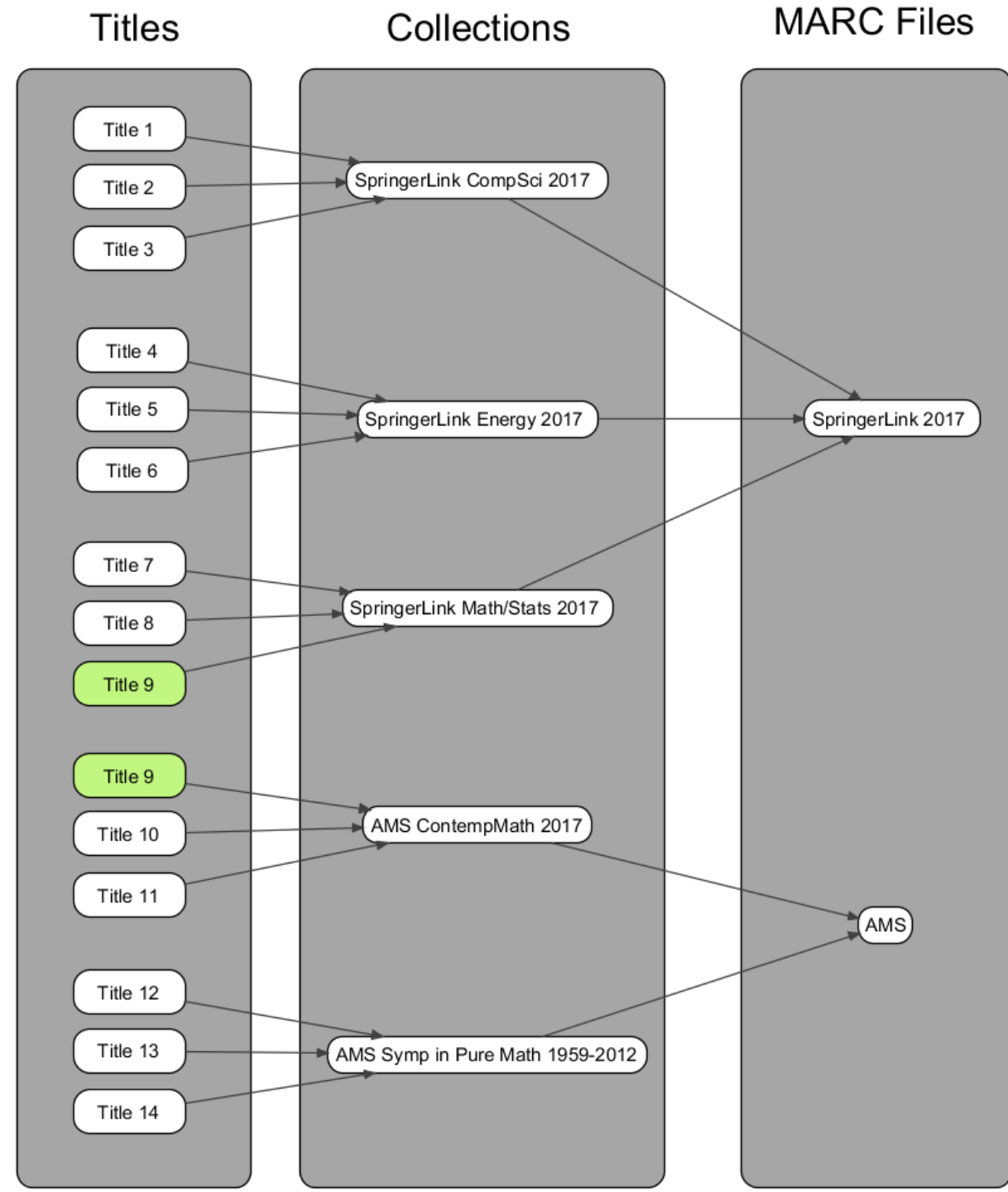
Rows 10 << 1 2 3 4 5 >>

File Name	File Type	Size	Post Date	Download Date	Action
metacoll.NOC.updates.D20170330.T183014.SPRolder.1.mrc	Records - Updated	196,255 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPR2016.1.mrc	Records - Updated	27,678 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.Video.1.mrc	Records - Updated	170 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SciDir.1.mrc	Records - Updated	12,001 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SerAsMono.1.mrc	Records - Updated	1,315 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPRgratis.1.mrc	Records - Updated	9,959 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPR2017.1.mrc	Records - Updated	8,912 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.DGprinceton.1.mrc	Records - Updated	14,780 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.RSC.1.mrc	Records - Updated	5,127 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPIE.1.mrc	Records - Updated	1,268 KB	03/31/2017		Download

Results 1 - 10 of 1674

Rows 10 << 1 2 3 4 5 >>

A little digression on Delivery File Output...





Our mess

WCM MARC-related

Our initial setup:

- Receive add/update/delete files
- Over 250 collections set up for MARC delivery
- Collections output to ~15 files for MARC delivery
- Some files being delivered monthly, others every 2 weeks



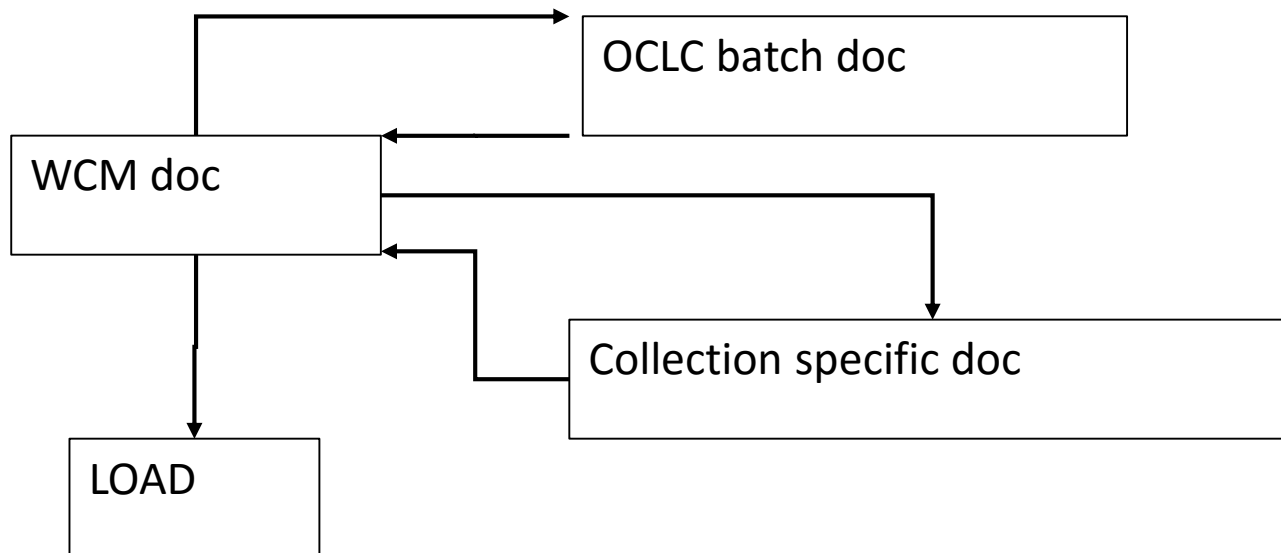
Our initial workflow:

- Use regular e-book batch load profile
- Some steps were the same for all WCM collections:
 - Identifying/retrieving files
 - Pre-processing files
 - Some editing steps:
 - Removal of “=598 Reason for updated record:”
 - Cleanup of OCNs
 - Addition of “wcm” suffix to OCNs
 - Routine cleanup done on all OCLC batch records



Our initial workflow:

- File/collection-specific steps
 - # of concurrent users
 - Perpetual access or not
 - Collection-specific ID suffix
- THEN...
 - Back to overall WCM workflow for load notes and procedure




And the complexity grew: What should be simple...

- Train staff to identify/download the correct/complete files...
 - For a given file/set (i.e. SciDir)...
 - For a given date


Downloads

Filter by [Reset](#)

Results 1 - 10 of 1674 

Rows [<<](#) [1](#) [2](#) [3](#) [4](#) [5](#) [>>](#)

File Name	File Type	Size	Post Date	Download Date	Action
metacoll.NOC.updates.D20170330.T183014.SPRolder.1.mrc	Records - Updated	196,255 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPR2016.1.mrc	Records - Updated	27,678 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.Video.1.mrc	Records - Updated	170 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SciDir.1.mrc	Records - Updated	12,001 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SerAsMono.1.mrc	Records - Updated	1,315 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPRgratis.1.mrc	Records - Updated	9,959 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPR2017.1.mrc	Records - Updated	8,912 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.DGprinceton.1.mrc	Records - Updated	14,780 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.RSC.1.mrc	Records - Updated	5,127 KB	03/31/2017		Download
metacoll.NOC.updates.D20170330.T183014.SPIE.1.mrc	Records - Updated	1,268 KB	03/31/2017		Download

Results 1 - 10 of 1674 

Rows [<<](#) [1](#) [2](#) [3](#) [4](#) [5](#) [>>](#)

And the complexity grew: Replacement: New + Delete

Ebook title in collection: Can cats and goats coexist?

Previous OCN association:

```
001 12345  
245 Engineering dictionary
```

Corrected OCN association:

```
001 98765  
245 Can cats and goats coexist?
```

In next record delivery after correction:

DELETE RECORD SENT

```
001 12345  
245 Engineering dictionary
```

NEW RECORD SENT

```
001 98765  
245 Can cats and goats coexist?
```


And the complexity grew: Replacement: Merge/update

Ebook title in collection: Tuxedo menagerie

OCN association: 123 in KB

Record merge in WorldCat!

OCN association: still 123 in KB

In next record delivery after correction:

WorldCat master record:

001 123
245 \$aTuxedo menagerie

Will this overlay work in your system?

WorldCat master record:

001 789
019 \$a456\$a123
245 \$aTuxedo menagerie

UPDATE RECORD SENT

At this point, I wrote the first iteration of a WCM pre-processing script.



And then things got worse...

2016-06-16

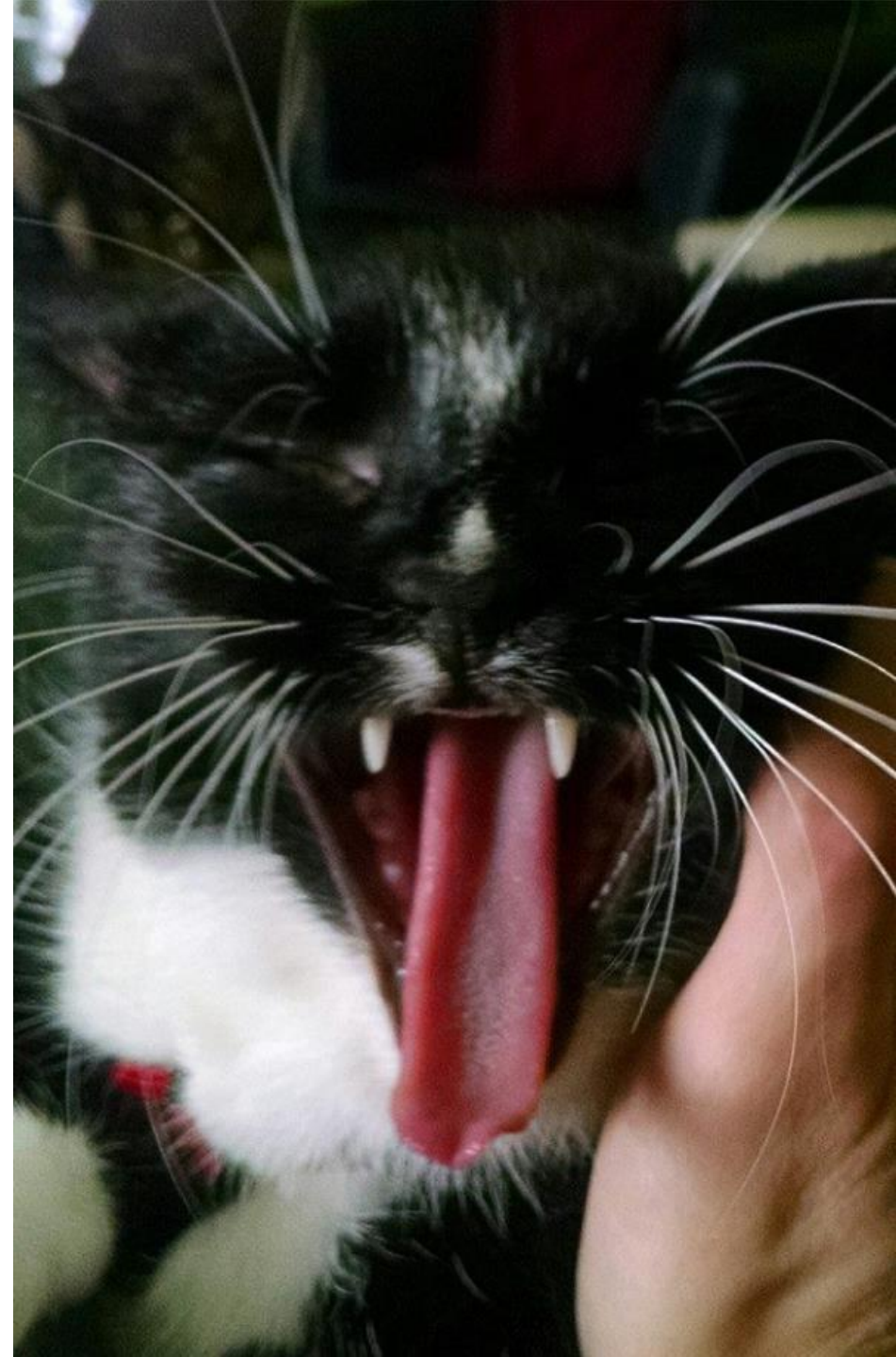
In Millennium

System went down, seemingly related to very long variable fields in overlaying records.

2016-09-16

We resume batch loads in production

^- - THREE MONTHS HAS ELAPSED



2016-09-16: WCM status



3 months
of
add+change+delete files
for
~15 WCM filesets

Cries for help



1. OCLC cannot output a fresh set of base files
2. OCLC **strongly advises** against de-selecting and re-selecting all MARC collections to get fresh record sets.
3. “Include a complete set of records with every delivery” option now available, but only at collection level.

The workable solution:

“Include a complete set of records with every delivery”

PROS

- It'll work
- Will always have latest complete files
 - Less tied to specific schedule of prep/load (if workflow is right)
- Less confusing for staff to get the right files
 - Mistakes easier to remedy



The workable solution:

“Include a complete set of records with every delivery”

CONS

- Reset delivery option on ~250 collections... manually
- Requires completely rethinking workflow
- Especially in light of...



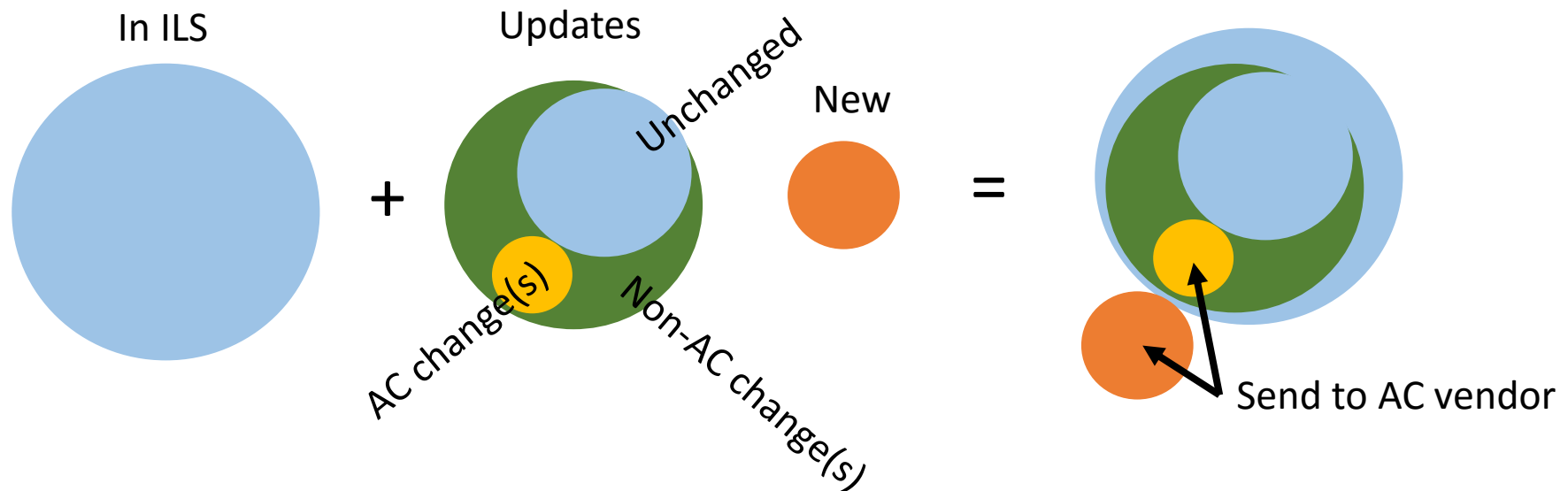
Another twist

minimize authority control vendor costs



Full-file workflow: what do we actually need?

- Only load updates that have changed
- Know which changes affected AC headings



- This is already too complex for a manual workflow.
- So, since I'll be writing some code...

Full-file workflow: what else would we like to do?

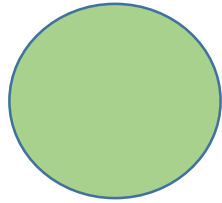
- No authority control for records w/ low encoding level (LDR/17)
- Have flexibility to follow our own processing/load schedule



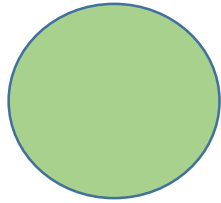
Expectation vs. reality

“Include a complete set of records with every delivery”

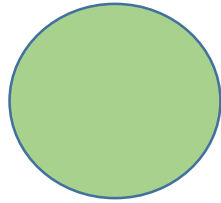
My expectation (weekly delivery):



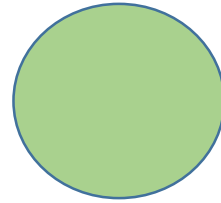
2016-03-06
Complete set



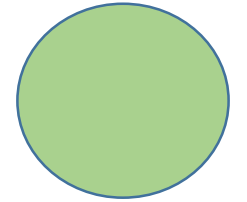
2016-03-13
Complete set



2016-03-20
Complete set

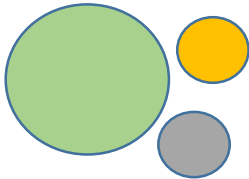


2016-03-27
Complete set

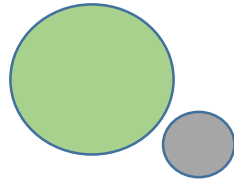


2016-04-03
Complete set

Reality (weekly delivery):



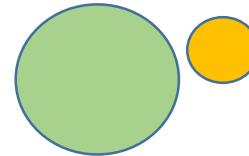
2016-03-06
Updates, News,
Deletes



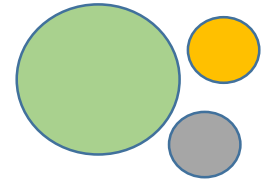
2016-03-13
Updates,
Deletes



2016-03-20
Updates only



2016-03-27
Updates, News

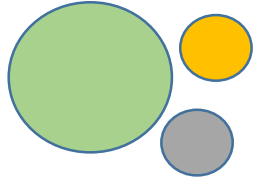


2016-04-03
Updates, News,
Deletes

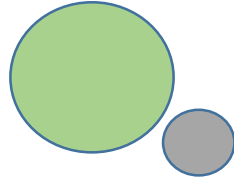
Expectation vs. our needs

Processing/loading of weekly record deliveries

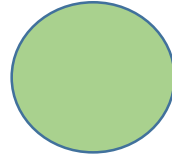
WCM expectation: each weekly set, loaded in order?



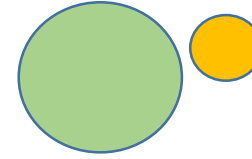
2016-03-06
Updates, News,
Deletes



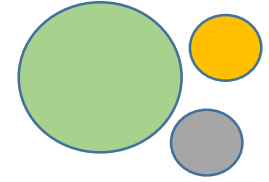
2016-03-13
Updates,
Deletes



2016-03-20
Updates only

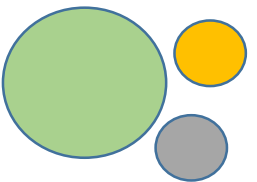


2016-03-27
Updates, News

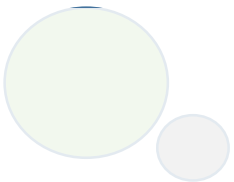


2016-04-03
Updates, News,
Deletes

Our needs: Routine monthly load for most sets, but weekly delivery to ensure a relatively fresh full set at any time



2016-03-06
Updates, News,
Deletes



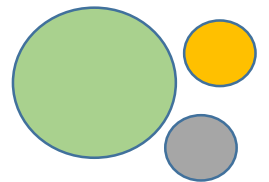
2016-03-13
Updates,
Deletes



2016-03-20
Updates only



2016-03-27
Updates, News



2016-04-03
Updates, News,
Deletes

Solution!

Script + new load profile + workflow



So I wrote (yet another) script

<https://github.com/UNC-Libraries/MARC-record-set-wrangler>

- Goal:
 - Create a solution to all pieces of the WCM problems I've mentioned...
 - That my staff can run in one step (complexity is mainly hidden)...
 - AND do it in such a way that hopefully other people can use it
 - Don't hard-code UNC-specific practices
 - Make it all configurable

So I wrote (yet another) script

<https://github.com/UNC-Libraries/MARC-record-set-wrangler>

- Result:
 - Just for us: utility will go far beyond WCM MARC processing
 - Complexity is all in setting up the configuration
 - But that part is my job, so it's ok
 - Is it potentially useful to others?
 - You tell me...

Name

- existing_marc
- incoming_marc
- output

config.yaml

MARC_set_wrangler.rb

← → ▾ ↑ OCLC-WorldShare-MARC-processing > existing_marc

Name	Size
SPR2016_D20161221_updates_1.mrc	25,695 KB

← → ▾ ↑ OCLC-WorldShare-MARC-processing > incoming_marc

Name	Size
SPR2016_D20170319_new_1.mrc	54 KB
SPR2016_D20170319_updates_1.mrc	28,594 KB

← → ▾ ↑ OCLC-WorldShare-MARC-processing > output

Name	Date modified
This folder is empty.	

Collection-level configuration

```
collections:
WCM - SPR2016:
  id affix value: 'SPR2016'
  add MARC field spec:
    - tag: '773'
      i1: '1'
      i2: ' '
      subfields:
        - delimiter: 't'
          value: 'OCLC WorldShare Collection Manager managed collection. SpringerLink 2016'
    - tag: '590'
      i1: ' '
      i2: ' '
      subfields:
        - delimiter: 'a'
          value: 'Provider: SpringerLink.'
    - tag: '506'
      i1: '1'
      i2: ' '
      subfields:
        - delimiter: 'a'
          value: 'Access limited to UNC Chapel Hill-authenticated users.'
        - delimiter: 'f'
          value: 'Unlimited simultaneous users'
    - tag: '949'
      i1: ' '
      i2: '1'
      subfields:
        - delimiter: 'g'
          value: '1'
        - delimiter: 'l'
          value: 'ebnb'
        - delimiter: 't'
          value: '11'
        - delimiter: 'r'
          value: 'n'
        - delimiter: 'h'
          value: '0'
        - delimiter: 'u'
          value: '-'
        - delimiter: 'n'
          value: 'Perpetual access'
```

workflows:

WCM - routine update using most recent whole file + last whole file:

clean ids:

- find: '^o(c[mn]|n)'
replace: ''
- find: '*\$'
replace: ''
- find: '\\\$'
replace: ''

use id affix: true

affix type: 'suffix'

id affix value: 'wcm'

use existing record set: true

overlay merged records: true

manipulate 019 for overlay: true

flag overlay type: true

set record status by file diff: true

report record status counts on screen: true

omit from comparison fields:

- tag: '005'
field has: 'Reason for updated record:'
- tag: '019'
- tag: '035'
- tag: '598'
- tag: '655'
field has: 'Electronic book'
- tag: '773'
field has: '\\(online collection\\)'
- tag: '938'
- tag: '994'
- tag: '999'

omit from comparison subfields:

- '040': 'cd'

Workflow-level configuration

```
flag rec status: true
incoming record output files:
  CHANGE: '_change'
  STATIC: '_static'
  NEW: '_add'
produce delete file: true
report delete count on screen: true
elvl sets AC status: true
add AC MARC fields: false
add noAC MARC fields: true
flag AC recs with changed headings: true
warn about non-e-resource records: true
warn about cat lang: true
write warnings to recs: true
log warnings: true
```

```

institution:
  main id: '001'
  merge id: '019a'
  cat lang:
    - 'eng'
  overlay type flag spec:
    - tag: '999'
      i1: '9'
      i2: '9'
      subfields:
        - delimiter: 'a'
          value: 'Overlay type'
        - delimiter: 'b'
          value: '[OVRTYPE]'
  warning flag spec:
    - tag: '999'
      i1: '6'
      i2: '6'
      subfields:
        - delimiter: 'a'
          value: 'WARNING'
        - delimiter: 'b'
          value: '[WARNINGTEXT]'
  rec status flag spec:
    - tag: '999'
      i1: '7'
      i2: '7'
      subfields:
        - delimiter: 'a'
          value: 'Record status'
        - delimiter: 'b'
          value: '[RECORDSTATUS]'
  elvl AC map:
    '2': 'noAC'
    '3': 'noAC'
    '4': 'noAC'
    '5': 'noAC'
    '7': 'noAC'
    '8': 'noAC'
    'K': 'noAC'
    'L': 'noAC'
    'M': 'noAC'
    'J': 'noAC'
    ' ': 'AC'
    '1': 'AC'
    'I': 'AC'
  fields under AC:
    - tag: '1..'
    - tag: '240'
    - tag: '4([04]0|1[01])'
    - tag: '490'
      i1: '0'
    - tag: '6([013].|5[01])'
      i2: '[ 012]'
    - tag: '6([013].|5[01])'
      i2: '7'
      field has: '\$2(lcsh|mesh|lcac)'
    - tag: '655'
      i2: '0'
      field does not have: '\$aElectronic book'
    - tag: '655'
      i2: '7'
      field has: '\$2(lcsh|mesh|lcfgt|gsafd)'
      field does not have: '\$aElectronic book'
    - tag: '7([03]0|1[01])' #700, 710, 711, 730
    - tag: '8([034]0|1[01])'

```

Institution-level configuration

```

add noAC MARC spec:
  - tag: '915'
    i1: ' '
    i2: ' '
    subfields:
      - delimiter: '9'
        value: 'NOTAUTHO'
  - tag: '949'
    i1: ' '
    i2: ' '
    subfields:
      - delimiter: 'a'
        value: '*b3=x'
changed heading MARC spec:
  - tag: '599'
    i1: ' '
    i2: ' '
    subfields:
      - delimiter: 'a'
        value: 'LTIEXP'

```

Using the thing...

```
PS Microsoft.PowerShell.Core\FileSystem: \\ad.unc.edu\lib\departments\TechServ\ESM\e-resources  
cataloging\E-book record loads\WCM - OCLC WorldShare Collection Manager\OCLC-WorldShare-MARC-  
processing> ruby .\MARC_set_wrangler.rb
```

- 1) WCM - routine update using most recent whole file + last whole file
 - 2) WCM - whole file - new or fresh reload
 - 3) WCM - routine update using most recent whole file + current ILS set
- Which of the above workflows do I use? 1

Using the thing...

```
1) WCM - ACS
2) WCM - AGU
3) WCM - AMS
4) WCM - CIAO
5) WCM - CUP
6) WCM - DGcolumbia
7) WCM - DGedition
8) WCM - DGharvard
9) WCM - DGpenn
10) WCM - DGprinceton
11) WCM - DynCUP
12) WCM - EMS
13) WCM - GSW
14) WCM - RSC
15) WCM - SPIE
16) WCM - SPR2016
17) WCM - SPR2017
18) WCM - SPRgratis
19) WCM - SPRolder
20) WCM - SciDir
21) WCM - SerAsMono
22) WCM - Video
which of the above collections do I use? 16
```

Using the thing...

```
Grabbed 6396 records from the following incoming_marc files:  
SPR2016_D20170319_new_1.mrc  
SPR2016_D20170319_updates_1.mrc
```

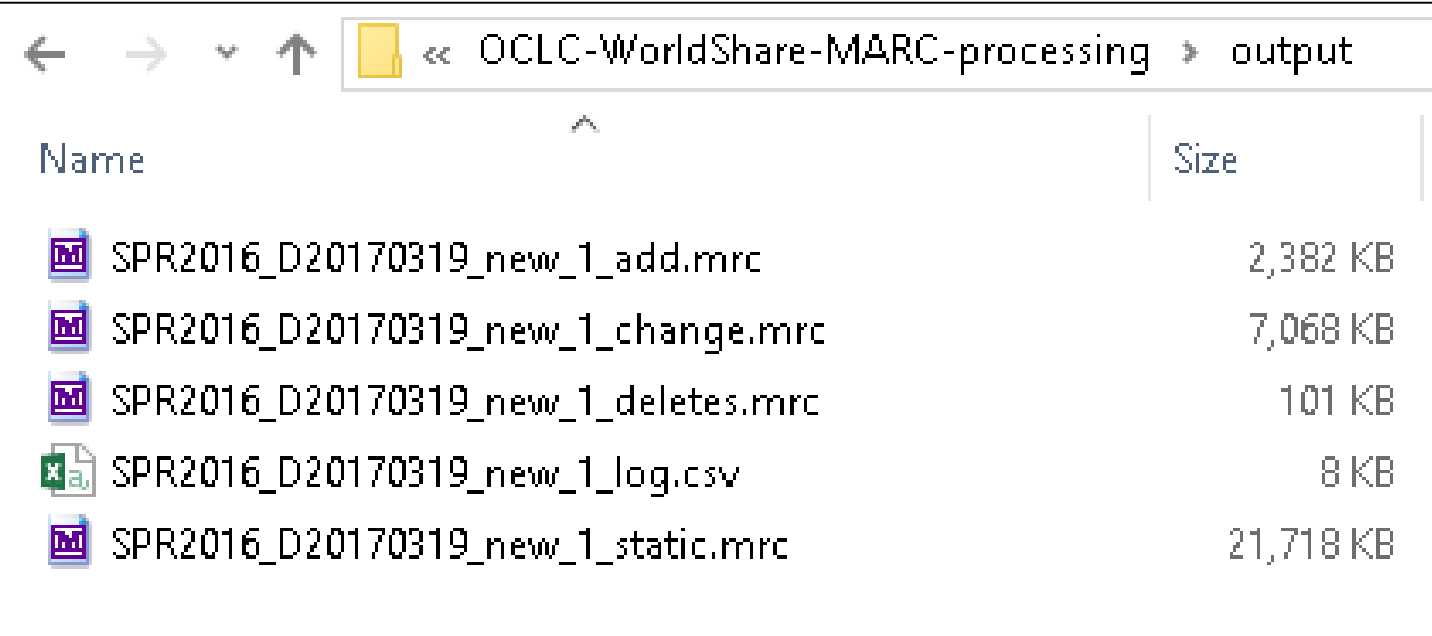
```
Grabbed 5937 records from the following existing_marc files:  
SPR2016_D20161221_updates_1.mrc
```

The suffix wcmSPR2016 will be added to 001, 019a.

```
-----  
Record status counts  
-----  
494 new -- 1335 change -- 4567 static  
35 delete
```

Done!

Results: the output folder



Name	Size
SPR2016_D20170319_new_1_add.mrc	2,382 KB
SPR2016_D20170319_new_1_change.mrc	7,068 KB
SPR2016_D20170319_new_1_deletes.mrc	101 KB
SPR2016_D20170319_new_1_log.csv	8 KB
SPR2016_D20170319_new_1_static.mrc	21,718 KB



Results: log file


	A	B	C	D	E
1	filename	recid	warning		
2	incoming_marc/SPR2016_D20170319_updates_1.mrc	186543952	Not our language of cataloging		
3	incoming_marc/SPR2016_D20170319_updates_1.mrc	701369029	Not an e-resource record?		
4	incoming_marc/SPR2016_D20170319_updates_1.mrc	944061786	Not our language of cataloging		
5	incoming_marc/SPR2016_D20170319_updates_1.mrc	944066684	Not our language of cataloging		
6	incoming_marc/SPR2016_D20170319_updates_1.mrc	946550919	Not our language of cataloging		
7	incoming_marc/SPR2016_D20170319_updates_1.mrc	950926105	Not our language of cataloging		
8	incoming_marc/SPR2016_D20170319_updates_1.mrc	950929612	Not our language of cataloging		
9	incoming_marc/SPR2016_D20170319_updates_1.mrc	950930675	Not our language of cataloging		
10	incoming_marc/SPR2016_D20170319_updates_1.mrc	950931650	Not our language of cataloging		
11	incoming_marc/SPR2016_D20170319_updates_1.mrc	950932625	Not our language of cataloging		
12	incoming_marc/SPR2016_D20170319_updates_1.mrc	950933662	Not our language of cataloging		


Results: MARC from Change file

=999 66\$aWARNING\$Not an e-resource record?

=LDR 03735cam a22006491a 4500
=001 701369029wcmSPR2016 ← ←
=003 OCoLC
=005 20170320101230.6
=008 110104s2011\\ne\\ab\\b\\001\\0\\eng\\d
=019 \\\$a704395405wcmSPR2016\$a704518941wcmSPR2016\$a771427754wcmSPR2016 ←
=020 \\\$a9400702760\$q(print)
=020 \\\$a9400702779\$q(electronic bk.)
=020 \\\$a9789400702769\$q(print)
=020 \\\$a9789400702776\$q(electronic bk.)
=020 \\\$z9789400702769
=035 \\\$a(OCoLC)701369029\$z(OCoLC)704395405\$z(OCoLC)704518941\$z(OCoLC)771427754
=040 \\\$aQE2\$beng\$epn\$cQE2\$dGW5XE\$dEBLCP\$dOCLCQ\$dVPI\$dUKMGB\$dOCLCQ\$dN{\$dollar}T\$dE7B
\$dDEBSZ\$dOCLCQ\$dIDEBK\$dJGO
=050 \4\$aRC152\$b.K35 2011
=245 00\$aKala azar in South Asia :\$bcurrent status and challenges ahead /\$cT.K. Jha, E. Noiri, editors.
=260 \\\$aDordrecht ;\$aNew York :\$bSpringer,\$c©2011.
=300 \\\$axi, 148 pages :\$billustrations, maps
=504 \\\$aIncludes bibliographical references and index.

Results: MARC from Change file

=506 1 \ \$aAccess limited to UNC Chapel Hill–authenticated users.\$fUnlimited simultaneous users 

=590 \ \ \$aProvider: SpringerLink. 


=650 \0\$aKala-azar\$xDiagnosis\$zSouth Asia.

=650 \0\$aKala-azar\$xTreatment\$xHistory.


=700 1 \ \$aJha, T. K.,\$cMRCP, DTM.


=700 1 \ \$aNoiri, E.\$q(Eisei)


=773 0 \ \$tSpringer e-books (online collection). Biomedical and life sciences. 2016


=773 1 \ \$tOCLC WorldShare Collection Manager managed collection. SpringerLink 2016 

=856 40\$uhttp://libproxy.lib.unc.edu/login?url=http://link.springer.com/10.1007/978-94-007

=949 \1 \$g1 \$lebnb\$t11 \$rn\$h0\$u-\$nPerpetual access 

=999 66\$aWARNING\$Not an e-resource record? 

=999 77\$aRecord status\$bCHANGE 

=999 99\$aOverlay type\$bOVERLAY on 001 

Results: Authority control changes

Previous/existing record:

```
=001  ocn958270280  
=100  1\ $aQuirk, Thomas Joseph, $eauthor.  
=700  1\ $aCummings, Simone, $eauthor.
```

Incoming -> output record:

```
=001  958270280wcmSPR2016  
=100  1\ $aQuirk, Thomas J., $eauthor.  
=599  \ $aLTIEXP ←  
=700  1\ $aCummings, Simone, $eauthor.
```

Results: Authority control status

Incoming -> output record:

```
=LDR  03770cam a2200673M 4500
=050  \4$aHC401
=245  00$aEconomic Crisis, Development and Competitiveness in
Southeastern Europe$bTheoretical Foundations and Policy Issues.
=260  \\$bSpringer Verlag$c2016.
=300  \\$a1 online resource
=700  1\ $aKarasavvoglou, Anastasios G.$4edt
=700  1\ $aKyrkilis, Dimitrios.$4edt
=700  1\ $aMakris, Georgios.$4edt
=700  1\ $aPolychronidou, Persefoni.$4edt
=915  \\$9NOTAUTHO ←
=949  \\$a*b3=x ←
```

New load profile

Main changes:

- **On overlay:**
 - **Protect CAT DATE**
 - **Do not add item record if one already exists**
 - **Protect 915 (option d)**
 - **Protect 599 (option n)**



Workflow

1. Retrieve files
2. Run through script using:
 - a. routine workflow config
 - b. appropriate collection config
3. Edit new/updated records for load
4. Load new + updated records
5. Post load queries/global updates
6. Deletes



Workflow

incoming elvl	script AC compare re	script AC fields added	in ILS	Type	after load	post load list logic	post load action	final coding	Will be pulled to send to LTI?
low	changes to headings	bcode3=x 599 LTIEXP 915 NOT AUTHO	bcode3=- 915 Under Authority Control	overlay	bcode3=x 599 LTIEXP 915 Under Authority	search on: overlay review file bcode3=x 915 Under Authority Control	Change bcode3: x -> -	bcode3=- 599 LTIEXP 915 Under Authority Control	y
low	no changes to headings	bcode3=x 915 NOT AUTHO	bcode3=- 915 Under Authority Control	overlay	bcode3=x 915 Under Authority Control	search on: overlay review file bcode3=x 915 Under Authority Control	Change bcode3: x -> -	bcode3=- 915 Under Authority Control	n
low	changes to headings	bcode3=x 599 LTIEXP 915 NOT AUTHO	bcode3=x 915 NOT AUTHO	overlay	bcode3=x 599 LTIEXP 915 NOT AUTHO	search on: overlay review file bcode3=x 915 NOT AUTHO 599 LTIEXP	Delete 599 LTIEXP	bcode3=x 915 NOT AUTHO	n
low	no changes to headings	bcode3=x 915 NOT AUTHO	bcode3=x 915 NOT AUTHO	overlay	bcode3=x 915 NOT AUTHO			bcode3=x 915 NOT AUTHO	n
high	changes to headings	599 LTIEXP	bcode3=x 915 NOT AUTHO	overlay	bcode3=- 599 LTIEXP 915 NOT AUTHO	search on: overlay review file bcode3=- 915 NOT AUTHO	Add y599 LTIEXP Delete 915 NOT AUTHO Delete duplicate field: 599	bcode3=- 599 LTIEXP	y
high	no changes to headings	.	bcode3=x 915 NOT AUTHO	overlay	bcode3=- 915 NOT AUTHO	search on: overlay review file bcode3=- 915 NOT AUTHO	Add y599 LTIEXP Delete 915 NOT AUTHO Delete duplicate field: 599	bcode3=- 599 LTIEXP	y
high	changes to headings	599 LTIEXP	bcode3=- 915 Under Authority Control	overlay	bcode3=- 599 LTIEXP 915 Under Authority			bcode3=- 599 LTIEXP 915 Under Authority Control	y
high	no changes to headings	.	bcode3=- 915 Under Authority Control	overlay	bcode3=- 915 Under Authority Control			bcode3=- 915 Under Authority Control	n
high	n/a	.	n/a	insert	CAT DATE NEW			CAT DATE NEW	y
low	n/a	bcode3=x 915 NOT AUTHO	n/a	insert	CAT DATE NEW bcode3=x 915 NOT AUTHO			CAT DATE NEW bcode3=x 915 NOT AUTHO	n
high	no changes to headings	.	bcode3=- 599 LTIEXP	overlay	bcode3=- 599 LTIEXP			bcode3=- 599 LTIEXP	y, though local changes were lost if they hadn't been also in WorldCat
high	changes to headings	599 LTIEXP	bcode3=- 599 LTIEXP	overlay	bcode3=- 599 LTIEXP 599 LTIEXP			bcode3=- 599 LTIEXP 599 LTIEXP	y

Limitations

and/or, future improvements



Limitation

Comparing previously downloaded file to most recent downloaded file:

- Avoids the complexity of comparing unedited records in downloaded files with edited + loaded records
- BUT
 - can't account for authority control or other changes made in ILS since load



Initial intention for this

- Use Sierra API to grab relevant MARC records from ILS on the fly and compare to downloaded file
- Set up a separate workflow config to omit from comparison:
 - Fields/subfields routinely manipulated in pre-load processing
- We've been on Sierra in production less than a month, so I haven't done this yet... plus...



Limitation: Workaround

- Export relevant record set from ILS via data exchange.
- Use workflow accounting for edited vs. non-edited files.
- BUT, for some sets:
 - Entire record sets came back as “Changes” when they looked identical.
- I know exactly what the problem is, but haven’t been able to figure out a way around it...



Known issues: character encoding will mess you up

ü



WCM MARC

LATIN SMALL LETTER U (U+0075)

+

COMBINING DIAERESIS (U+0308)

`\x75\xCC\x88`

MARC from III Data Exchange

LATIN SMALL LETTER U WITH
DIAERESIS' (U+00FC)

`\xC3\xBC`

Other things I want to add



- Have log file tell you what file a reported record was written *to* in addition to source file
- Check file for records with duplicate main ids
- Automated testing (but...)
- Ideally, an easier user experience (but...)
- etc.

Questions?

Comments?
Discussion?



Gimlet



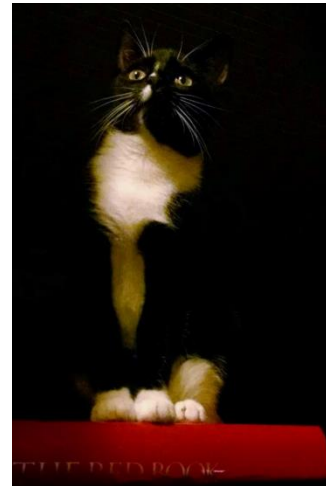
Delilah



Emerson



Toddy



Hermes



Hestia

Thank you!