



# Wrangling E-Resource Catalog Records

AT TEXAS STATE UNIVERSITY

## Who We Are



**Misty Hopper**

Head of Cataloging &  
Metadata Services

Has also worked at:

- University of Texas at Austin–  
Perry Castañeda Library
- University of Illinois Urbana-  
Champaign--Main Library



**Mary Aycock**

Database & Metadata  
Management Librarian

Has also worked at:

- University of Missouri at  
Columbia
- Missouri University of Science &  
Technology

# About our Library

## **Texas State University** **University Libraries Five Branches**

- ▶ Alkek Library
- ▶ Archives & Research Center repository
- ▶ Round Rock Campus Library
- ▶ University Archives
- ▶ The Wittliff Collections

# Meeting the Needs of our Library Staff

## ▶ Acquisitions

- ▶ Separate records for each vendor
- ▶ Separate records (even if same vendor)
  - ▶ Purchased
  - ▶ Leased
  - ▶ Demand driven
- ▶ Easy collection retrieval in the catalog
- ▶ Library of Congress classification for subject analysis
- ▶ Fund codes

## ▶ Subject Librarians

- ▶ Is resource purchased or leased?
- ▶ User access notes
- ▶ Library of Congress classification
- ▶ LCSH subject headings
- ▶ Local subject headings
- ▶ Local collection headings
- ▶ Fund codes

# Tricks we use to meet library staff needs

Separate Records for each vendor/purchased/leased/Demand

- 001 – Bib ID number alpha code prefix is added
  - Alexander Street Press **leased** records: asp12345678
  - Alexander Street Press **purchased** records: aspp12345678

Easy collection retrieval in catalog

- 730 collection titles! Added by Collection Manager or batch processing
  - Example: 730 0 Springer Education Collection.

Fund codes

- Searchable 691 field coded with year, fund code, staff code: **1930L**

Purchased or leased notes in bibliographic records (500 note & permanent 9XX fields)

User access notes in bibliographic records (500 note & permanent 9XX fields)

# What do our researchers need?

- ▶ One record per work
- ▶ Robust bibliographic description
  - ▶ Accurate titles
  - ▶ Table of contents
  - ▶ Summaries
  - ▶ Subjects (both keyword and LCSH)
  - ▶ Authors
  - ▶ Uniform titles (particularly for music works)
- ▶ Links that actually work

# Merging Records from a Single Vendor

- ▶ Collection Manager Delivery Files
  - ▶ Single vendor – Same format – Purchased or Leased
    - ▶ Code single delivery file
    - ▶ For each CM collection, program the addition of a 730 collection access point

**Delivery File Output**

☐ Deliver records for this collection in the same file as other collections

☒ Deliver records for this collection in a separate file

Springer ?

**MARC Custom Text**

Define Custom Text Labels

Text Label	
	60/512
	Springer Biochemical Engineering & Biotechnology Collection.
730	

# All Springer CM Records Delivered to Springer Delivery File

730 Series	Added in CM?	Collection Name on Worldshare Collection Manager
Springer Biomedical and Life Sciences Collection.	yes	Springer Biomedical and Life Sciences eBooks (YYYY) English+International
Springer Earth and Environmental Science Collection.	yes	Springer Book Archive - Earth and Environmental Science
Springer Earth and Environmental Science Collection.	yes	Springer Earth and Environmental Science eBooks (YYYY) English+International
Springer Business and Economics Collection.	yes	Springer Business and Economics eBooks (YYYY) English+International
Springer Business and Economics Collection.	yes	Springer Business and Management eBooks (YYYY) English+International
Springer Business and Economics Collection.	yes	Springer Economics and Finance eBooks (YYYY) English+International
Springer Chemistry and Materials Science Collection.	yes	Springer Chemistry and Materials Science eBooks (YYYY) English+International
Springer Complete Protocols Collection.	yes	Springer Complete Protocols
Springer Computer Science Collection.	yes	Springer Computer Science eBooks (YYYY) English+International
Springer Professional and Applied Computing Collection.	yes	Springer Professional and Applied Computing eBooks (YYYY) English+International
Springer Education Collection.	yes	Springer Education eBooks (YYYY) English+International
Springer Energy Collection.	yes	Springer Energy eBooks (YYYY) English+International
Springer Engineering Collection.	yes	Springer Engineering eBooks (YYYY) English+International
Springer Humanities, Social Sciences, and Law Collection.	yes	Springer Humanities Social Sciences and Law eBooks (YYYY) English+International



# Communication / Communication

## What is Cataloging/Acquisitions doing?

- ▶ How we simplified communication between departments
  - ▶ Spreadsheet in shared folder documenting collection details
    - ▶ Records: Vendor ; Collection Name ; How cataloged ; Purchased/Leased/Demand driven/Evidence based/Open source information ; Format ; Notes about the collection
  - ▶ Internal Cataloging spreadsheet to denote how to process each collection
    - ▶ Coding match points, collection access points, purchased/leased information, etc.

CM MARC Record File Name	Cataloger	Overlays (OCLC # or 856 Link # Identifier)	001 Prefix	Authority (YY/MM)	691	500 "Multi-user access"	500 "Single-user access"	500 "Purchased title"	500 "Leased title"	500 "Open access"
--------------------------	-----------	--	------------	-------------------	-----	-------------------------	--------------------------	-----------------------	--------------------	-------------------

- ▶ Sharepoint Task Form to assign Cataloging staff work:
  - ▶ New/existing collection work
  - ▶ Collections to be cancelled

## The Wild West of Cataloging E-Resources



Title or Collection \*

Vendor

How Cataloged? \*

Priority

Task Status

"New Order - MH" in Task Status means assign to Misty Hopper

Description

[Click for help about adding basic HTML formatting.](#)

For "New Order- MH" describe how MARC records will be retrieved. Collection Manager (manually selected or auto supplied by vendor), Vendor Records, Title by Title.

Comments

691

Tracings

- ☒ None
- ☐ Add Reference Collection tracing
- ☐ Add Single-user access
- ☐ Add Multi-user access
- ☐ Add Career Guide
- ☐ Add Purchased title
- ☐ Add Leased title
- ☐ Add Patron Driven (905)
- ☐ Specify your own value:

Indicates which tracings or notes to add to the MARCs

Public Performance Rights?

Not applicable 

Expiration date?



Date collection will expire online (not needed for renewable subscriptions)

Format \*

E-book 

Overlay old record(s)? \*

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Assigned To \*

Enter a name or email address...

# Receive WorldShare Collection Manager Records

## ► Advantages

- Free! With full membership
- Can update records (metadata or knowledge base)
- Customizable
  - Choose frequency of records
  - Choose which updates to trigger
  - Choose to receive full files each month

The screenshot displays the WorldShare Collection Manager web interface. At the top, the WorldShare logo is visible. Below it, a navigation bar contains tabs for Metadata, Analytics, and Admin. The main content area is titled 'Collection Manager' and includes a search section with a dropdown for 'Scope' (set to 'Collection'), a text input for 'Search Term(s)', and a 'Limit by' checkbox for 'My Selected Collections'. Below the search section are buttons for 'Activity History', 'Approve Changes to Global Collections', 'View Recent Collections', 'Create a Collection', and 'Institution Settings'. At the bottom, there is a 'My Files' section with a right-pointing arrow.

# Trigger examples

Customize Records

WorldCat Updates

Record Delivery

Cataloging Partner Options

## WorldCat Updates

\* Required field

Update Option

Knowledge base and query collections

Update Conditions

☐ Receive updates for any change to the WorldCat master record



☒ Receive updates for selected criteria:

Include

Exclude

## Updates Will Include

Field Changes in Master Record

Tag	5	0	5	Ind 1	Any	Ind 2	Any	 	Reset
Tag	5	2	0	Ind 1	Any	Ind 2	Any	 	
Tag	1	0	0	Ind 1	Any	Ind 2	Any	 	
Tag	7	0	0	Ind 1	Any	Ind 2	Any	 	



# MARC delivery

Customize Records

WorldCat Updates

Record Delivery

Cataloging Partner Options

## Record Delivery

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

**Format**

MARC 21 with MARC-8

**Delivery Frequency**

Monthly



**File Output**



Deliver my serials in a separate file (.ser)



Deliver records for WorldCat updates in a separate file (.wcholdings)

**Uniform Resource Identifiers (URIs)**

Name Headings Identity Services



GND Authorities



LC/NACO Authority File

Subject Headings Identity Services



GND Authorities



LC Subject Authority File



MeSH Authorities

# Processing with CM files

- ▶ Process files via a Ruby program called MARC\_set\_wrangler (credit to Kristina Spurgin, Library Data Strategist, of University of North Carolina, Chapel Hill) <https://github.com/UNC-Libraries/MARC-record-set-wrangler>
- ▶ Compares full collection loads of the current month with the previous month.
  - ▶ Generates new files specific to our parameters

- ▶ New files
- ▶ Updated file
- ▶ Delete files

**omit from comparison fields:**

```
- tag: '65[35]'  
i2: '[1234567]'  
field does not have: 'lcgft|gsafd'
```

# Ruby program: Features we Use

- ▶ Add fields to records
- ▶ Clean up & add prefix to id fields (001 in cataloging speak)
- ▶ Authority control: Flags records that have changes to controlled access points (author, subject, series) which reduces the number of records sent to authority control vendor.

```
1) WCM - routine update using most recent whole file + last whole file (Not Under Authority Control)
2) WCM - routine update using most recent whole file + last whole file (Under Authority Control)
3) WCM - whole file - new or fresh reload
4) WCM - routine update using most recent whole file + current ILS set
which of the above workflows do I use? 3
1) WCM - ACSArchives
2) WCM - ACSFrontFile
3) WCM - ACSESSDigital
4) WCM - AlexAudioPurchased
5) WCM - AlexEbookPurchased
6) WCM - AlexMusicAudio
7) WCM - AlexMusicScore
8) WCM - AlexVideo
9) WCM - AlexVideoPurchased
10) WCM - BloomsburyDesign
11) WCM - EbookCentralAcademic
```

Operates on  
command line



# Final Processing of Records

- ▶ Electronic Resources cataloger plays key role as last stop quality control/troubleshooting. Some activities:
  - ▶ VERY IMPORTANT: Verifies URLs are working
  - ▶ Eliminates fields that can cause catalog problems (029)
  - ▶ Ensures correct collection tracings appears
  - ▶ LC call number (core subject)
    - ▶ When possible, matches call number with current tangible holdings
  - ▶ Subject headings (core subjects)
  - ▶ Diacritic maintenance & other cleanups
- ▶ Loads the records

# Case Study: Upgrading Proquest records

- ▶ Purchased records from Acquisitions vendor
  - ▶ Advantages:
    - ▶ Fairly consistent quality
    - ▶ Ease of cataloging
  - ▶ Disadvantages:
    - ▶ Standards evolve (RDA, Provider Neutral record), records become outdated
    - ▶ Missing out on community enhancements, such as TOC or summaries
  - ▶ Added Challenge:
    - ▶ Records include number of users notes in MARC record (Single-user) that need to be retained/updated

# Phase 1: Create consistent matchpoint to overlay vendor records with CM records

- ▶ 001 of purchased vendor records involved identifier (usually found in URL)
  - ▶ ebc**1075080** →
  - ▶ <https://ebookcentral.proquest.com/lib/txstate/detail.action?docID=1075080>

# Step 1: Delete existing 001

- ▶ Easily convert identifier (001) in records from CM using URL via MarcEdit
  - ▶ Delete existing 001 of OCLC # (Using Add/Delete field)

## *Add/Delete Field Utility*

Field:

001

Field Data:

Delete Field

## Step 2: Build New Field in MarcEdit

Build New Field Tool works by utilizing patterns to generate a new MARC field from existing data in a record. Enter the desired Field pattern, using field/subfield combinations when data from within a record should be used as part of the new field..

Examples:

- \* =856 41\$uhttps://mylibrary.proxy.edu?{856\$u}\$zConnect to this Resource
- \* =099 \a{050\$a}\$b{050\$b}

**Pattern:**

=001 ebc{856\$u}

# Step 3: Regular Expression replacement

- ▶ Using Edit Field, do a replacement to eliminate all parts of the links except for the vendor ID.
- ▶ Before:  
`ebcp`<http://libproxy.txstate.edu/login?url=https://ebookcentral.proquest.com/lib/txstate/detail.action?docID=408518>
- ▶ After: `ebcp408518`

***Edit Field Data***

Field:	Find:
<input type="text" value="001"/>	<input type="text" value="http.*\="/>
<input type="text"/>	Replace:
<input type="text"/>	<input type="text"/>

Options

☐ Match Case

# Phase 1, Matchpoint, continued

- ▶ Added Challenge: Proquest has absorbed so many vendors over the years making it difficult to find official “Proquest” identifier for them (RIP ebrary, etc.)
  - ▶ Proquest spreadsheet to the rescue: had many of these identifiers listed
  - ▶ For those that weren’t found, used macro to obtain URL
  - ▶ Needed to dedup many records



# Deduplication of records: Low-tech method Using Excel

- ▶ 1<sup>st</sup>: Find matchpoint to dedup records (001, URL, etc)
- ▶ 2<sup>nd</sup>: Export selected fields from OPAC records in package
- ▶ 3<sup>rd</sup>: Open in Excel and sort by matchpoint
- ▶ 4<sup>th</sup>: Setup 2 columns with “Exact” formula
- ▶ 5<sup>th</sup>: Copy “True” values into tab
- ▶ 6<sup>th</sup>: Re-sort and select dups to delete
- ▶ 7<sup>th</sup>: Copy dups & save to file to import into ILS



Credit to Matthew Inman for “Tumbeasts”



## Spreadsheet containing potential dups

	A	B	C	D	E
1	1 RECORD #(BIBLIC UPDATED(BIB CREATED(BIBLIO 245 a				
2	EDZ0000018140	b21010249	4/7/2014	4/29/2011	Occasionalism
3	EDZ0000018144	b21010250	9/27/2013	4/29/2011	The unity of consciousness
4	EDZ0000018145	b21010262	3/4/2015	4/29/2011	Truth-conditional pragmatism
5	EDZ0000018146	b21010274	9/27/2013	4/29/2011	Sexual solipsism
6	EDZ0000018147	b21010286	9/27/2013	4/29/2011	Freedom and belief
7	EDZ0000018272	b21010298	9/27/2013	4/29/2011	Semantics vs. pragmatics
8	EDZ0000018152	b21010304	9/27/2013	4/29/2011	Contractualism and the force of law
9	EDZ0000018153	b21010316	9/27/2013	4/29/2011	Truth through proof
10	EDZ0000018156	b21010328	9/27/2013	4/29/2011	Practical tortoise raising, or the limits of reason
11	EDZ0000018170	b2101033x	9/27/2013	4/29/2011	Social epistemology
12	EDZ0000018174	b21010341	9/27/2013	4/29/2011	Partiality and impartiality
13	EDZ0000018287	b21010353	9/27/2013	4/29/2011	Essays on skepticism
14	EDZ0000018183	b21010365	9/27/2013	4/29/2011	Essays in the metaphysics of mind
15	EDZ0000018184	b21010377	1/30/2015	4/29/2011	The domain of reasons
16	EDZ0000018190	b21010389	10/29/2014	4/29/2011	Intuition, imagination, and the limits of reason
17	EDZ0000018294	b21010390	9/27/2013	4/29/2011	Eternal God
18	EDZ0000018194	b21010407	9/27/2013	4/29/2011	Attempts
19	EDZ0000018195	b21010419	9/27/2013	4/29/2011	Procreation and parenthood
20	EDZ0000018204	b21010420	1/23/2014	4/29/2011	Self, no self?

	A	B
1	1	
2	EDZ0000014301	=exact(A2,A3)
3	EDZ0000014302	
4	EDZ0000014303	
5	EDZ0000014304	
6	EDZ0000014305	

	A	B	C
1	1		
2	EDZ0000014301	FALSE	
3	EDZ0000014302		=exact(A3,A2)

## Sort

?

×

+A/z Add Level
✕ Delete Level
📄 Copy Level
▲ ▼ Options...
☒ My data has headers

Column	Sort On	Order
Sort by (Column B) ▼	Values ▼	Largest to Smallest ▼
Then by (Column C) ▼	Values ▼	Largest to Smallest ▼

226	EDZ0001906993	FALSE	TRUE	b47126760	8/9/2018	8/9/2018	Gothic architecture /
227	EDZ0001906994	FALSE	TRUE	b47126772	8/9/2018	8/9/2018	The American West /
228	EDZ0001909089	FALSE	TRUE	b47126784	8/9/2018	8/9/2018	Glasgow /
229	EDZ0001909102	FALSE	TRUE	b47126796	8/9/2018	8/9/2018	Brazilian cinema /
230	EDZ0001909103	FALSE	TRUE	b47126802	8/9/2018	8/9/2018	Katharine Hepburn /
231	EDZ0001909104	FALSE	TRUE	b47126814	8/9/2018	8/9/2018	Invasion of the body sn
232	EDZ0001909105	FALSE	TRUE	b47126826	8/9/2018	8/9/2018	Repo man /
233	EDZ0000014301	FALSE	FALSE	b3501524x	10/12/2016	10/12/2016	Dress and ethnicity

2	A	B	C	D	E	F	G	/ Amer
1	1			RECORD #(BIBLIO)	UPDATED(BI	CREATED(BIBLIO)	245   a	
2	EDZ0000023516	FALSE	TRUE	b46995791	8/2/2018	7/25/2018	Access to health care	
3	EDZ0000023516	TRUE	FALSE	b46984847	8/2/2018	7/25/2018	Access to health care	
4	EDZ0000023517	FALSE	TRUE	b46995808	8/2/2018	7/25/2018	Addiction	
5	EDZ0000023517	TRUE	FALSE	b46984859	8/2/2018	7/25/2018	Addiction	
6	EDZ0000023518	FALSE	TRUE	b4699581x	8/2/2018	7/25/2018	Community air pollution	
7	EDZ0000023518	TRUE	FALSE	b46984860	8/2/2018	7/25/2018	Community air pollution	
8	EDZ0000023519	FALSE	TRUE	b46995821	8/2/2018	7/25/2018	Health promotion	
9	EDZ0000023519	TRUE	FALSE	b46984872	8/2/2018	7/25/2018	Health promotion	
10	EDZ0000023520	FALSE	TRUE	b46995833	8/2/2018	7/25/2018	Ambient air quality stand	
11	EDZ0000023520	TRUE	FALSE	b46984884	8/2/2018	7/25/2018	Ambient air quality stand	
12	EDZ0000023521	FALSE	TRUE	b46995845	8/2/2018	7/25/2018	Behavior	
13	EDZ0000023521	TRUE	FALSE	b46984896	8/2/2018	7/25/2018	Behavior	

## Phase 2: Resolve User Notes

- ▶ Out of sync with reality due to multiple mergers, database entropy

**500** **Single-user access.**

- ▶ Vendor spreadsheet to the rescue

**1-user,1-user,1-user,1-user,1-user,1-user**

- ▶ Added metadata to local 9XX fields

- ▶ Added Challenge:

- ▶ Different models, such as non-linear purchase models (325 use limit annually)
- ▶ Compromise: Displayed user note as:

**500** **"Multi-user access; 325 use limit"**

# Phase 3: Overlay records with CM records

- ▶ Change 001 of incoming records through MarcEdit
- ▶ Overlay records (Run test beforehand)
- ▶ User access in 970 field: Globally inserted notes after overlay

- ▶ Example: 970 2 user

<b>Misc.</b>	970			2 user
--------------	-----	--	--	--------

- ▶ 500 "Two-user access"

<b>Note</b>	500			"Two-user access"
-------------	-----	--	--	-------------------

# Database Rule: Do No Harm

- ▶ Some bad apples: Overall records improvement, but 5% (~1,000) of CM records worse than existing records
  - ▶ Bare bones minimal records
  - ▶ Subject headings were sub-standard (no LCSH).
  - ▶ 505 & 520 often seemed questionable.

=505 0\Salntro; Half Title; Title Page; Copyright; Contents;

# Conclusion

- ▶ Hopes to someday be able Use Collection Manager to combine records for the same e-resource title
- ▶ Future Library Systems on the market or in development

# Questions? Comments? Ideas?

Misty Hopper

Head of Cataloging & Metadata Services

[MH27@txstate.edu](mailto:MH27@txstate.edu)

512-245-3005

Mary Aycok

Database & Metadata Management Librarian

[MBA18@txstate.edu](mailto:MBA18@txstate.edu)

(512) 245-0510