

IGeLU 2020 Digital Conference and Developers' Day
September 17, 2020

Structured Collection Notes and Dashboard for Electronic Collections Management

Siôn Romaine, Director of Acquisitions & Rapid Cataloging Services (sromaine@uw.edu)

Erin Grant, Director of Cataloging & Metadata Services (egrant2@uw.edu)

Nicholas Schockmel, Acquisitions Metadata Specialist (nicsch@uw.edu)



Outline

- > Session Goals
- > Context
- > The Problem
- > The Dream
- > The Solution
 - Note Structure, SQL, Implementation
- > Benefits and Drawbacks
- > Q&A



Session Goals

- > What we hope you learn/take away
 - Ability to use Alma & Analytics to control data about collection records
 - Ideas for how to implement at your institution at a scale that works for *you*



Context

- > University of Washington Libraries
 - Member of Orbis Cascade Alliance (Network Zone)
 - Alma CZ and OCLC KB preferred over vendor recs and manual cataloging
 - Management of electronic resources split across departments
 - > Acquisitions & Rapid Cataloging (ordering & copy cataloging)
 - > Collection Analysis & Strategy (licensing)
 - > Cataloging & Metadata (original cataloging & quality analysis)
 - > East Asia Library
 - > Law Library



The Problem

- > Hundreds of electronic collections
- > Collection info needed by multiple departments
- > Data needs for each collection
 - Set numbers, record sources, record load frequency, workflows, metadata, etc.
- > Data about collections tracked and duplicated in multiple places
 - Inconsistencies in data
 - Data updates difficult
 - Difficult to get statistics
- > Ensure data is migratable in future



The Problem

- > Spreadsheet-like internal documentation page
 - Silo, hard to update, repurposed from microform set mgmt

Number	Millennium ERM#	Lifeboated	Alma Loading Situation	OCLC KB	Set name	Source	Restrictions	Call #	Record Count from III
Set 001	e10057274	yes	N/A		Pre-1900 Canadiana [microform] (no recs for units 35-36, all subjects; no fiche or recs for units 28-34)	OCLC	No 3rd party dist.	M-2501	14,869
Set 002	e10057262	yes	N/A		History of women [microform]	OCLC WCS	No 3rd party dist.	A8266	9,415
Set 003	e10057134	yes	N/A		Landmarks of science I [microform]	OCLC WCS	No 3rd party dist.	Microcards	3,912
Set 004	e10057250	yes	N/A		Major studies of the LRS/CRS [microform]	OCLC WCS	No 3rd party dist.	A6815	239
Set 005	e10057249	yes	N/A		Major studies and issue briefs of the CRS. Supplement [microform]	OCLC WCS	No 3rd party dist.	A6816	18,906
Set 006	e10057237	yes	N/A		American periodical series I [microform]	OCLC WCS	No 3rd party dist.	A122	101
Set 007	e10057237	yes	N/A		American periodical series II [microform]	OCLC WCS	No 3rd party dist.	A444	1364
Set 008	e10057237	yes	N/A		American periodical series III [microform]	OCLC WCS	No 3rd party dist.	A6680	190
Set 009	e10057225	yes	N/A		Early American imprints, 2nd series	OCLC WCS	No 3rd	Microcards	36,750



The Problem

- > E-resources spreadsheet from previous institution
 - Acquisition, subscription, metadata info - useful to many units
 - But did not want to invest time recreating another external silo

Collection	Acquisition Method	Status of Subscription	Collection Still Being Added To?	CKB or Records Loaded	Vendor Reliability	URLs Stable	Selective vs. Aggregate	Record Update Regularity	Record Update Notification?	Record Retrieval	Type of Content	940	Alma Electronic Collection
Kanopy Patron Driven Acquisition	PDA and one-off purchases		Yes	Records loaded				Quarterly	Yes for FDA	http://emory.screenstepslive.com/s/alma/m/ecr/1/631979-woodruff-kanopy-dda-setup ; https://help.kanopystreaming.com/hc/en-us/articles/210927677-How-to-manage-my-kanopy-marc-records	Streaming video	KANOPY, KANOPY DDA CANDIDATE, or KANOPY DDA PURCHASE	Kanopy Streaming (Kanopy DDA)
Knovel e-books		Unsubscribed						monthly		Records come automatically monthly through OCLC World Collection Sets. If not in record, add proxy prefix to URL and do not add puris. We do not "own" these. Periodically we remove some of the titles	e-book		
Knowledge Unlatched								?		Round 1 records added. Round 2 records incomplete. Possibly finished in Sept 2016 ; updated to 11/7/2016; no new records as of 11/8/2017. Go to: http://collections.knowledgeunlatched.org/downloads-2/	e-book	KNOWUNLATCH	No
Landmarks of science		subscription ended						biannual?		Records come automatically through WorldCat Collection Sets. Notification of availability sent to Marian Kelley. Use instructions on F drive to process records.	microfiche	LANDMARK	
Loeb classical library								biannual?		Look for updates at: http://www.loebclassics.com/pagefaq/frequently-asked-questions.jsessionid=8415C9FAE8E431BA93147CDC19733E8#18 ; last updated 2/22/17 ; last checked 2/21/2017	e-book	LOEBCLASSICAL	No
Lynne Jaffe music outsourcing								monthly		Records are sent in e-mail. Notification sent to Marian Kelley.	CD-SOUND, scores	No	No
Medici TV								monthly		Ordered through WorldCat; per e-mail, will be sent file of records monthly ; loaded to 7/2017	streaming media	MEDICITY	No
Met Opera on Demand								?		Retrieve records in OCLC using the following search: am:metopera and am:on-demand and cs=NYMOT	streaming video and audio	METOPERA	No
Morgan & Claypool Synthesis Computer & Information Science Collections 1-5		complete unless we order set 6						?		Check each date list for title not held. Those records are the ones that need to be loaded. Loaded to 2/21/2017 ; last checked for updates on 11/8/2017	e-book	MORGANCLAYPOOL	No
Open Book Publishers								quarterly		Open access titles; complete set as of March 2017 added	e-book	OPENBOOKPUB	No
Oxford Clinical Psychology								monthly		Go to: http://www.oxfordclinicalpsych.com/pageMARC30020Records/marc-records ; last loaded to Sept 2017	e-book	OXFORDPSYCH	No
Oxford Handbooks online (Economics and Finance, History, Literature, Philosophy, Political Science, Religion collections)	Subscribed	Added Economics and Finance in Fall 2015	Yes, new articles/chapters added monthly	Records loaded	Reliable	Oxford has recently changed its URL structure, but appears to be stable		Monthly		MARC updates available at: http://www.oxfordhandbooks.com/page/marc_updates ; updated to 9/2017	e-book	OXFHANDARCHAEOLOGY or OXFHANDLIT	No
Oxford Reference Online	Subscribed		Yes, regularly	Records loaded	Reliable	Oxford has recently changed its URL structure, but appears to be stable		According to website "regularly throughout the year"		Not sure if more titles are coming ; mk, 4/28/2015; all records added to 4/19/2016	e-book	OXFORDREF	No
Project Muse		As of 3/14, no plans to get 2014 collections						monthly?		we subscribe to: 2010-2013 Complete Collections, 2011-2013 Complete Supplement, 2012 Complete Supplement II. An e-mail from Project MUSE <muse@press.jhu.edu> sent monthly with updates - not updated every month however. Last updated 3/2014 ; Project Muse claimed that there was a 2013 update in 10/14 and 11/14, but the record [same one sent	e-book	PROJECTMUSE	No



The Dream

- > Data stored in a single place
 - Feed into Analytics dashboard
 - No creation/maintenance of external, multiple silos
- > Data entry
 - Easily updated
 - Consistent format, wording
- > Data reporting
 - No need to consult multiple silos
 - No need to create ad hoc reports for daily work
- > Flexibility/extensibility
 - Data can be expanded or collapsed to suit different needs
 - Ability to answer many questions, ones we might not have even thought of yet



The Solution

- > Input structured notes into the Internal description field of the electronic collection records that follow a specific format then use SQL to parse the data into useful data points in Analytics
- > I'll briefly outline the following:
 - Technical details
 - Implementation
 - Benefits and drawbacks








The Solution

> The goal is for this...

Internal description

Set Number: Set 174 | Record Source: OCLC KB | Collection ID: springerlink.protocols | Note: Open collection - annual subscription | Updates: Weekly

> To become this...

Electronic Collection Public Name	Set ID	Record Source	Collection Id (for KB collections, click to view collection details)	Source Note	Note	▲▼	Updates	Total Active Portfolios	Click For a List of Portfolios
ACLS Humanities Ebooks	Set 129	OCLC KB	ACLS.heb				Weekly	5,197	
AccessMedicine	Set 137	OCLC KB	McGrawHillACM				Weekly	189	
American Mathematical Society Ebooks	Set 323	OCLC KB	AMS.egsm2011; AMS.egsm2016; AMS.egsm2017; AMS.etrans234; AMS.ecln26; AMS.ecln2016; AMS.ecln2017; AMS.backfileapplied; AMS.epsapm2014; AMS.epsapm2016; AMS.backfilepure; AMS.epspum2013; AMS.epspum2014; AMS.epspum2016; AMS.epspum2017; AMS.estm2014	Multiple collection IDs on one set			Weekly	692	
American Society Of Civil Engineers ASCE Proceedings	Set 138	OCLC KB	asce.conf				Weekly	684	
American films scripts online	Set 065	OCLC KB	alexanderpress.AmericanFilmScriptsOnline				Monthly	1,024	



Note Structure

- > The internal description field of the electronic collection record uses the pipe character to separate information in the field into usable data
- > This structure can be used in Alma Analytics to parse the data into individual fields

Internal
description

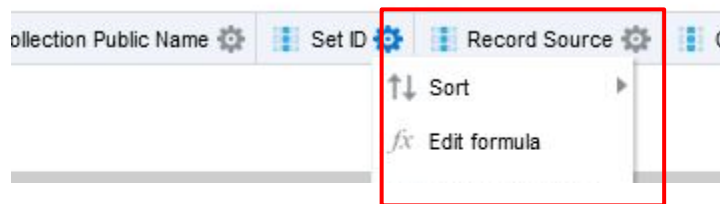
Set Number: Set 174 | Record Source: OCLC KB | Collection ID: springerlink.protocols | Note: Open collection - annual subscription | Updates: Weekly

- > Each note created or modified needs to be in this specific format and any deviation will cause the data to parse incorrectly



SQL Formulas

- > SQL functions are used to parse the internal description into discrete fields in Analytics by editing the column formula:
 - CASE: Uses a sequence of conditions and executes a corresponding statements (basically, if/else)
 - SUBSTRING: Creates a new string starting from a fixed number of characters into the original string.
 - LOCATE: returns the index position of a desired string



- > There are other SQL options to parse strings (like regex)



SQL Code

- > The goal: locate the data point we're trying to isolate and extract the string between that data point and the next data point in the structured note

```
CASE WHEN "Electronic Collection"."Electronic Collection Linked To CZ" = 'Yes' THEN  
NULL WHEN "Electronic Collection"."Electronic Collection Internal Description" NOT  
LIKE '%Set Number%' THEN NULL
```

```
ELSE SUBSTRING("Electronic Collection"."Electronic Collection Internal Description"  
FROM (LOCATE('Set Number: ', "Electronic Collection"."Electronic Collection Internal  
Description")+12) FOR ( LOCATE( '| Record Source: ', "Electronic Collection"."Electronic  
Collection Internal Description")-(LOCATE('Set Number: ', "Electronic  
Collection"."Electronic Collection Internal Description")+12)))
```

```
END
```



Generic SQL Example

- > We created a generic example and placed it in the Alma Analytics Community Folder
 - /shared/Community/Reports/Institutions/University of Washington/IGeLU 2020/Internal Note Parsing Example
- > Use it to test SQL parsing on the internal description of your local electronic collections

Internal Note Parsing Example						
Electronic Collection Internal Description	First Example Field	Second Example Field	Hello FROM Pos	Hello FOR Pos	World FROM Pos	World FOR Pos
First Example Field: Hello Second Example Field: World!	Hello	World!	22	6	51	7



Hello World! Example

> Some of the code for the “First Example Field”

```
SUBSTRING("Electronic Collection"."Electronic Collection Internal Description" FROM  
(LOCATE('First Example Field: ', "Electronic Collection"."Electronic Collection Internal  
Description")+21) FOR ( LOCATE( ' Second Example Field: ', "Electronic  
Collection"."Electronic Collection Internal Description")-(LOCATE('First Example Field: ',  
"Electronic Collection"."Electronic Collection Internal Description")+21)))
```

We can see “Hello” is the parsed output for the “First Example Field:” which starts FROM the 22nd character in the string and continues FOR 6 characters.

Electronic Collection Internal Description	First Example Field	Second Example Field	Hello FROM Pos	Hello FOR Po	World FROM Pos	World FOR Pos
First Example Field: Hello Second Example Field: World!	Hello	World!	22	6	51	7



Other SQL Resources

- > A great resource on SQL formulas in Alma Analytics
 - Rogers, Mike (2016) "Tips and Tricks for Alma Analytics," *Ex Libris Bluegrass Users Group Newsletter*: Vol. 2016 , Article 5.

Available at:

https://digitalcommons.wku.edu/ebug_newsletter/vol2016/iss1/5



Implementation

- > Conservative estimate of ~20 hours of staff time to implement the basic infrastructure which consisted of 4 basic steps:
 1. Extract existing internal notes and online documentation into spreadsheets
 2. Normalize and unify data from existing internal notes with the information in our internal web documentation
 3. Copy and paste normalized, structured notes into collection record
 4. Write SQL code and setup Analytics basic reports and dashboards



Benefits

- > The most important benefits we found:
 - Maintain one record
 - Separate columns means more useful Analytics reports & dashboards
 - Can be used to more easily filter Analytics reports
 - Stores data in a way that is more easily migratable



Benefits -Analytics Reports and Dashboards

- > We can view all e-collections on a dashboard that summarises all the relevant metadata and workflow considerations in one place

Electronic Collection Public Name	Set ID	Record Source	Collection Id (for KB collections, click to view collection details)	Source Note	Note	Updates	Total Active Portfolios	Click For a List of Portfolios
ACLS Humanities Ebooks	Set 129	OCLC KB	ACLS.heb			Weekly	5,197	
AccessMedicine	Set 137	OCLC KB	McGrawHill.ACM			Weekly	189	
American Mathematical Society Ebooks	Set 323	OCLC KB	AMS.egsm2011; AMS.egsm2016; AMS.egsm2017; AMS.etrans234; AMS.ecln26; AMS.ecln2016; AMS.ecln2017; AMS.backfileapplied; AMS.epsapm2014; AMS.epsapm2016; AMS.backfilepure; AMS.epspum2013; AMS.epspum2014; AMS.epspum2016; AMS.epspum2017; AMS.estm2014	Multiple collection IDs on one set		Weekly	692	
American Society Of Civil Engineers ASCE Proceedings	Set 138	OCLC KB	asce.conf			Weekly	684	
American films scripts online	Set 065	OCLC KB	alexanderpress.AmericanFilmScriptsOnline			Monthly	1,024	
Asian Film Online	Set 296	OCLC KB	alexanderpress.asianfilmNA; alexanderpress.asianfilmv2	Multiple collection IDs on one set	Volumes 1-2	Weekly	1,002	

Filter by record source/workflow

CZ
 OCLC KB
 Manual Cataloging
 Vendor

Apply Reset



Benefits -Analytics Reports and Dashboards

> Analytics drilling and action links

Electronic Collection Public Name	Set ID	Record Source	Collection Id (for KB collections, click to view collection details)	Source Note	Note	Updates	Total Active Portfolios	Click For a List of Portfolios
ACLS Humanities Ebooks	Set 129	OCLC KB	ACLS.heb			Weekly	5,197	

UNIVERSITY of WASHINGTON

WorldCat > 1 collection results

Search Browse

Browse: Collections Journals

[ACLS Humanities Ebooks](#)
Collection
Titles subscribed: 5,514

MMS id	Portfolio id	Title (click to view title in WorldCat)	ISBN	Publication Date	Local Param 06	Portfolio Activation Date	Portfolio URL Type	Portfolio URL	Lifecycle
99105911390001452	53493100740001452	Heidegger's technologies : postphenomenological perspectives /	0823235282; 9780823235281; 0823233762; 9780823233766; 0823248887; 9780823248889; 0823233774; 9780823233779; 9780823233762; 0823233766	2010.	1 online resource (xv, 155 pages)	6/19/2013	Static URL	http://hdl.handle.net/2027/heb.09414	In Repository
99105912720001452	53509368830001452	The murals of Eugene Delacroix at Saint-Sulpice /		©1967.	1 online resource (171 pages) : illustrations (some color).		Static URL	http://hdl.handle.net/2027/heb.06156	In Repository
99105913470001452	53506142160001452	The politics of retribution in Europe : World War II and its aftermath /	1400832055; 9781400832057; 9780991099544; 0991099546; 9781282459859; 1282459855; 0612459856; 9786612459851; 0691109530; 9780991099537	©2000.	1 online resource (xv, 337 pages)		Static URL	http://hdl.handle.net/2027/heb.04986	In Repository
99105913480001452	53495372610001452	Antoine Le Pautre : a French architect of the era of Louis XIV /		©1969.	1 online resource (xvi, 162 pages) : illustrations, plans.		Static URL	http://hdl.handle.net/2027/heb.04021	In Repository



Benefits -Analytics Reports and Dashboards

- > Statistics by record source and cataloging workflow

Ebook Portfolios Activated by ...

Portfolio Activation Fiscal Year	Record Source	Activated Portfolios
FY-2020	Manual Cataloging	4,548
	OCLC KB	79,509
	Vendor	144,707
Grand Total		228,764

[Edit](#) - [Refresh](#) - [Print](#) - [Export](#) - [Add to Briefing Book](#) - [Copy](#)



Benefits -Analytics Reports and Dashboards

- > Analytics quality control reports for various workflows

OCLC KB eBook Metadata Screen by Load Date

MMS Id	Title	Bibliographic Level	Language Of Cataloging	MARC 300	OCLC Control Number (035a)	Electronic Collection Public Name	Set Number	Problem Type	Portfolio Id	Portfolio Activation Date	Portfolio URL
99162367110801452	INTRODUCTION TO NEUROPHARMACOLOGY.	m	None			ScienceDirect Ebooks	Set 151	Language of cataloging is not english	53743388550001452	9/9/2020	jkey=https://www.s/science/book/97E
99162367427301452	Digital Foundations: Introduction to Media Design with the Adobe Creative Cloud, Revised Edition	s	English		1151533794	Open Textbook Library	Set 250	Serial bib	53743481650001452	9/5/2020	jkey=https://open.i/opentextbooks/te

Edit - Refresh - Print - Export - Add to Briefing Book - Copy



Benefits - Data Flexibility

- > Structured data with a delimiter makes it easy to move around
- > Data can be exported into a spreadsheet without any SQL
- > Data points can be easily parsed for porting to a new field or system



Drawbacks

- > There are some drawbacks that need to be considered against the benefits:
 - Takes some effort to implement
 - Internal descriptions require staff diligence to be maintained correctly
 - Some Analytics performance issues
 - Would something more simple been better?



Thank you!

Questions?



Suzzallo Library, Jonathan Steinberg, 2017