Ingest Institutional Repository Content to Primo VE via OAI-PMH

Lori Stethers
Systems Librarian, Wesleyan University
lstethers@wesleyan.edu

ELUNA 2023

Agenda

Background

Discovery import setup

Test and troubleshoot

Deleted records

Norm rules on github

Google Doc of example records

Why

Increase visibility of collections

Reduce silos for patrons

What

Make Primo a one-stop search tool for local resources

How

You aren’t interested in managing the titles in Alma

Why

What

How

Sources with an OAI-PMH provider

- LibGuides
- ArchivesSpace
- Institutional repositories
- Data repositories

Sources that can export XML or MARC files

Why

What

How

Primo's harvesting process

OAI - List records

OAI – Get records

Discovery Normalization rules

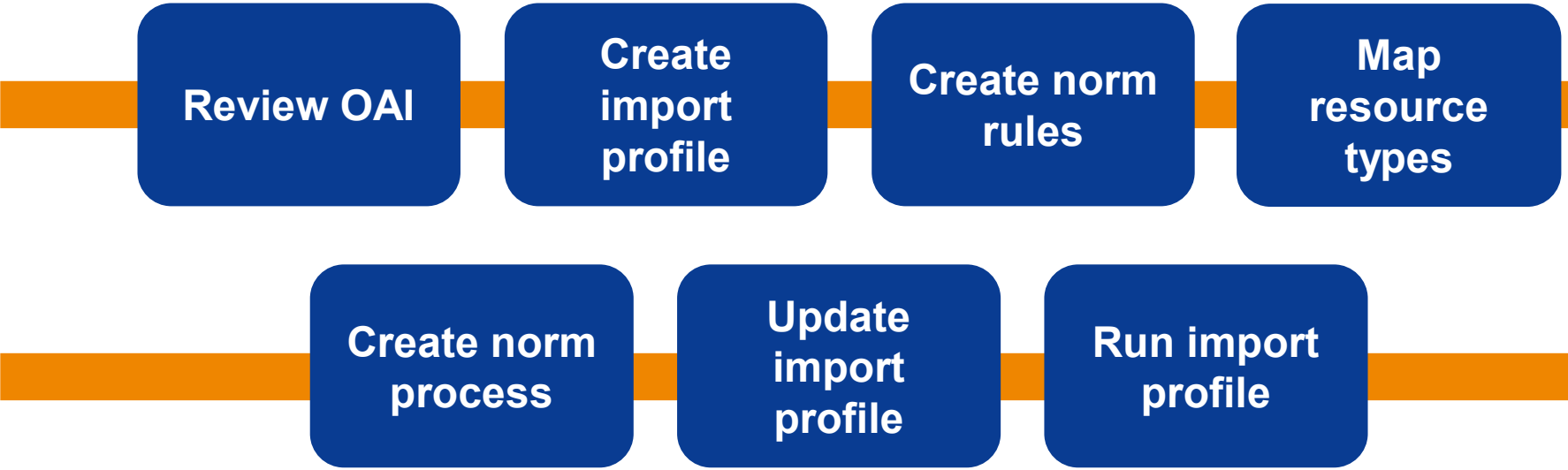
Assign Resource Types

Delivery fields

Publish to Primo

Index

Discovery Import Setup



Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Metadata formats available

Number of records

Fields and values

Links

Record deletion method

OAI verbs to use:

- ListMetadataFormats
- Identify
- ListRecords
- GetRecord

https://digitalcollections.wesleyan.edu/oai/request?**verb=ListRecords**&metadataPrefix=oai_dc

```
      <...snip...>


      <dc:identifier>arupc_b00137_f022_s001_001</dc:ident
      <dc:relation>University Photographer's Collection, Box 137, Folder
      22</dc:relation>
      <dc:rights>In Copyright - Non-Commercial Use Permitted</dc:rights>
    </oai_dc:dc>
  </metadata>
 </record>
 <resumptionToken completeListSize="10367" cursor="0"
 expirationDate="2023-03-17T15:57:08Z">9177</resumptionToken>
</ListRecords>
</OAI-PMH>
```

https://digitalcollections.wesleyan.edu/oai/request?**verb=GetRecord**&metadataPrefix=oai_dc&identifier=oai:digitalcollections.wesleyan.edu:node-1461


```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">

<responseDate>2023-03-17T14:59:24Z</responseDate>
<request verb="GetRecord" metadataPrefix="oai_dc">https://i2-prod.wesleyan.edu/oai/request</request>
<GetRecord>
<record>
 <header>
   <identifier>oai:i2-prod.wesleyan.edu:node-1076</identifier>
   <datestamp>2022-12-05T16:23:48Z</datestamp>
   <setSpec>oai_pmh:all_repository_items</setSpec>
 </header>
 <metadata>
   <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
   http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
     <dc:title>Sermons in Latin : manuscript</dc:title>
     <dc:coverage>England, United Kingdom</dc:coverage>
     <dc:description>Manuscript volume of 87 sermons written on parchment; 30 lines to the page with occasional red or blue initials ; each page
     handnumbered in ms.</dc:description>
     <dc:date>1400</dc:date>
     <dc:format>1 volume (pages 49-492) : parchment ; 19 x 14 cm</dc:format>
     <dc:type>sermons</dc:type>
     <dc:type>Collection</dc:type>
     <dc:language>lat</dc:language>
     <dc:contributor>Everett, James, 1784-1872</dc:contributor>
     <dc:contributor>Clarke, Adam, approximately 1762-1832</dc:contributor>
     <dc:rights>No Copyright - United States</dc:rights>
     <dc:subject>Sermons, Latin -- Early works to 1800</dc:subject>
     <dc:subject>Manuscripts, Latin -- England -- Early works to 1800</dc:subject>
```

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

If modifications are needed...

make them in source system or Primo VE?

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Alma Config > Discovery > Loading External Data Sources > Discovery Import Profiles > Add New Profile

Profile Details | Normalization | Delivery

Create Import Profile

Profile Details

Profile name * **I2-PROD**

Profile description **Islandora2/Islandora8 Dublin Core Prod harvest**

Data Source Code * **I2-PROD**

Data Source Label * **Wesleyan Digital Collections**

Originating system * **Other**

File name patterns

Import Protocol **OAI**

Physical source format **XML**

Source format **Generic XML**

Target format **Dublin Core**

Status **Active**

File splitter parameters

Root element tag * **ListRecords**

Record elements tag * **record**

XPath to the identifier tag * **//record/header/identifier**

XPath to the location of the de

Delete record regular

Discovery Import Setup

OAI Details

Create Import Profile

OAI Base URL * https://digitalcollections.wesleyan.edu/oai/request

Authentication

Connect and Edit

Repository Name Islandora 8

Earliest Date Stam

Granularity YYYY-MM-DDThh:mm:ssZ

Admin EMail/

Metadata Prefix oai_dc

Set All Sets

Identifier Prefix

Harvest Start Date 03/23/2023 07:57

Harvest End Dat

Encode Date

Open Test Page

Discovery Import Setup

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

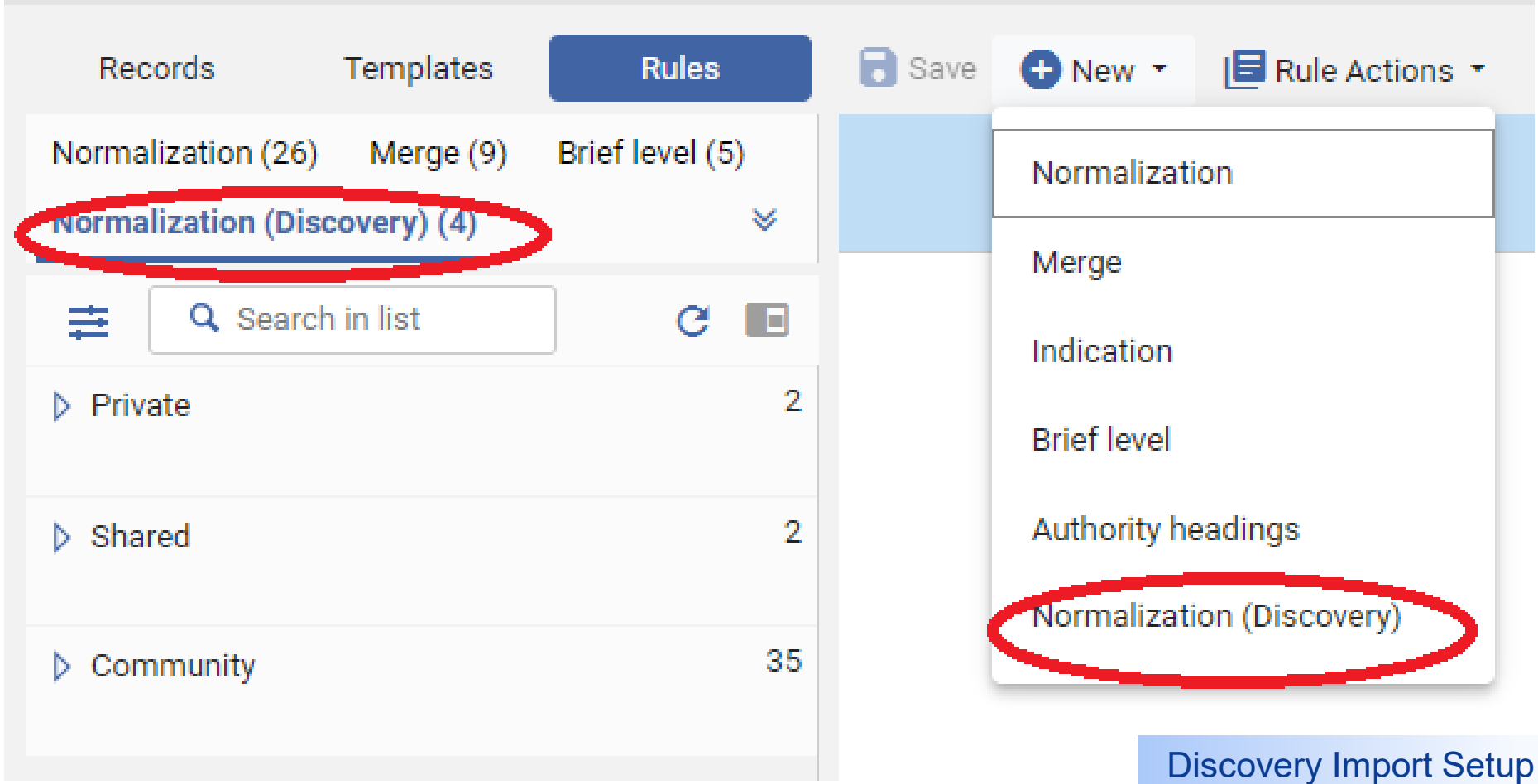
Update import profile

Run import profile

https://github.com/lstethers/primove-normrules

Alma Config > Discovery > Loading External Data Sources > Normalization rules for external data sources

Create Normalization Rules



Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Source format:

Generic XML or

Dublin Core?

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Primo VE Resource Types: 3 ways!

1. discovery:resourceType
2. dcterms:type + Dublin Core Type to Discovery Type Mapping
3. dc:type + Dublin Core Type to Discovery Type Mapping

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Alma Config > Discovery > Loading External Data Sources > Normalization process task > Add process

Create Norm Process Task

Basic Parameters | General Information | Task List | Task Parameters

	Name	Description
1.	DCImportDroolNormalization	DC Drool Normalization

Basic Parameters | General Information | Task List | Task Parameters

Alt + 1

DC DROOL NORMALIZATION

Drools File Key **ArchivesSpace Normalization (DC)**

Review OAI

Create import profile

Create norm rules

Map resource types

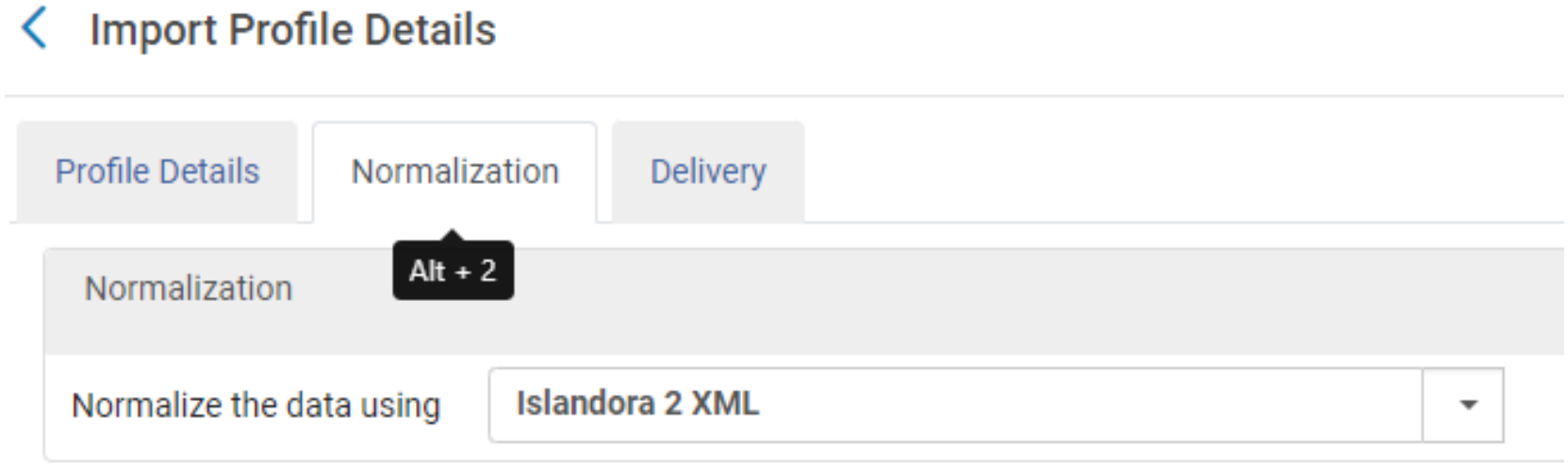
Create norm process

Update import profile

Run import profile

Alma Config > Discovery > Loading External Data Sources > Discovery Import Profiles > (edit your profile)

Update Import Profile: Normalization



Update Import Profile: Delivery: Static URL

Profile Details | Normalization | Delivery

Link to Resource

Template | Static URL from source

Dublin Core tag * dc:identifier

Link Label Online access

Update Import Profile: Delivery: URL Template

Profile Details | Normalization | Delivery

Link to Resource

(•) Template ( ) Static URL from source

Template: **$$LinkingParameter1**

Link Label: **Online access**

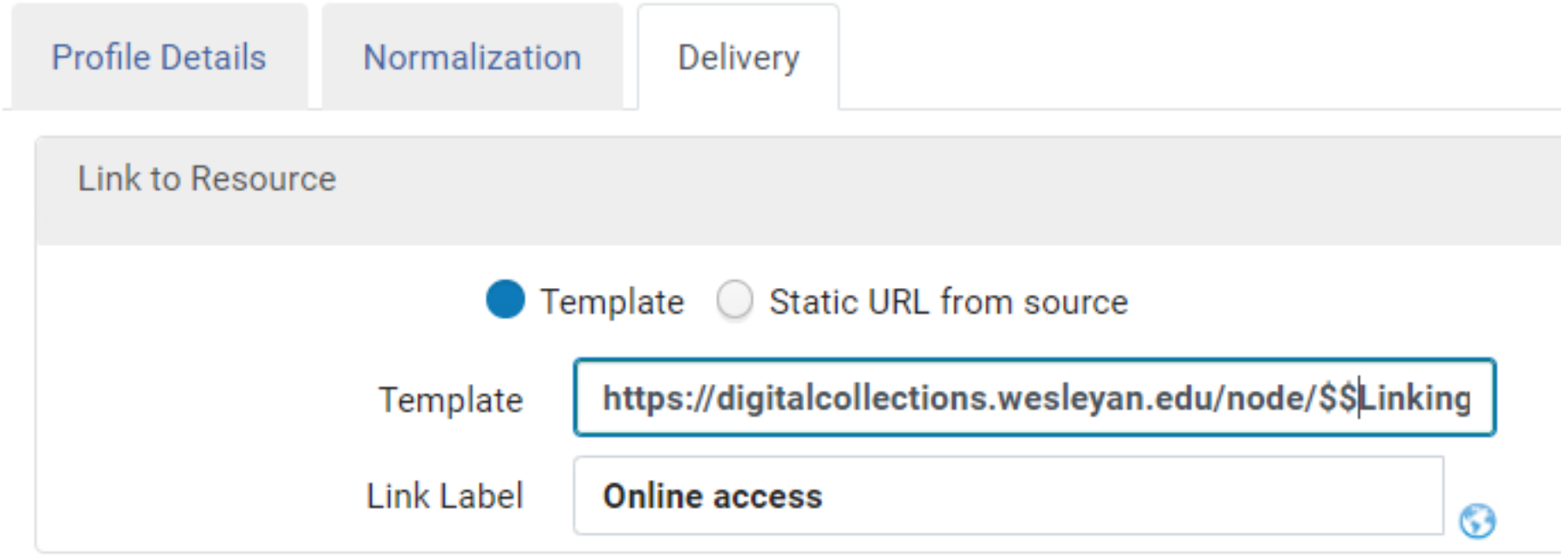
Update Import Profile: Delivery:Linking Parameters

Edit LinkingParameter1

Source Tag *

Use source tag	Matching string using a regular expression
Matching String	^(?:http(s)?:\/\/).*
Normalization source tag	No normalization

Update Import Profile: Delivery: URL Template



Update Import Profile: Delivery:Linking Parameters

Edit LinkingParameter1

Source Tag *

Use source tag	Matching string using a regular expression
Matching String	^oai:digitalcollections.wesleyan.edu:node-.*$
Normalization source tag	Normalize using a regular expression
Normalization Pattern	^oai:digitalcollections.wesleyan.edu:node-(.+)$

Review OAI

Create import profile

Create norm rules

Map resource types

Create norm process

Update import profile

Run import profile

Alma Config > Discovery > Loading External Data Sources > Discovery Import Profiles > Run

Discover Import Profile: Run Import

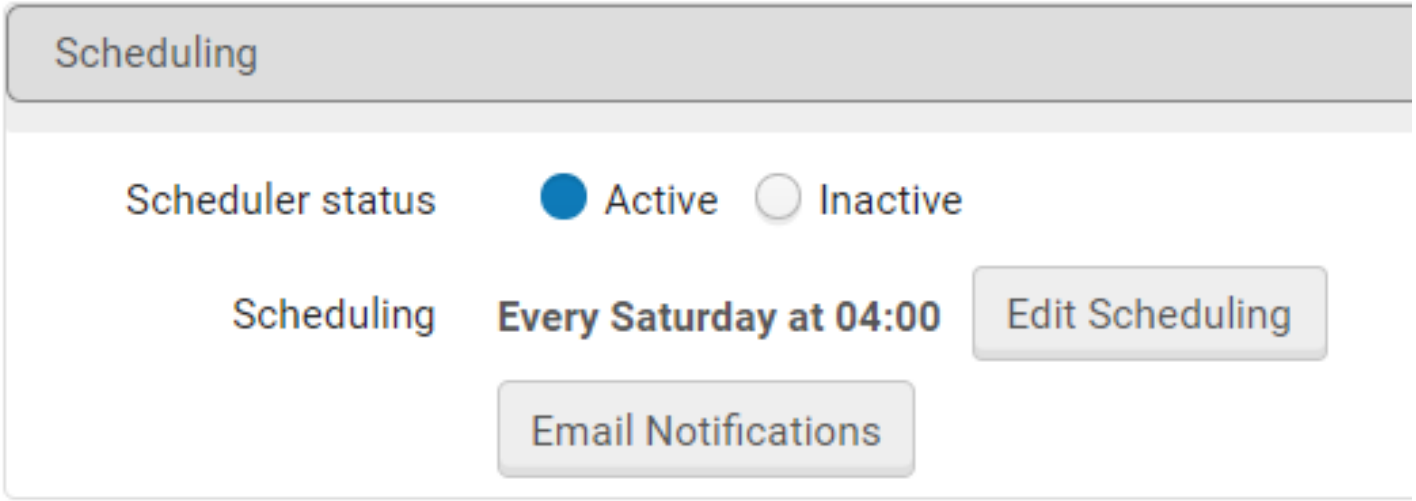
Add New Profile

us	Profile name	Profile description	Profile Type	Contributed by	
e	01CTW_WU_ARCHIVESSPACE	-	Discovery	-	•••
e	01CTW_WU_ISLANDORA	Islandora via MODS	Discovery	-	•••
e	01CTW_WU_ISLANDORA_2	Islandora 2.0 connector	Discovery	-	
e	01CTW_WU_ISLANDORA_DELETES	Islandora via MODS - delete records only	Discovery	-	
e	01CTW_WU_LIBGUIDES	-	Discovery	-	

- Edit
- View
- History
- Run
- Reload
- Reload and Delete
- Copy
- Contribute
- Delete

Discovery Import Setup

Discovery Import Profile: Schedule Import



Test & Troubleshoot

- OAI feed
- Alma OAI test
- Harvest subsets
- Norm rules & norm process
- Reload options
- Logs
- Primo VE records

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records

From a browser

- connect
- how many records
- record fields

OAI feed

Alma OAI test

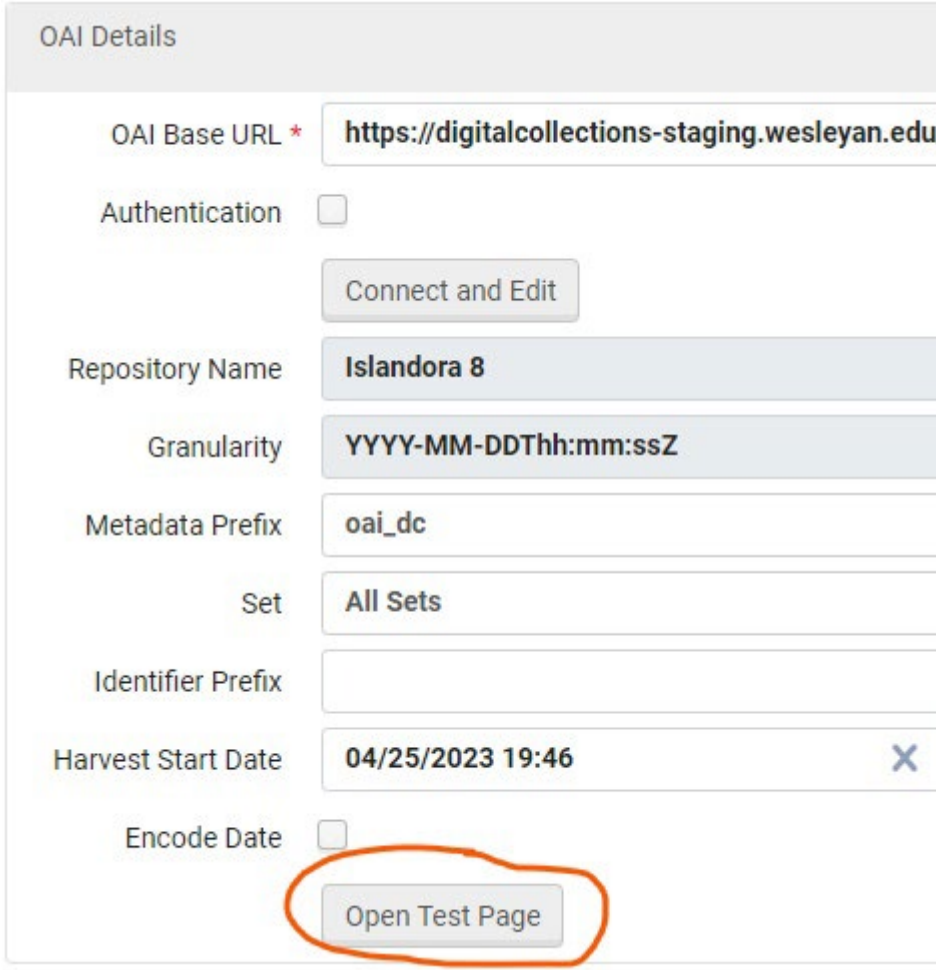
Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records



Test & Troubleshoot

Testing: Alma OAI Test

OAI Test

File was not normalized

Test By ID

ID * oai:digitalcollections-staging.wesley

Test First Record

Test Random Record

Test

Record Identifier

Status PASSED

Source Record

```
<record>
  <header>
    <identifier>oai:digitalcollections-staging.wesleyan.edu:node-175</identifier>
    <datestamp>2023-03-08T21:49:54Z</datestamp>
```

DC Record

```
<record><header><identifier>oai:digitalcollections-staging.wesleyan.edu:node-175</identifier>
<datestamp>2023-03-08T21:49:54Z</datestamp><setSpec>oai_pmh:all_repository_items</setSpec>
</header><metadata><oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
```

Test & Troubleshoot

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records

Limit harvest start and end dates

Use sets

Testing: Harvest Subsets

OAI Base URL * https://digitalcollections-staging.wesleyan.edu/oai/i

Authentication ☐

Connect and Edit

Repository Name: Islandora 8

Earliest Date Stamp: 09/06/2022 19:46

Granularity: YYYY-MM-DDThh:mm:ssZ

Admin EMail/s: @wesleyan.edu

Metadata Prefix: oai_dc

Set: All Sets

Identifier Prefix:

Harvest Start Date: 09/06/2022 19:46

Harvest End Date:

Encode Date ☐

Open Test Page

Test & Troubleshoot

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

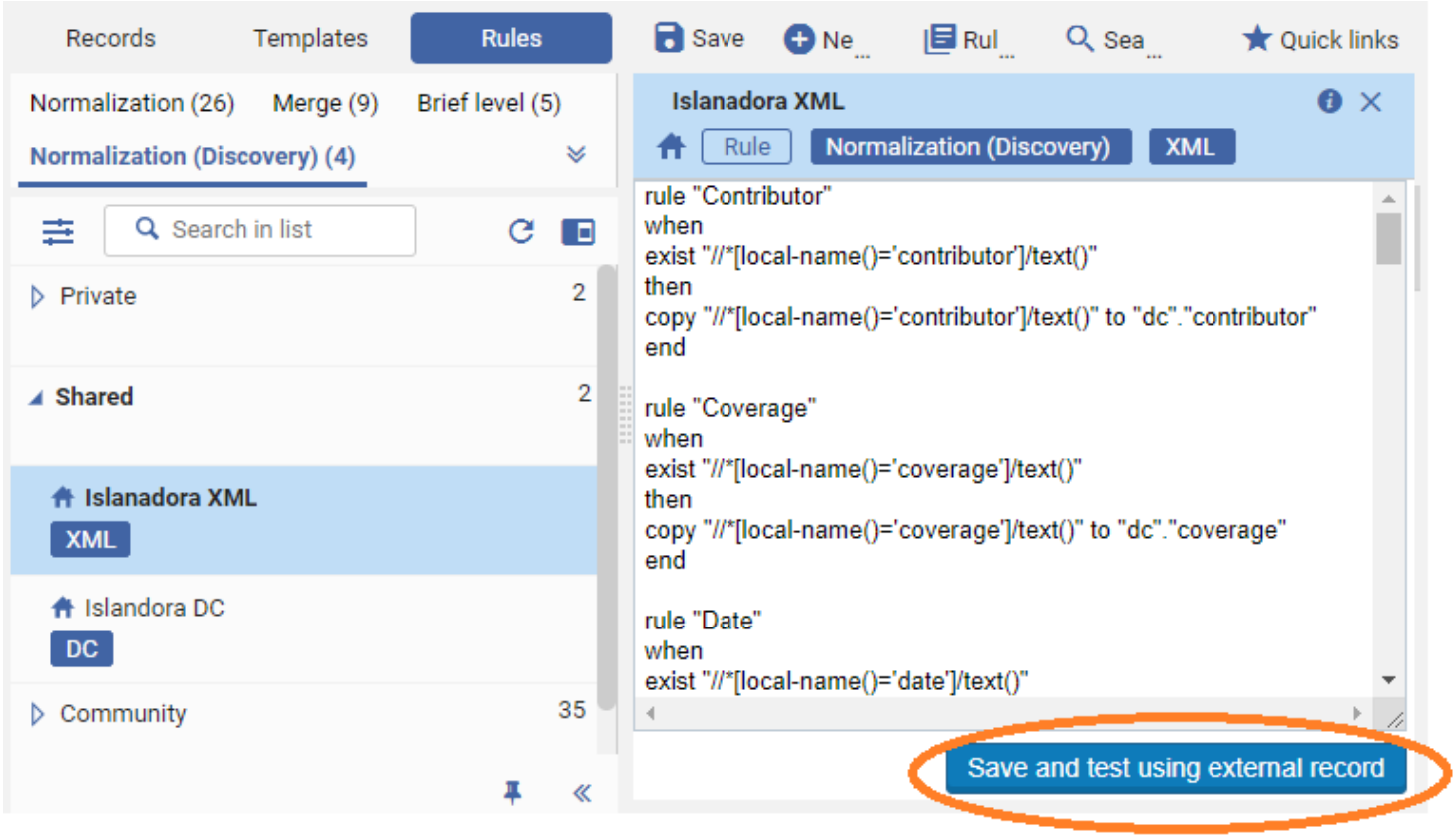
Primo VE records

Alma Config > Discovery > Loading External Data Sources > Normalization Rules for External Data Sources

OR

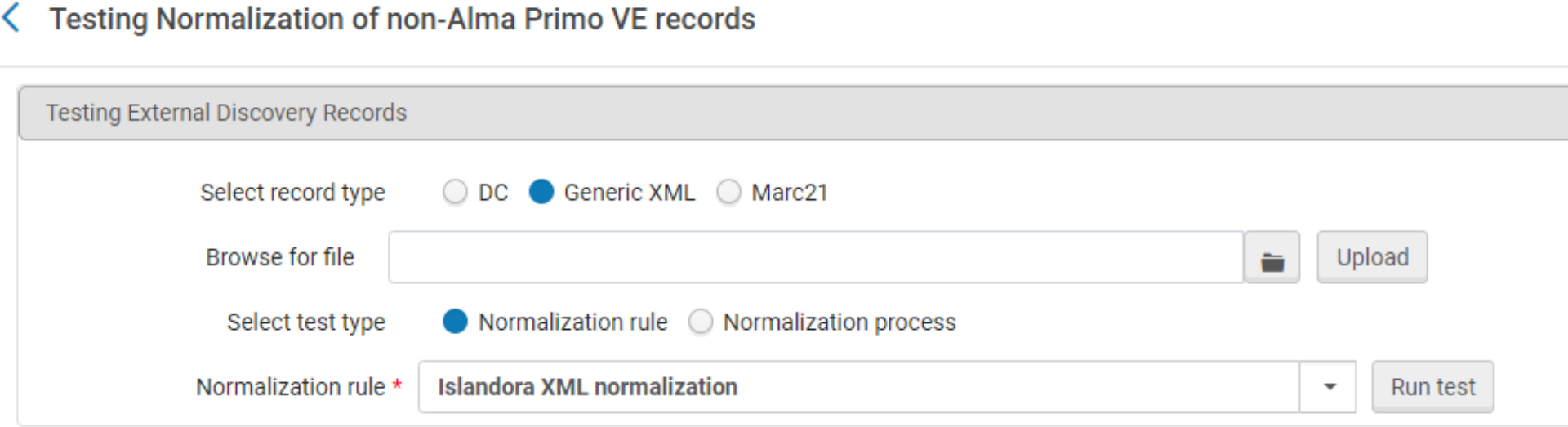
Alma Config > Discovery > Loading External Data Sources > Test Normalization Rules

Testing: Norm Rules & Norm Process



Test & Troubleshoot

Testing: Norm Rules & Norm Process



Test & Troubleshoot

Testing: Norm Rules & Norm Process

Source Record


```
<record>
<header>
<identifier>oai:i2-prod.wesleyan.edu:node-300</identifier>
<datestamp>2022-11-23T14:29:58Z</datestamp>
<setSpec>oai_pmh:all_repository_items</setSpec>
</header>
<metadata>
<qdc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:doc="http://www.lyncode.com/xoai"
xmlns:qdc="http://www.openarchives.org/OAI/2.0/qdc/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/qdc/
http://www.openarchives.org/OAI/2.0/qdc.xsd">

<dc:title>02.003.006 Unidentified group photo of Wesleyan women</dc:title>
<dc:date>1893-09-01</dc:date>
<dc:type>photographs</dc:type>
<dc:language>eng</dc:language>
<dc:format>nonprojected graphic</dc:format>
<dc:rights>No Copyright - United States</dc:rights>
</qdc:dc>
</metadata>
</record>
```


Normalized Record

dc:type	photographs
dc:title	02.003.006 Unidentified group photo of Wesleyan w
dcterms:accessRights	Unrestricted online access
dc:rights	No Copyright - United States
dc:language	eng
dc:identifier	oai:i2-prod.wesleyan.edu:node-300
dc:identifier	oai:i2-prod.wesleyan.edu:node-300
dc:format	nonprojected graphic
dc:date	1893-09-01

Test & Troubleshoot

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records

Reload

Reload and Delete

Delete External Data Sources Job

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records

Alma Config >

Discovery >

Loading External Data Sources >

Discovery Import Profiles >

choose History from the profile menu

Testing: Logs: Report

Job History

Back

1 - 6 of 6 All

Refresh

Filter : All

	Job ID	Status	User	Time Started	Time Ended	Files Processed
1	1641643160003766	Completed with Errors	AlmaLtdAdm1n	2022-12-02 12:19:09	2022-12-02 12:33:15	100%
2	1641604050003766	Completed Successfully	AlmaLtdAdm1n	2022-09-28 15:34:21	2022-09-28 15:34:21	100%
3	1641602480003766	Completed with Errors	AlmaLtdAdm1n	2022-09-28 15:24:11	2022-09-28 15:24:25	100%
4	1641583380003766	Completed with Errors	AlmaLtdAdm1n	2022-09-16 13:50:54	2022-09-16 13:51:00	100%
5	1641583080003766	Completed Successfully	AlmaLtdAdm1n	2022-09-16 13:45:24	2022-09-16 13:45:24	100%
6	1641582780003766	Completed Successfully	AlmaLtdAdm1n	2022-09-16 13:42:56	2022-09-16 13:42:58	100%

Report

View file details

Events

Testing: Logs: Report

Job Report

Failed

Import Data to Primo VE

Process ID	1647666470003766	Started on	03/10/2023 06:25:27 EST
Finished on	03/10/2023 06:25:42 EST	Total run time	15 Seconds
Status	Failed	Status date	03/10/2023 06:25:42 EST
Records processed	0	Records with exceptions	-
Job ID	161		

Alerts

The job completed with errors. For more information view the report details (or contact Support using the process ID).

Errors

General error **Premature end of file.**

Testing: Logs: Report

Job Results Summary

1	Total records processed	19897
2	Records successfully processed	19897
3	Records failed	0

Report Table

	Description	Count
1	New records added	19897
2	Updated records	0
3	Deleted records	0
4	Records deleted not in file	0
5	Records with deleted status but not found in database	0
6	Records that failed to load	0

Testing: Logs: Report

Back

	19897
	19897
	0

	Count	
	19897	...
	0	Preview records
	0	...
	0	...
ut not found in database	0	...
	0	...

Testing: Logs: Report

Events Report

Back

1 - 20 of 19,897

	Mms ID	Originating System ID
1	99244658103766	oai:digitalcollections.wesleyan.edu:node-3563
2	99244657903766	oai:digitalcollections.wesleyan.edu:node-3564
3	99244658503766	oai:digitalcollections.wesleyan.edu:node-3561
4	99244658303766	oai:digitalcollections.wesleyan.edu:node-3562
5	99244658703766	oai:digitalcollections.wesleyan.edu:node-3560
6	99244656903766	oai:digitalcollections.wesleyan.edu:node-3569
7	99244657303766	oai:digitalcollections.wesleyan.edu:node-3567
8	99244651903766	oai:digitalcollections.wesleyan.edu:node-3600
9	99244657103766	oai:digitalcollections.wesleyan.edu:node-3568
10	99244657703766	oai:digitalcollections.wesleyan.edu:node-3565
11	99244657503766	oai:digitalcollections.wesleyan.edu:node-3566
12	99244655903766	oai:digitalcollections.wesleyan.edu:node-3574
13	99244655803766	oai:digitalcollections.wesleyan.edu:node-3575
14	99244656303766	oai:digitalcollections.wesleyan.edu:node-3572
15	99244656103766	oai:digitalcollections.wesleyan.edu:node-3573
16	99244656703766	oai:digitalcollections.wesleyan.edu:node-3570
17	99244656503766	oai:digitalcollections.wesleyan.edu:node-3571
18	99244655503766	oai:digitalcollections.wesleyan.edu:node-3578
19	99244655403766	oai:digitalcollections.wesleyan.edu:node-3579
20	99244655703766	oai:digitalcollections.wesleyan.edu:node-35

Testing: Logs: Events

Events Report

1 - 2 of 2

Filter : All

	Event Description	Event Date	Severity
1	Bulk was finished successfully{BULK_ID=2199980003766, JOB_ID=1641583380003766, F...	09/16/2022 13:50:54 EDT	Information
2	Records that failed to load{BULK_ID=2199980003766, JOB_ID=1641583380003766, ERRO...	09/16/2022 13:50:54 EDT	Information

OAI feed

Alma OAI test

Harvest subsets

Norm rules & norm process

Reload options

Logs

Primo VE records

Search by identifier prefix

Search by individual identifier

Go directly to MMSID

OAI feed

Alma OAI test

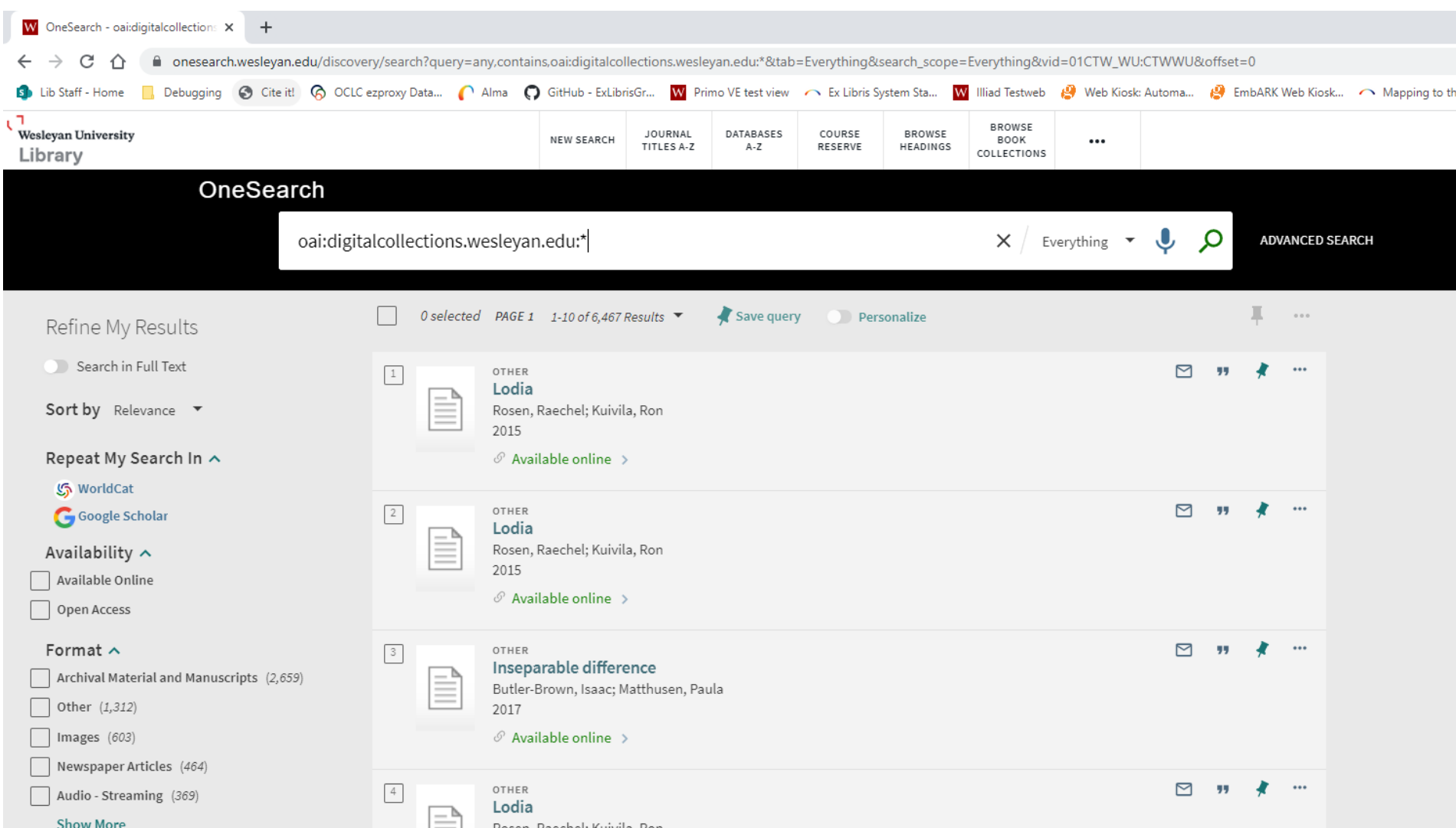
Harvest subsets

Norm rules & norm process

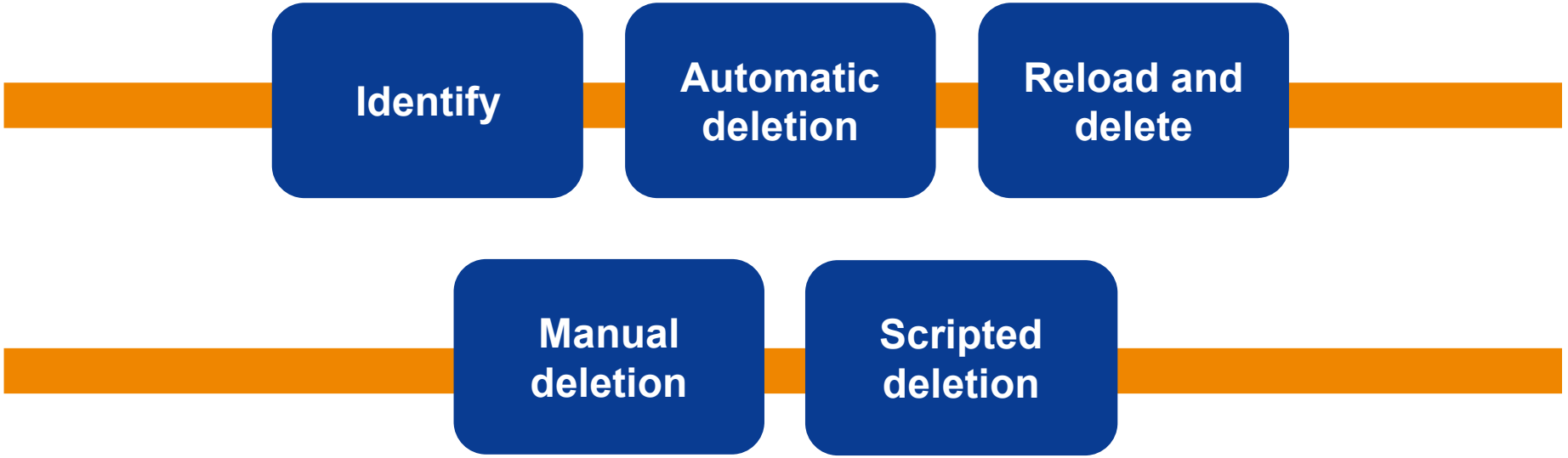
Reload options

Logs

Primo VE records


OneSearch - oai:digitalcollections
onesearch.wesleyan.edu/discovery/search?query=any,contains,oai:digitalcollections.wesleyan.edu:*&tab=Everything&search_scope=Everything&vid=01CTW_WU:CTWWU&offset=0
Lib Staff - Home
Debugging
Cite it!
OCLC ezproxy Data...
Alma
GitHub - ExLibrisGr...
Primo VE test view
Ex Libris System Sta...
Illiad Testweb
Web Kiosk: Automa...
EmbARK Web Kiosk...
Mapping to th
Wesleyan University
Library
NEW SEARCH
JOURNAL TITLES A-Z
DATABASES A-Z
COURSE RESERVE
BROWSE HEADINGS
BROWSE BOOK COLLECTIONS
OneSearch
oai:digitalcollections.wesleyan.edu:*
Everything
ADVANCED SEARCH
Refine My Results
Search in Full Text
Sort by Relevance
Repeat My Search In
WorldCat
Google Scholar
Availability
Available Online
Open Access
Format
Archival Material and Manuscripts (2,659)
Other (1,312)
Images (603)
Newspaper Articles (464)
Audio - Streaming (369)
Show More
0 selected
PAGE 1
1-10 of 6,467 Results
Save query
Personalize
1
OTHER
Lodia
Rosen, Raechel; Kuivila, Ron
2015
Available online
2
OTHER
Lodia
Rosen, Raechel; Kuivila, Ron
2015
Available online
3
OTHER
Inseparable difference
Butler-Brown, Isaac; Matthusen, Paula
2017
Available online
4
OTHER
Lodia

Deleted Records



Identify

Automatic deletion

Reload and delete

Manual deletion

Scripted deletion

What record deletion info does your source system send?

Deleted records

Send an Identify request:

https://digitalcollections-staging.wesleyan.edu/oai/request?**verb=Identify**

Response will include:

<deletedRecord>*some value here*</deletedRecord>

Possible values:

no
persistent
transient

Identify

Automatic deletion

Reload and delete

Manual deletion

Scripted deletion

Alma knows what to do with the persistent or transient delete records

Timing is important for transient info

Deleted records

Identify

Automatic deletion

Reload and delete

Manual deletion

Scripted deletion

Periodically run the Reload and delete option from import profile

Deleted records

Identify

Automatic deletion

Reload and delete

Manual deletion

Scripted deletion

Only makes sense in certain scenarios

1. create an import profile for deletions

2. build & load an XML file of the deleted IDs

Deleted records

Deleted Records: Manual deletion

Profile Details | Normalization | Delivery

Profile Details

Profile name *	01CTW_WU_ISLANDORA_DELETES
Profile description	Islandora via MODS - delete records only
Data Source Code *	01CTW_WU_ISLANDORA2
Data Source Label *	Wesleyan Digital Collections Deletions
Originating system *	Other
File name patterns	*.xml
Import Protocol	Upload File/s
Physical source format	XML
Source format	Dublin Core
Target format	Dublin Core
Status	Active

Deleted Records

```
<ListRecords>
  <record>
    <header status="deleted">
      <identifier>
        oai:digitalcollections.wesleyan.edu:ir_121
      </identifier>
    </header>
  </record>
  <record>
    <header status="deleted">
      <identifier>
        oai:digitalcollections.wesleyan.edu:ir_1471
      </identifier>
    </header>
  </record>
</ListRecords>
```

Identify

Automatic deletion

Reload and delete

Manual deletion

Scripted deletion

Script comparing current source record list to past source record list

Deleted records

Questions

Norm rules on github

Google Doc of example records

Lori Stethers
Systems Librarian, Wesleyan University
lstethers@wesleyan.edu