

Summer 7-31-2015

Harvesting a Local Digital Repository: Experiences and Lessons Learned

Ping Fu

Central Washington University, pingfu@cwu.edu

Follow this and additional works at: <http://digitalcommons.cwu.edu/libraryfac>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Fu, Ping, "Harvesting a Local Digital Repository: Experiences and Lessons Learned" (2015). *Faculty Scholarship from the Library*. Paper 29.

<http://digitalcommons.cwu.edu/libraryfac/29>

This Article is brought to you for free and open access by the Library at ScholarWorks@CWU. It has been accepted for inclusion in Faculty Scholarship from the Library by an authorized administrator of ScholarWorks@CWU.

Harvesting a Local Digital Repository: Experiences and Lessons Learned

Ping Fu

Associate Professor/Head of Library Technology Services

Central Washington University

Minneapolis, MN, May 5-8, 2015

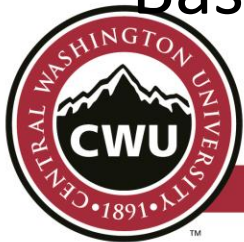
ELUNA 2015



LEARN. DO. LIVE.

Agenda

- Overview
- Digital Commons Repository
- New Data Source
- New Scope
- Normalization Rule
- Define a Pipe
- Run a Pipe
- Create a Tab and a Scope
- Basic Search and Advanced Search Configuration



Audience

- New Primo users
- Plan to buy or migrate to Primo
- Plan to harvest local digital repositories to Primo
- Librarians
- Library IT
- People who are interested in the topic



Overview

- Harvesting data to Primo
- Normalizing the records
- Using pipes
- Harvesting methods
- Create a tab and/or a scope in your institutional view
- More information please see Chapter 2, Primo Interoperability Guide and Primo System Administration Guide.



Digital Commons Repository - ScholarWorks

- <http://digitalcommons.cwu.edu/>
- Digital Showcase
- Scholarly Output
- Open Access



Setup New Data Source

Primo Home > Ongoing Configuration Wizards > Pipe Configuration Wizard > Data Sources Configuration ExLibris Primo

Home | Deploy & Utilities | General | Local Data | Publishing | FE & Delivery | Search Engine Hello ping.fu

> Data Sources

Data Source Attributes for CWU_ScholarWorks

Source Description	Owner: CWU - Central Washington University	Source name: CWU_ScholarWorks	Source code: CWU_ScholarWorks
	Source format: DC	Description:	Original Source Code
Source Definition	Source system: Other	Character Set: UTF-8	Version
	File Splitter: OAI splitter	Transformation file name	Input Record Path oai_dc:dc
	<input type="checkbox"/> File needs transformation		XREF keys LCCN OCLC ISSN ISBN

Created Oct 02, 2014. Last updated Dec 08, 2014 By primo

<https://developers.exlibrisgroup.com/primo/integrations/bo/filesplitter>



LEARN. DO. LIVE.

Setup New Scope Value

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Pipe Configuration Wizard](#) > [Scope Values Configuration](#)

ExLibris

[Home](#) | [Deploy & Utilities](#) | [General](#) | [Local Data](#) | [Publishing](#) | [FE & Delivery](#) | [Search Engine](#)

Hello p

> Edit Scope

General Attributes for Scope CWU - ScholarWorks

Owner: [CWU - Central Washington University](#)

Scope Value Name:

Scope Value Code:

Manually defined Search Scope

Description:

Scope Type:

Shared by other Institutions

Use Scope for: Search Restricted Search Restricted Delivery



LEARN. DO. LIVE.

Setup Normalization Rule

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Pipe Configuration Wizard](#) > Normalization Rules configuration

ExLibris Primo

[Home](#) | [Deploy & Utilities](#) | [General](#) | [Local Data](#) | [Publishing](#) | [FE & Delivery](#) | [Search Engine](#)

Hello ping.fu

> Normalization Rules Sets

Normalization Rules Sets

Owner: CWU - Central Washington University **Display Template NR**

Owner	Normalization Rules Set Name	Version	Description	Normalization Rules Format	
01ALLIANCE-Production	ALLIANCE_ALMA_Jan23	24		marc	View
01ALLIANCE-Production	ALLIANCE_ALMA_CP	24		marc	View
CWU - Central Washington University	CWU_Scholarworks	7		dublin_core	Edit Deploy Test Delete



LEARN. DO. LIVE.

Display Type

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Pipe Configuration Wizard](#) > [Normalization Rules configuration](#)

ExLibris **Primo**

[Home](#) | [Deploy & Utilities](#) | [General](#) | [Local Data](#) | [Publishing](#) | [FE & Delivery](#) | [Search Engine](#)

Hello ping.f

> Manage Normalization Rules

Normalization Rules Set: CWU_Scholarworks
Empty PNx fields

PNx Section: Display



Display

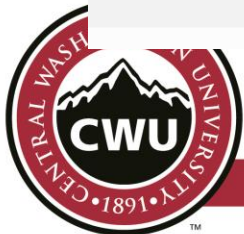
Normalization Rules Set Attributes

Set Name: CWU_Scholarworks
Version: 7

Set Description:
Enable:

Normalization Rules for PNx Section display in Normalization Rules set CWU_Scholarworks

PNx Field	Version	Source Field	Action	Updated By	Updated Date
display:type	1	Type Field XML dc:format	Attribute Value OR- Add single PNx field		
		Type Field XML dc:type	Attribute Value OR- Add single PNx field		
		Type Value			



LEARN. DO. LIVE.

Define Pipe

Source	Normalization and Enrichment	Errors
Data Source: * CWU_ScholarWorks	Normalization Mapping Set: * CWU_Scholarworks	Maximum error threshold: 100 % (1-100)
Harvesting method: * OAI	Enrichment Set: * No Enrichments - Template	
Server: * ions.cwu.edu/do/oai/	Harvested File Format: * *.tar.gz	<input type="checkbox"/> Force DEDUP ⓘ
Metadata format: * dcq		<input type="checkbox"/> Force FRBR ⓘ
Set: <input type="text"/>	<input checked="" type="checkbox"/> Include DEDUP	
Start harvesting files/records from: * Feb 01, 2014	<input checked="" type="checkbox"/> Include FRBR	
Start time: 14 : 36 : 26		



Pipe Monitoring

[Primo Home](#) > [Monitor Primo Status](#) > Pipe Monitoring

ExLibris Primo

[Home](#) [Deploy & Utilities](#) [General](#) [Local Data](#) [Publishing](#) [FE & Delivery](#) [Search Engine](#)

Helloping.fr

> Pipes List

Owner: CWU - Central Washington University **Display Template Pipes**

Owner	Pipe Name	Pipe Type	Stage	Status				
CWU - Central Washington University	CWU_ScholarWorks	Regular	completed successfully	completed	Edit	Execute	History	Errors

Process Status :



LEARN. DO. LIVE.

Harvest Report

[Primo Home](#) > [Monitor Primo Status](#) > Pipe Monitoring

[Home](#) | [Deploy & Utilities](#) | [General](#) | [Local Data](#) | [Publishing](#) | [FE & Delivery](#) | [Search Engine](#)

> **Job details**

Job Details for [2015.02.24-14.15.39-901410816](#)

CWU_ScholarWorks [\[configure\]](#)

[\[View Log\]](#)[\[View Harvest Log\]](#)

Job Id: 2015.02.24-14.15.39-901410816
Stage: completed successfully **Status:** completed
Start Time: 02/24/15 14:16:05 **End Time:** 02/24/15 14:54:22

Harvest

Files Copied: 111	Failed Records: 0
Normal Records: 11078	Deleted Records: 0
Normal Bulks: 12	Deleted Bulks: 0
Num Inserts: 394	Num Updates: 10684 Num Duplications: 0
Failed Files: 0	Empty Files: 0
Start time: 02/24/15 14:16:05	End time: 02/24/15 14:36:40



LEARN. DO. LIVE.

Harvest Report (cont.)

NEP

Succeeded Records: 9646

Failed Records: 1430

Start time:

02/24/15 14:36:36

End time:

02/24/15 14:54:22

Dedup

Single: 0 **Member:** 0 **Failed Records:** 0

Merge: 0 **Vectors:** 0

Start time:

End time:

FRBR

Single: 0 **Member:** 0 **Failed Records:** 0

Keys: 9646

Start time:

End time:



Add a Scope

Orbis Cascade Alliance - Primo Production Environment

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Views Wizard](#)

ExLibris

[Home](#) [Deploy & Utilities](#) [General](#) [Local Data](#) [Publishing](#) [FE & Delivery](#) [Search Engine](#)

Hello pi

> Search Scope List

View :

Search Scope List

Name	Description	Search Type		
primo_central	CWU Primo Central	Remote	Edit	Delete
CWU_ScholarWorks	CWU ScholarWorks	Local	Edit	Delete
cwu_cr	CWU - Course Reserves	Local	Edit	Delete
default_scope	Entire Local Repository and EBSCO and WorldCata	Both	Edit	Delete
cwu_alma_summit	CWU Alma + Summit	Both	Edit	Delete
cwu_alma	CWU Alma	Both	Edit	Delete
ebSCO	CWU Ebsco	Remote	Edit	Delete
worldcat	Worldcat	Remote	Edit	Delete

Add Search Scope

Search Scope Name: Description [Add](#)



LEARN. DO. LIVE.

Add a Tab

Orbis Cascade Alliance - Primo Production Environment

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Views Wizard](#)

ExLibris Primo

[Home](#) [Deploy & Utilities](#) [General](#) [Local Data](#) [Publishing](#) [FE & Delivery](#) [Search Engine](#)

Hello ping.fu

> Tabs Configuration

View : CWU

Tabs List for View CWU

No.	Tab Code	Description			
1	default_tab	Library Resources	Edit	Delete	▲▼
2	cr_tab	Course Reserves	Edit	Delete	▲▼
3	scholarworks	ScholarWorks	Edit	Delete	▲▼

Create a New Tab

Tab Code: Description: [Create](#)



LEARN. DO. LIVE.

Add a Scope to a Tab

Orbis Cascade Alliance - Primo Production Environment

[About](#) [Logout](#) [Help](#)

[Primo Home](#) > [Ongoing Configuration Wizards](#) > Views Wizard

ExLibris Primo

[Home](#) [Deploy & Utilities](#) [General](#) [Local Data](#) [Publishing](#) [FE & Delivery](#) [Search Engine](#)

Hello ping.fu

> Edit View Attributes

View : Interface Language:

General Attributes of scholarworks Tab

Tab Name:
Description:
ToolTip:
Display Find In Database:

Scopes for scholarworks Tab

No.	scope Name	
1	CWU_ScholarWorks	Delete

Add a New Scope

Scope Name:



LEARN. DO. LIVE.

Basic Search

Primo Admin

alliance-primo-bo.hosted.exlibrisgroup.com:1601/primo_publishing/admin/action/views/tilelist.do

Primo Home > Ongoing Configuration Wizards > Views Wizard

Home Deploy & Utilities General Local Data Publishing FE & Delivery Search Engine

ExLibris Primo Hello ping.fu

Edit Basic Search Attributes

View :CWU Title : Basic Search

Texts of Basic Search Tile Interface Language: English

Code	English Label
Opening Text	<input type="text"/>
Ending Text	<input type="text"/>

Values for Resource type pre-filter drop-down

Media Type	default_tab		cr_tab		scholarworks	
	Display	Default	Display	Default	Display	Default
All items	<input checked="" type="checkbox"/>	<input checked="" type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="radio"/>
eBooks	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Articles	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Journals	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Audio & Video	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Conference Proceedings	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Databases	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Dissertations	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Government Documents	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Images	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Legal Documents	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Maps	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Newspaper Articles	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Reference Entries	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Cancel & Go back	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
To Tiles List Scores	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>

Save & Continue To Tiles List



LEARN. DO. LIVE.

Basic Search

The screenshot displays the Primo Admin interface for configuring search fields. The browser address bar shows the URL: alliance-primo-bo.hosted.edibrisgroup.com:1601/primo_publishing/admin/action/views/tilelist.do. The page is titled "Primo Admin" and contains a search bar.

Legal Documents

Document Type	Display	Default	Display	Default	Display	Default
Maps	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Newspaper Articles	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Reference Entries	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Reviews	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Scores	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Web Sites	<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Rare Books	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Statistical Data Sets	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
Research Datasets	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
eBooks (old - do not use)	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
eAudio (do not use)	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>
eVideo (do not use)	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>

Search operators

Field	default_tab		cr_tab		scholarworks		
	Display	Default	Display	Default	Display	Default	
anywhere in the record	<input checked="" type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	Delete

Create a new Search Field Search Field Value:

Search Target Fields of Basic Search tile (Combo)

	default_tab	cr_tab	scholarworks
Precision Operator	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Starts With (Title left anchored search) Operator	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

To Tiles List To Tiles List



LEARN. DO. LIVE.

Advanced Search

> Edit Advanced Search Attributes

View : CWU

Title : Advanced Search

Texts of Advanced Search Tile Interface Language: English

Code

English Label

Opening Text

Ending Text

Define search pre-filters

Search Type	Search Status	default_tab	cr_tab	scholarworks			
Complex Search Line 3	User Defined	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Edit	Delete	▲▼
Complex Search Line 1	User Defined	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Edit	Delete	▲▼
Publication Date	Default	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Edit	Delete	▲▼
Material Type 1	User Defined	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Edit	Delete	▲▼
Languages 1	User Defined	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Edit	Delete	▲▼
Date Range	Default	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		Delete	▲▼
Complex Search Line 2	User Defined	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Edit	Delete	▲▼
Complex Search Line 9	User Defined	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Edit	Delete	▲▼

Create a new search option

Search Type: Date Range

Create



LEARN. DO. LIVE.

Don't forget "Deploy"

Orbis Cascade Alliance - Primo Production Environment

[About](#) | [Logout](#) | [Help](#)

[Primo Home](#) > [Ongoing Configuration Wizards](#) > [Views Wizard](#)

ExLibris Primo

[Home](#) | [Deploy & Utilities](#) | [General](#) | [Local Data](#) | [Publishing](#) | [FE & Delivery](#) | [Search Engine](#) | Hello ping.fu

> Tiles Configuration

To have the View Configuration take affect in the Front End UI the View Configuration must be deployed.

If you do not deploy the configuration now it can be deployed at any time.

Deploy Now






LEARN. DO. LIVE.

Basic Search

comparative analysis

[Advanced Search](#)
[Browse Search](#)

-  [rss](#)
-  [Save query](#)
-  [Add page to e-Shelf](#)

Refine My Results

Topic

- [Accounting students](#) (1)
- [library staffing modes](#) (1)
- [next-generation ILS](#) (1)
- [Client/server computing](#) (1)
- [Higher Education](#) (1)

[More options](#) ▾

Creator

- [Thompson, James H.](#) (1)
- [Fitzgerald, Moira](#) (1)
- [Fu, Ping](#) (1)

[More options](#) ▾

Suggested New Searches

by this author/creator:
[Thompson, James H.](#)

Show **bX** Hot Articles ▾

2 Results for CWU ScholarWorks


Sorted by: Relevance ▾



Article

A Comparative Analysis of the Effect of the Integrated Library System on Staffing Models in Academic Libraries

Fu, Ping; Fitzgerald, Moira
Information Technology & Libraries

 [Online access](#)


[View Online](#) [Details](#) [More](#) [Report Problem](#)



Article

A Comparative Empirical Analysis of Characteristics Associated with Accounting Internships

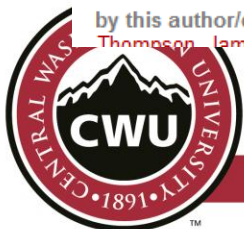
Thompson, James H.
International Journal of Business, Humanities and Technology

 [Online access](#)

[View Online](#) [Details](#) [More](#) [Report Problem](#)

2 Results for CWU ScholarWorks

Sorted by: Relevance ▾



LEARN. DO. LIVE.

Advanced Search

CWU Central Washington University Libraries

My Account Sign Out

New Search | Databases | Periodical A-Z | Research Guides | Chat | Help

OneSearch | Course Reserves | **ScholarWorks**

Any contains AND

Any contains AND

Any contains

Search Clear Simple Search

What is ScholarWorks?

ScholarWorks is a digital showcase of the research and scholarly output of members of the Central Washington University community. Search here to find scholarly articles, presentations, newspapers, images, artwork, and historical materials produced at this university and in the surrounding region.

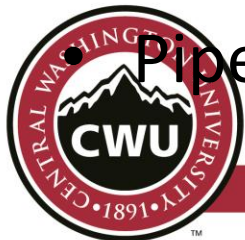
[Learn more...](#)



LEARN. DO. LIVE.

Experience & Lessons Learned

- Correct configuration order
- Correct server name
- Display Type issue
- Metadata format: oai_dc or dcq
- Digital Commons: default metadata prefix is "oai_dc" but offer extended metadata mapping through the prefixes "dcq" & "dcs."
- No enrichments



• Pipe monitoring: stopped error - threshold exceeded

LEARN. DO. LIVE.

Questions? Comments?

pingfu@cwu.edu