

Failed Jobs:
Importing Data to
Primo VE -
Institutional
Repository edition

Alie Visser (avisser9@uwo.ca)

Christina Zoricic (czorici@uwo.ca)

Metadata Management Librarians
Western University



Summary

- Background and Context
- Migration promises and assumptions
- Import profile configuration
- Failed jobs
- What we tried
- Help

Background

- Pre migration platforms:
 - PQ Summon
 - II Sierra
 - BE Press
- 2012ish – stopped cataloguing ETDs since it was indexed in Discovery layer and IR was managed by a Metadata Librarian
- 2018ish – Organizational Renewal (ORI) - responsibility for IR moved to new team (Research and Scholarly Communication)
- 2019/20 and beyond (aka the dark times) = perfect storm (post ORI, migration, pandemic)

IR in Context

- Institutional repository, Scholarship@Western, is not managed by our metadata department, and we do not catalogue theses/dissertations
- User generated metadata: a lot of copying/pasting from Word creates encoding errors that stop the import process
- Repository includes:
 - EBook collections
 - Reports & working papers
 - Conferences & symposiums
 - **Electronic theses & dissertations**
 - Digitized special collections



Migration promises and assumptions



Discovery Import Profile – Part 1



Profile Details	Normalization	Delivery
Profile Details		
Profile name *	Scholarship @ Western	
Profile description	BePress	
Data Source Code *	BEPRESS_SCHOLARSHIP_WESTERN	Data Source Label * Scholarship @ Western
Originating system *	Other	File name patterns
Import Protocol	OAI	
Physical source format	XML	
Source format	Dublin Core	Target format Dublin Core
Status	Active	
Share with Network	<input type="checkbox"/>	

Discovery Import Profile – Part 2

OAI Details

OAI Base URL *	<input type="text" value="https://ir.lib.uwo.ca/do/oai/"/>		
Authentication	<input type="checkbox"/>		
	<input type="button" value="Connect and Edit"/>		
Repository Name	<input type="text" value="Scholarship@Western"/>	Earliest Date Stamp	<input type="text" value="01/18/2000 19:00"/>
Granularity	<input type="text" value="YYYY-MM-DDThh:mm:ssZ"/>	Admin EMail/s	<input type="text" value="dc-support@bepress.com"/>
Metadata Prefix	<input type="text" value="oai-dc"/> ▼		
Set	<input type="text" value="All Sets"/> ▼		
Identifier Prefix	<input type="text"/>		
Harvest Start Date	<input type="text" value="01/01/2005 09:31"/>	Harvest End Date	<input type="text"/>
Encode Date	<input type="checkbox"/>		
	<input type="button" value="Open Test Page"/>		

Failed Job

↕ Name	↕ Process ID	Records processed	Records with exceptions	↕ Job Category	↕ Operator	↕ Submit Date	↕ Start Date	▼ End Date	↕ Status	
Import Data to Primo VE	504333904000051...	0	-	Discovery	System	05/13/2023 04:00:02	05/13/2023 04:01:03	05/13/2023 04:19:34	 Failed	

Alerts



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

Errors

General error **An invalid XML character (Unicode: 0xb) was found in the element content of the document.**

Ex Libris Response:

First response: *"Regarding the job failure, I know that the job is confirmed to run until it finishes, or encounters a corrupted character, in which case the entire import process fails when encountering that character."*

Second response: *"Our Product Management Team has reviewed your request and determined that Alma is currently not designed to provide the functionality you requested, so the next step is to pursue this as an enhancement."*

While we were not able to identify an open-source validator that was able to catch these corrupted characters in your OAI Import profile/repository, it should be possible to manually check for corrupted or invalid characters and replace them with the appropriate character."

What we tried

- Discovery import rather than Resources import profile
 - We don't have to import the records into an Alma collection this way
- Asked Ex Libris to ignore validation errors and continue the import
- Asked BePress for help: "I was able to harvest the entirety of the /etd series via a command-line OAI tool that dumps XML, so this might be more of an issue with how they ingest the data than how we serve it. Is there perhaps a different way for you to try and ingest the data?"

Help?


- Are our partner schools using a Resources > Digital import profile instead?
 - We are trying this, but....
 - “Completed with errors”

S@W test

1 - 1 of 1 All

Refresh

Filter : All

Job ID	Status	User	Time Started	Time Ended	Files Processed
1. 50793981910005163	Completed with Errors	020053557	2023-05-30 14:23:55	2023-05-30 14:43:19	100% 

< Job Report



Metadata Import: S@W test

**Completed
with Errors**

Process ID 50793981910005163
Finished on 05/30/2023 14:43:19
Status **Completed with Errors**
Files Processed 184

Job ID 160

Started on 05/30/2023 14:23:55
Total run time 19 Minutes 24 Seconds
Status date 05/30/2023 14:43:19
Records with
exceptions 0

Counters

Total records processed **9188**
Total records imported 0
Total records deleted 0
Total records not imported/deleted **9188**
Total electronic portfolios activated 0
Total Bibliographic records created 0

< Records Not Imported Due To Validation Errors

Records skipped by validation

1 - 20 of 33,559

	Record ID	Error Message
1	unknown	Field dc:thesis.degree.name is invalid
2	unknown	Field dc:thesis.degree.level is invalid
3	unknown	Field dc:contributor.role is invalid
4	unknown	Field dc:thesis.degree.discipline is invalid
5	unknown	Field dc:thesis.degree.name is invalid
6	unknown	Field dc:thesis.degree.level is invalid
7	unknown	Field dc:contributor.role is invalid
8	unknown	Field dc:thesis.degree.discipline is invalid
9	unknown	Field dc:thesis.degree.name is invalid
10	unknown	Field dc:thesis.degree.level is invalid
11	unknown	Field dc:contributor.role is invalid
12	unknown	Field dc:thesis.degree.discipline is invalid
13	unknown	Field dc:thesis.degree.name is invalid
14	unknown	Field dc:thesis.degree.level is invalid
15	unknown	Field dc:date.available is invalid



Discussion

- Give us your feedback!
- Share your ideas, please