

Infrastructure - Task #2815

EZID - create script and sample DOI batch spreadsheet

2012-05-24 22:28 - Ben Leinfelder

Status: Closed	Start date: 2012-05-24
Priority: Normal	Due date:
Assignee: Ben Leinfelder	% Done: 100%
Category:	Estimated time: 0.00 hour
Target version: Sprint-2012.35-Block.5.2	Story Points:
Milestone: CCI-1.0.0	
Product Version: *	
Description The spreadsheet should include the following columns: EZID account DOI URL (CN.resolve endpoint) Title (for EML docs) Creator (first listed creator)	
Related issues: Related to Infrastructure - Story #2720: KNB, LTER, PISCO, ESA, SANPARKS roll... Closed 2012-05-08	

History

#1 - 2012-05-25 19:08 - Ben Leinfelder

First draft has been sent to Greg at EZID.

There has been mention of including publisher and pub date as well. These fields may be difficult to procure for every object, even EML where it is optional metadata.

#2 - 2012-05-25 23:25 - Ben Leinfelder

- Status changed from New to In Progress

columns now are:
ezid_account

doi
url
title

creator
publisher

pub_date

I've sent a sample to Greg and John at CDL

#3 - 2012-06-04 18:47 - Ben Leinfelder

script is in Metacat SVN:

https://code.ecoinformatics.org/code/metacat/trunk/src/doi_registration.sql

or for tagged versions:

https://code.ecoinformatics.org/code/metacat/tags/METACAT_2_0_0_RC1/src/doi_registration.sql

#4 - 2012-07-05 20:57 - Ben Leinfelder

Helpful pointers for required metadata:

<http://n2t.net/ezid/doc/apidoc.html#internal-metadata>

<http://dublincore.org/documents/dcmi-terms/#H7>

Decided to add some more columns and make adjustments:

- use dc namespace for column headers
- include resource type column (Dataset/DATA, Dataset/METADATA)
- include [DataONE] objectFormat column
- translate DNs to names using LDAP dump
- Data file - see about using entity name
- add obsoletes/obsoleted by column
- add resource map id/location (using ore namespace)
- add public/private access column

#5 - 2012-08-30 15:56 - Ben Leinfelder

After much discussion, I think we are almost ready to do the batch registration. The major contention has been for public vs. "protected" records. Many holdings in the KNB are not entirely visible to the general public and so should not have even their basic record metadata available (name, title, publisher). We have excluded those entries from the batch registration. This means, of course, that those DOIs will not be resolvable, but no one except a select group of a few people should ever encounter those DOIs anyway. Awaiting comments from Kunze and Starr.

#6 - 2012-08-31 01:59 - Chris Jones

- Target version changed from *Sprint-2012.29-Block.4.3* to *Sprint-2012.37-Block.5.3*

#7 - 2012-08-31 17:35 - Ben Leinfelder

- Parent task deleted (#2720)

Moving this task to be stand-alone rather than under the initial Metacat-based MN rollout.

#8 - 2013-01-08 18:56 - Ben Leinfelder

- translation missing: *en.field_remaining_hours* set to 0.0

- Status changed from *In Progress* to *Closed*

The DOIs have been registered with EZID and DataCite. Only bad thing is that they UPPERCASED them all because dois are case-insensitive and that is their standard rendering of them. The resolution URLs still use the original case we sent - thankfully - since our identifiers are very much case-sensitive.