

Member Nodes - Task #3724

MNDeployment # 3552 (Closed): USGS CSAS

Re-typing ORE documents

2013-04-23 19:18 - Skye Roseboom

Status:	Closed	Start date:	2013-04-23
Priority:	Normal	Due date:	
Assignee:	Skye Roseboom	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			
Description Similiar issue to ORNL Daac ORE doc issue where some documents were synchronized with the incorrect formatId ('application/octet-stream'). Need to resolve in the same manner. Updating the formatId on the CN, and on the MN manually. For dataone to update the CN we need a list of pids to correct.			
Related issues: Related to Member Nodes - Task #3839: ORE documents contain references to non... <div>Closed2013-06-24</div>			

History

#1 - 2013-05-02 15:13 - Ranjeet Devarakonda

Refer to <https://redmine.dataone.org/issues/3725>

#2 - 2013-05-14 14:43 - Skye Roseboom

Hi Ranjeet

I created a list of pids from the CN solr index, filtering by USGS CSAS, documents that are not ORE, and Ids that start with resourceMap:

https://cn.dataone.org/cn/v1/query/solr/?rows=1000&q=datasource:urn\:node\:USGSCSAS%20id:resourceMap*%20-formatType:RESOURCE&fl=id,formatId

for a list of just pids:

https://cn.dataone.org/cn/v1/query/solr/?rows=1000&q=datasource:urn\:node\:USGSCSAS%20id:resourceMap*%20-formatType:RESOURCE&fl=id&wt=csv

It looks to be 304 documents. Some of these might also be on the archived list. Can you confirm this list looks good? If so, we will proceed with updating the formatId for these documents - from 'application/octet-stream' to '<http://www.openarchives.org/ore/terms>'.

Ranjeet indicated all these resource maps are to be archived, so updating the formatId for these are not priority. The new resource maps that need formatId updated have DOI identifiers - so I could not locate them by naming scheme in the search index.

#3 - 2013-06-21 17:35 - Ranjeet Devarakonda

- Assignee changed from Ranjeet Devarakonda to Skye Roseboom

All the PID's you sent should be archived (as we added DOI's to all resourceMaps).

Here are the PIDs that should have the formatType changed to "<http://www.openarchives.org/ore/terms>"
https://cn.dataone.org/cn/v1/query/solr/?rows=1000&q=datasource:urn\:node\:USGSCSAS%20id:resourceMap_doi*&fl=id&wt=csv

Also, I won't be able to set the PID's to archived state..I'd need your help in doing this directly at the CN.

#4 - 2013-06-21 17:59 - Skye Roseboom

- File pids to retype to ORE.txt added

- File pids to retype to ORE.txt added

Created text file from solr query provided by Ranjeet: pids_to_retype_to_ORE.txt

#5 - 2013-06-21 18:03 - Skye Roseboom

- Assignee changed from Skye Roseboom to Chris Jones

Issue is complicated by the fact that the original documents were not copied to the CN as they were typed as a 'data' type document. When these documents are modified to be an ORE, they need to copy to the CN for indexing. The current CN operational process does not allow formatID to be changed - and so the backdoor method of changing the formatID in system metadata on the CN, does not fully effect the change needed - that the document now be copied to the CN.

Just fyi, if the documents had been 'obsoleted' by the MN with new documents with proper formatID, this issue would not exist. The new documents would be copied to the CN and indexed and the old documents would remain as 'data' type documents.

#6 - 2013-06-24 22:37 - Skye Roseboom

- Assignee changed from Chris Jones to Skye Roseboom

- Status changed from New to In Progress

#7 - 2013-06-24 22:37 - Skye Roseboom

- Status changed from In Progress to Testing

#8 - 2013-06-24 22:38 - Skye Roseboom

- Status changed from Testing to In Review

#9 - 2013-06-24 22:39 - Skye Roseboom

This issue should be closed.

These documents were always originally uploaded as ORE packages. The issue is that the ORE document contents do not seem to specify existing science metadata documents. Will open new issue to capture the ORE content issue.

#10 - 2013-06-25 14:26 - Skye Roseboom

- Status changed from In Review to Closed

- translation missing: en.field_remaining_hours set to 0.0

Files

pids to retype to ORE.txt	8.67 KB	2013-06-21	Skye Roseboom
---------------------------	---------	------------	---------------