

Member Nodes - Task #4140

MNDeployment # 3118 (Operational): Dryad Member Node

ORE documents reference pids that do not appear in object list

2013-10-29 20:05 - Skye Roseboom

Status:	Closed	Start date:	2013-10-29
Priority:	Normal	Due date:	
Assignee:	Skye Roseboom	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	Operational		
Story Points:			
Description			
After 10/28 test run:			
https://dev.datadryad.org/mn/object/?count=10&formatId=http://www.openarchives.org/ore/terms reports: 1931			
https://cn-dev-ucsb-1.test.dataone.org/cn/v1/object/?count=10&formatId=http://www.openarchives.org/ore/terms reports: 1913			
https://cn-dev-ucsb-1.test.dataone.org/cn/v1/query/solr/?q=formatType:RESOURCE%20datasource:urn\:/node\:/mnTestDRYAD&fl=id&rows=0 reports: 1254			
While most ORE documents synchronized to the CN, it appears about 1/3 of the successfully synchronized ORE - about 650 have not indexed.			
Triaging some of the ORE that have not indexed, it appears they are referencing identifiers that to not present on the dryad /object list endpoint:			
ORE http://dx.doi.org/10.5061/dryad.5p6b1?format=d1rem&ver=2012-06-29T12:01:50.683-04:00 missing: http://dx.doi.org/10.5061/dryad.5p6b1/1/bitstream			
ORE http://dx.doi.org/10.5061/dryad.8164?format=d1rem&ver=2011-09-02T13:08:10.676-04:00 missing: http://dx.doi.org/10.5061/dryad.8164/5/bitstream http://dx.doi.org/10.5061/dryad.8164/2/bitstream http://dx.doi.org/10.5061/dryad.8164/3/bitstream http://dx.doi.org/10.5061/dryad.8164/4/bitstream http://dx.doi.org/10.5061/dryad.8164/6/bitstream			
ORE http://dx.doi.org/10.5061/dryad.8m8r1?format=d1rem&ver=2012-06-26T15:24:29.813-04:00 missing: http://dx.doi.org/10.5061/dryad.8m8r1/1/bitstream http://dx.doi.org/10.5061/dryad.8m8r1/2/bitstream http://dx.doi.org/10.5061/dryad.8m8r1/3/bitstream			
ORE http://dx.doi.org/10.5061/dryad.f385721n?format=d1rem&ver=2012-06-09T13:47:30.984-04:00 missing: http://dx.doi.org/10.5061/dryad.f385721n/1/bitstream http://dx.doi.org/10.5061/dryad.f385721n/1?ver=2012-06-09T13:17:44.181-04:00			
I do not yet see a pattern. It appears the dryad /object list may not be presenting all identifiers in use by the ORE docs. All the reported 'missing' identifiers can be found via dryad's /meta/{pid} and /object/{pid} services but I do not find them when using /object?formatId= or just object list slicing.			

History

#1 - 2013-11-19 07:00 - Ryan Scherle

We determined that most of these errors were due to items with restricted access. These items were correctly filtered out of the object lists, but not out of the resource maps. We have improved the filtering to make the resource map and object list consistent.

A few other items were the result of dirty data on our development server.

Please re-test.

#2 - 2013-11-19 07:03 - Ryan Scherle

- Assignee changed from Ryan Scherle to Skye Roseboom

#3 - 2013-11-19 23:28 - Skye Roseboom

- Status changed from New to In Progress

#4 - 2013-11-19 23:29 - Skye Roseboom

- Status changed from In Progress to Testing

#5 - 2013-11-22 21:02 - Skye Roseboom

- Assignee changed from Skye Roseboom to Ryan Scherle

Test run 11/21.

<https://dev.datadryad.org/mn/object/?count=0&formatId=http://www.openarchives.org/ore/terms>

-- shows 1932

<https://cn-dev-ucsb-1.test.dataone.org/cn/v1/query/solr/?q=datasource:urn\:\node\:\mnTestDRYAD%20formatType:RESOURCE>

-- shows 1586

So we got a couple hundred more ORE to index.

Inspecting ORE that did not process, seem to indicate this problem still exists:

ORE: <http://dx.doi.org/10.5061/dryad.505?format=d1rem&ver=2011-07-28T14:52:47.474-04:00>

references data files:

<http://dx.doi.org/10.5061/dryad.505/1/bitstream>

<http://dx.doi.org/10.5061/dryad.505/2/bitstream>

however neither of these pids were harvested by the CN and neither appear on dev.datadryad.org's object list although /meta{pid} and /object/{pid} work.

ORE: <http://dx.doi.org/10.5061/dryad.389?format=d1rem&ver=2011-07-28T15:13:04.871-04:00>

references data file:

<http://dx.doi.org/10.5061/dryad.389/1/bitstream>

however: <https://dev.datadryad.org/mn/object/?count=1000&formatId=text/html> yields 0 results. (does not appear on object list). /meta{pid} and /object/{pid} work.

ORE: <http://dx.doi.org/10.5061/dryad.509?format=d1rem&ver=2011-07-28T14:53:53.030-04:00>

references data file:

<http://dx.doi.org/10.5061/dryad.509/1/bitstream>

however this pid does not seem to appear on dev.datadryad.org's object list. /meta{pid} and /object/{pid} work however.

Seems to be the same type of error as before. Are these particular ORE meant to be working?

#6 - 2013-11-22 21:02 - Skye Roseboom

- *Status changed from Testing to In Progress*

#7 - 2013-12-12 03:31 - Ryan Scherle

- *Assignee changed from Ryan Scherle to Skye Roseboom*

We have corrected an inconsistency in the object list. Please re-test this.

#8 - 2014-02-03 16:03 - Laura Moyers

- *Target version changed from Deploy by end of Y5Q2 to Deploy by end of Y5Q3*

#9 - 2014-03-14 18:22 - Laura Moyers

- *Target version changed from Deploy by end of Y5Q3 to Operational*

#10 - 2014-03-14 21:29 - Skye Roseboom

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

- *Status changed from In Progress to Closed*

Closing this issue as Dryad is now in production and these issues appear resolved.