

Member Nodes - Task #3917

MNDeployment # 3118 (Operational): Dryad Member Node

ORE documents reference science metadata identifiers that do not exist

2013-08-13 22:18 - Skye Roseboom

Status:	Closed	Start date:	2013-08-13
Priority:	Normal	Due date:	
Assignee:	Skye Roseboom	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	Operational		
Story Points:			
Description			
Similar to issue documented here: https://redmine.dataone.org/issues/3854#note-4			
Dryad's ORE seem to define science metadata identifiers that do not exist for the 'dryadDataPackage' files:			
Example: ORE uses '&' in the sci-metadata's identifier - lines 6,24,28,30 http://dev.datadryad.org/mn/object/http://dx.doi.org/10.5061/dryad.8570?format=d1rem&ver=2011-09-02T16:12:04.933-0400 line 30: http://dx.doi.org/10.5061/dryad.8570&ver=2011-09-02T16:12:04.933-0400/dcterm:identifier			
Although dryad meta and object paths respond to both versions of the identifier (one with '&' and one with '?'): https://dev.datadryad.org/mn/meta/http://dx.doi.org/10.5061/dryad.8570&ver=2011-09-02T16:12:04.933-0400 https://dev.datadryad.org/mn/meta/http://dx.doi.org/10.5061/dryad.8570?ver=2011-09-02T16:12:04.933-0400			
It appears only the '?' version is provided via the object list service: (search for 'dryad.8570') https://dev.datadryad.org/mn/object/?formatId=http://datadryad.org/profile/v3.1&count=100&start=400			
Corroborated by the CN - the '?' (%3F) version exists, the '&' version (%26) does not: https://cn-dev-unm-1.test.dataone.org/cn/v1/meta/http://dx.doi.org/10.5061/dryad.8570%3Fver=2011-09-02T16:12:04.933-0400 https://cn-dev-unm-1.test.dataone.org/cn/v1/meta/http://dx.doi.org/10.5061/dryad.8570%26ver=2011-09-02T16:12:04.933-0400			
So it would seem the ORE is using the wrong identifier. This ultimately prevents the CN indexing process from properly indexing the data package.			

History

#1 - 2013-08-13 22:19 - Skye Roseboom

- Description updated

#2 - 2013-08-13 22:20 - Skye Roseboom

- Description updated

#3 - 2013-09-24 20:46 - Ryan Scherle

- Status changed from New to Closed
- translation missing: en.field_remaining_hours set to 0.0

The Dryad dev server has been updated to correct this issue.

#4 - 2013-10-01 20:05 - Skye Roseboom

- Estimated time set to 0.00
- Status changed from Closed to In Progress

Re-opening on 10/1/13

Cleared CN-dev environment and re - synchronized/indexed dev.dryad.org MN. Synchronization overall appeared to run well. However indexing ORE ran into an issue that prevents ORE from indexing due to mis-aligned timestamp argument on document identifiers:

It appears ORE documents contain references to identifiers that do not appear on the dryad's object list response and thus do not appear to exist to the CN infrastructure. Issue is with the timestamp value that appears at the end of the identifiers:

A couple examples:

ORE: <http://dx.doi.org/10.5061/dryad.4f1r5vg8?format=d1rem&ver=2012-01-13T11:42:38.588-0500>

contains references to pids:

<http://dx.doi.org/10.5061/dryad.4f1r5vg8/1/bitstream>

<http://dx.doi.org/10.5061/dryad.4f1r5vg8/1?ver=2012-01-13T11:42:38.588-0500>

<http://dx.doi.org/10.5061/dryad.4f1r5vg8?ver=2012-01-13T11:42:38.588-0500>

<http://dx.doi.org/10.5061/dryad.4f1r5vg8?format=d1rem&ver=2012-01-13T11:42:38.588-0500>

However:

<http://dx.doi.org/10.5061/dryad.4f1r5vg8/1?ver=2012-01-13T11:42:38.588-0500>

does not appear in object list on dev.dryad.org. This pid does:

<http://dx.doi.org/10.5061/dryad.4f1r5vg8/1?ver=2011-12-17T15:07:22.498-0500>

This appears to be the same document but with slightly different timestamp. Could not find the version as it appears in the ORE on the object list.

Another example:

ORE: <http://dx.doi.org/10.5061/dryad.8671?format=d1rem&ver=2011-09-02T16:26:38.662-0400>

references pids:

<http://dx.doi.org/10.5061/dryad.8671/1/bitstream>

<http://dx.doi.org/10.5061/dryad.8671/1?ver=2011-09-02T16:26:38.662-0400>

<http://dx.doi.org/10.5061/dryad.8671?ver=2011-09-02T16:26:38.662-0400>

However:

<http://dx.doi.org/10.5061/dryad.8671?ver=2011-09-02T16:26:38.662-0400>

does not appear to in dev.dryad's object list - while:

<http://dx.doi.org/10.5061/dryad.8671/1?ver=2011-07-03T12:04:03.127-0400>

does. These appear to be the same document with different timestamps.

These errors appear to be systematic - it does not appear that any ORE documents were able to index successfully due to references to documents/pids that are not found.

#5 - 2013-10-23 15:26 - Ryan Scherle

- Assignee changed from Ryan Scherle to Skye Roseboom

This issue should now be fixed.

#6 - 2013-10-29 14:51 - Skye Roseboom

Resource Maps/ ORE docs are processing through indexing properly now.
Closing issue.

#7 - 2013-10-29 14:52 - Skye Roseboom

- *Status changed from In Progress to Closed*

#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*