

Member Nodes - Task #3595

MNDeployment # 3552 (Closed): USGS CSAS

Unresolvable content on CN from USGS Clearinghouse

2013-02-19 00:59 - Skye Roseboom

Status:	Closed	Start date:	2013-02-19
Priority:	Normal	Due date:	
Assignee:	Chris Jones	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			
Description			
Noticed that some pids that have been harvested by the production CN for USGS Clearinghouse MN are not resolving: www1.usgs.gov_metadata_mdata_NPS_nps_d_metaapisfield.xml www1.usgs.gov_metadata_mdata_NPS_VegMap_nps_d_metabandbdy.xml www1.usgs.gov_metadata_mdata_NPS_nps_d_metablcaaa.xml www1.usgs.gov_metadata_mdata_NPS_VegMap_nps_d_metabrcabdy.xml Meta data records are served, for example: http://mercury-ops2.ornl.gov/clearinghouse/mn/v1/meta/www1.usgs.gov_metadata_mdata_NPS_VegMap_nps_d_metabrcabdy.xml however the /object endpoint does not: http://mercury-ops2.ornl.gov/clearinghouse/mn/v1/object/www1.usgs.gov_metadata_mdata_NPS_VegMap_nps_d_metabrcabdy.xml Appears that pids beginning with 'www1' are effected. Need to determine if these pids are no longer being used and should be archived or whether this content is just temporarily missing.			
Related issues:			
Duplicated by Infrastructure - Task #3789: USGSCSAS query failure to retrieve...		Closed	2013-06-05

History

#1 - 2013-02-22 18:03 - Skye Roseboom

- Target version set to Operational

#2 - 2013-02-22 18:03 - Skye Roseboom

- Parent task changed from #3568 to #3552

- Assignee deleted (Giri Palanisamy)

- Category deleted (296)

- Project changed from Infrastructure to Member Nodes

#3 - 2013-04-23 19:16 - Skye Roseboom

- Assignee set to Ranjeet Devarakonda

Talked with Ranjeet and Chris Jones today:

Decided that USGS CSAS will set the archive flag/element in the system metadata for the objects that are no longer available and also update the system metadata update time ('dateSysMetadataModified') - so these records are harvested to the dataone CN.

Once this is done, lets confirm the changes have made it to the CN (objects should be removed from search index)

#4 - 2013-04-23 19:20 - Ranjeet Devarakonda

Test

#5 - 2013-05-09 14:32 - Skye Roseboom

- File *To_be_archived.txt* added
- File *To_be_archived.txt* added

Moved this list from redmine 3725 to this issue which is tracking pids that are no longer used by USGS Clearinghouse MN

#6 - 2013-05-09 14:35 - Skye Roseboom

- Assignee changed from *Ranjeet Devarakonda* to *Chris Jones*

Ranjeet supplied this list of pids that should be archived from the USGS Clearinghouse MN. The system metadata no longer exists on the MN so the change needs to be made at the CN.

These pids can be run through the script which makes system metadata changes in hazelcast directly to archive.

#7 - 2013-06-21 23:32 - Chris Jones

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

I've updated all of the system metadata on the CNs for the pids in the attached file and have set them to be archived. I'll follow up with SKye to ensure these aren't in the Solr index on the CNs.

#8 - 2013-06-24 22:18 - Skye Roseboom

- File *USGSCSAS-NON-ORE-TO-BE-ARCHIVED.pids* added
- Estimated time set to *0.00*
- File *USGSCSAS-NON-ORE-TO-BE-ARCHIVED.pids* added

attached a second list of pids to be archived for USGS CSAS. this time its a list of pids that start with resource*, are from usgs, but are not typed as an ORE

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

300 pids

#9 - 2013-06-24 22:19 - Skye Roseboom

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

Re-opened with new list of pids to archive for USGS

NOT A COMPLETE LIST OF PIDS TO ARCHIVE.....this is just a list of 'old' resource maps to remove.

#10 - 2013-06-25 15:27 - Chris Jones

I've archived the pids in the new list added by Skye, so I'll close this ticket again. If there are other issues with USGS archival, just reopen again. I'll confirm with Skye that the objects are not indexed.

#11 - 2013-06-25 15:27 - Chris Jones

- Status changed from In Progress to Closed

#12 - 2013-06-25 16:04 - Skye Roseboom

- File USGS_to_archive_nrdata.txt added
- File USGS_to_archive_www1.txt added
- File USGS_to_archive_nrdata.txt added
- Status changed from Closed to In Progress
- File USGS_to_archive_www1.txt added

Adding two more text files containing pids to archive.

66 pids to archive that are from USGS CSAS and have pids that start with 'nrdata'
https://cn-ucsb-1.dataone.org/cn/v1/query/solr/?rows=1000&q=datasource:urn\:node\:USGSCSAS%20id:nrdata*&fl=id

237 pids to archive that are from USGS CSAS and have pids that start with 'www1'
https://cn-ucsb-1.dataone.org/cn/v1/query/solr/?rows=1000&q=datasource:urn\:node\:USGSCSAS%20id:www1*&fl=id

These pids no longer exist on the source MN.

#13 - 2013-06-25 16:32 - Chris Jones

- Status changed from In Progress to Closed

All pids look to be archived now that don't exist on the MN. Closing.

Files			
To_be_archived.txt	5.26 KB	2013-05-09	Skye Roseboom
USGSCSAS-NON-ORE-TO-BE-ARCHIVED.pids	15.2 KB	2013-06-24	Skye Roseboom
USGS_to_archive_nrdata.txt	2 KB	2013-06-25	Skye Roseboom
USGS_to_archive_www1.txt	9.55 KB	2013-06-25	Skye Roseboom