Member Nodes - Task #4295

MNDeployment # 3221 (Operational): EDAC member node

address discovered content issues in the index

2014-03-04 20:55 - Rob Nahf

Status:	Closed	**Start date:**	2014-03-04	
Priority:	Normal	**Due date:**		
Assignee:	Rob Nahf	**% Done:**	100%	
Category:		**Estimated time:**	0.00 hour	
Target version:	Operational			
Story Points:				

Description

duplicate and triplicates were found in STAGE ONEMercury:

There are I think three sets of science metadata that looks like duplicates.
An example of a pair can be seen in one-mercury:
https://cn-stage-orc-1.test.dataone.org/onemercury/send/facetsQuerry2?term1=*&term1attribute=text&op1=&term3attribute=overlaps&term3=%2C%2C%2C&op3=&term6attribute=datasource&op6=+OR+&term6=urn%5C%3Anode%5C%3AEDACGSTORE&term8=collection&pageSize=10&start=0&sortattribute=default&facetattribute=author&facet=New%20Mexico%20Tech,%20Department%20of%20Earth%20and%20Environmental%20Science

Maybe a triplicate here:
https://cn-stage-orc-1.test.dataone.org/onemercury/send/facetsQuerry2?term1=*&term1attribute=text&op1=&term3attribute=overlaps&term3=%2C%2C%2C&op3=&term6attribute=datasource&op6=+OR+&term6=urn%5C%3Anode%5C%3AEDACGSTORE&term8=collection&pageSize=10&start=0&sortattribute=default&facetattribute=keywords&facet=Aerial%20Compliance

I think these are what Soren is seeing. And there does look like a synchronization error occurred. From the first link, the duplicates have identifiers:
08d2c688-19fd-4cd5-88ff-4ed76bf40332
62140ca7-2ad7-42b3-a447-c3a2e539fea6

On gstore.unm.edu the first identifier is obsoleted
https://gstore.unm.edu/dataone/v1/meta/08d2c688-19fd-4cd5-88ff-4ed76bf40332
but the cn isn't reflecting this:
https://cn-stage-orc-1.test.dataone.org/cn/v1/meta/08d2c688-19fd-4cd5-88ff-4ed76bf40332
further, the identifier that obsoletes the first:
https://gstore.unm.edu/dataone/v1/meta/71fd14aa-8c7e-4e89-a3f3-1df2a011db2a
does not appear on the cn at all:
https://cn-stage-orc-1.test.dataone.org/cn/v1/meta/71fd14aa-8c7e-4e89-a3f3-1df2a011db2a - yields NotFound

So it seems the obsolete chain that would remove the 'duplicate' in the first example isn't being reflected on the CN - and there are some items seem missing.

In the second example the identifiers shown in order of obsoletes chain are:
f23fbc28-8fda-45da-b201-bb8c584fb273
610eb76c-4bea-496d-bbd5-a0afc70f6c9b
ff54d33c-38a6-4158-a7cb-a7974a91b480

With f23fbc28-8fda-45da-b201-bb8c584fb273 as the start of the chain and ff54d33c-38a6-4158-a7cb-a7974a91b480 as the most recent/tail of the chain. But the proper obsoletes chain is not reflected on the CN with the 'obsoletedBy' value unset on the first 2 identifiers - causing the duplicates to appear in oneMercury.

Rob - Maybe resetting the harvest date for the edac-gstore node, and restarting d1-processing to trigger a re-sync would pick up those differences in obsolete chain and add the missing doc.

Related issues:

Related to Infrastructure - Bug #6895: Complete obsolescence chain is being d...	**Rejected**	**2015-03-17**

History

#1 - 2014-03-04 20:56 - Rob Nahf

- Description updated

#2 - 2014-03-04 21:00 - Rob Nahf

- % Done changed from 0 to 100

- translation missing: en.field_remaining_hours set to 0.0

- Status changed from New to Closed

after registering GSTore in DEV and reharvesting, checked situations listed (followed links substituting cd-dev fro cn-stage-orc-1) and not finding duplicates or triplicated in DEV.

Conclude that these issues were artifacts of incremental fixes and harvesting while we were shaking out content issues in STAGE.

#3 - 2014-05-01 15:51 - Laura Moyers

- Target version changed from Deploy by end of Y5Q3 to Deploy by end of Y5Q4

#4 - 2014-06-30 17:52 - Laura Moyers

- Target version changed from Deploy by end of Y5Q4 to Operational

#5 - 2015-03-17 18:29 - Mark Servilla

- Related to Bug #6895: Complete obsolescence chain is being displayed for LTER content in ONEMercury added