Member Nodes - Task #3980

MNDeployment # 3558 (Operational): CDL Merritt

Missing content

2013-09-17 20:04 - Skye Roseboom

Status:	New		**Start date:**	2013-09-17
Priority:	Normal		**Due date:**	
Assignee:	Chris Jones		**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:	Operational			
Story Points:				

Description

Started noticing some 404 - NotFound errors from CDL/Merrit MN today.   It appears that the previous content of documents has been removed from the MN. Appears to correspond with rollout of corrected ORE/RDF resource maps -   this is a guess based on the pid structure and content of the following example and the date of the 'new' version of content.

These two documents appear to have the same content:
https://cn.dataone.org/cn/v1/object/ark:/13030/m5416x9f/1/cadwsap-s2010005-011.xml
https://cn.dataone.org/cn/v1/object/ark:/13030/m5416x9f/2/cadwsap-s2010005-011.xml

however the first (with the '1' before 'cadwsap') does not appear on the source MN:
https://merritt.cdlib.org:8084/knb/d1/mn/v1/object/ark:/13030/m5416x9f/1/cadwsap-s2010005-011.xml
Although the second does ('2'):
https://merritt.cdlib.org:8084/knb/d1/mn/v1/object/ark:/13030/m5416x9f/1/cadwsap-s2010005-011.xml
Similar response for /meta requests on the same pids.

Another example (this time an RDF/ORE):
https://cn.dataone.org/cn/v1/meta/ark:/13030/m5s46rvg/1/mrt-dataone-map.rdf
https://cn.dataone.org/cn/v1/meta/ark:/13030/m5s46rvg/2/mrt-dataone-map.rdf

Same pattern - first version with the ('1') does not appear on MN:
https://merritt.cdlib.org:8084/knb/d1/mn/v1/meta/ark:/13030/m5s46rvg/1/mrt-dataone-map.rdf
although the second version ('2') does appear on the MN:
https://merritt.cdlib.org:8084/knb/d1/mn/v1/meta/ark:/13030/m5s46rvg/2/mrt-dataone-map.rdf

The old content is still present on the CN due to it not being 'archived'. This results in these documents continuing to appear in the CN search index and object list - although they no longer appear to exist on the source MN.

Need to discuss possible solutions. Possibly generate list of 'old' pids to be archived at the CN to cleanup content that no longer appears in the MN.

History

#1 - 2013-09-25 21:46 - Skye Roseboom

- File Merrit-PIDs.txt added

- File OneShare-PIDs.txt added

- File Merrit-PIDs.txt added

- File OneShare-PIDs.txt added

Adding file listing of pids that should be archived for Merritt/CDL and ONESHare.

#2 - 2013-09-25 21:48 - Skye Roseboom

- Assignee changed from John Kunze to Chris Jones

Chris can we 'archive' these pids for Merritt and ONEShare with the script you created?

Files

Merrit-PIDs.txt	4.01 MB	2013-09-25	Skye Roseboom
OneShare-PIDs.txt	2.66 KB	2013-09-25	Skye Roseboom