Member Nodes - Task #7081

MNDeployment # 3552 (Closed): USGS CSAS

Archive CSAS content (metadata)

2015-05-06 18:45 - Laura Moyers

Status:	Closed		**Start date:**	2016-03-11
Priority:	Normal		**Due date:**	
Assignee:	Mark Servilla		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:				
Story Points:				

Description
USGS CSAS (or the Clearinghouse) has gone away, so we need to archive its content. Plan how to and who will execute this. CSAS doesn't hold content but provided metadata with links out to the content held elsewhere. With archival, the metadata is not discoverable via ONEMercury, but given a DOI, a user may still access the content at it's original host location. (This next is a USGS action) For National Park Service content, redirect to a "tombstone" indicating that the content is no longer available via this route.

Subtasks:	
Task # 7672: Un-approve the CSAS MN in production	**Closed**
Task # 7673: Remove the CSAS MN from the MN dashboard	**Closed**

History

#1 - 2015-09-21 20:56 - Laura Moyers

- % Done changed from 0 to 30

- Status changed from New to In Progress

UPDATE: (ref https://epad.dataone.org/pad/p/USGS_CSAS_and_SDC)

Originally (9/9/15), we were going to get sysmeta from CN for all content, update sysmeta with archive=true, change systmeta authoritative MN to one of the dedicated Replication Nodes, then delete/remove CSAS.

The current plan (9/21/15) is to get sysmeta from CN for all content, ARCHIVE all content and set CSAS's status to "down". This ensures that, if anyone had accessed content and has the DOI, they may still access it even though CSAS is "down". However, a MN that is "down" still shows up in the list of MNs in ONEMercury, which might be misleading. If the MN is "unapproved", then content will not resolve.

The proposed NEW plan (later 9/21/15) is to get sysmeta from CN for all content, ARCHIVE all content, REPLICATE all content at a MN/RN willing to host the content, change authoritative MN to the host of the replicates, then change CSAS's status to "down" and "unapprove" it.

See documentation:
https://releases.dataone.org/online/api-documentation-v1.2.0/apis/MN_APIs.html#MNStorage.archive

USGS CSAS (Clearinghouse) - proposed sequence of events to archive CSAS's (metadata) content:
All content, including National Park Service content, should be replicated somewhere. A logical target would be one of the replication nodes such as mnORC1. The Member Node (CSAS) will change the system metadata for each object to indicate that it is to be replicated and specify a "preferred" target.

replication policy in sysmeta for each object
setReplicationPolicy call on CN, with the MN cert (
https://releases.dataone.org/online/api-documentation-v1.2.0/apis/CN_APIs.html#CNReplication.setReplicationPolicy);

replication policy payload is an XML document of the type Types.ReplicationPolicy (
https://releases.dataone.org/online/api-documentation-v1.2.0/apis/Types.html#Types.ReplicationPolicy) - call getSystemMetadata on CN to get the most current serial version of the sysmeta

The authoritative Member Node for the content should be changed to the replication target where it now resides. This is specified in the system metadata for each object. Under the current API, the Coordinating Nodes "own" system metadata; therefore, DataONE administrator (maybe Skye) should make this change. Under API version 2.0, the Member Nodes will "own" the system metadata for their content, and the Member Node administrator would make this change.

Summary on DataONE Object States:
https://docs.google.com/document/d/1u6L8kmGRjcQ94DmWIbee9Jn46mYVdrXs_n2_ZOczYFE

#2 - 2015-11-13 16:07 - Laura Moyers

In conversations with Mike Frame, we think we want to archive CSAS content to the new SDC MN. Currently SDC is a Tier 1 MN, so we'll have to evaluate the ramifications of taking SDC to Tier 4.

#3 - 2016-02-11 15:55 - Laura Moyers

After further discussion, this is the plan of attack:

• First, ARCHIVE all CSAS content so that it is undiscoverable.
• Second, if desired, change the ACCESS POLICY on each object so that it is non-public, so that in the very slight chance that someone has an identifier, if they should try to resolve it, they'd get a "not authorized" message. This is up to Mike/Ranjeet as the CSAS MN owner.
• Lastly, UNAPPROVE the CSAS MN in production so that it does not show up on lists (such as the list of MNs in search, the dashboard, etc.)

Some of the content is National Park Service vegetation data, which the NPS has asked the Clearinghouse (and hence the CSAS MN) to not serve up.

A minority of the content on CSAS has a later, updated version at SDC. We thought about indicating an obsolescence chain for this content, crossing MNs, but the more straightforward process is simply to archive the CSAS content and un-approve the CSAS MN.

#4 - 2016-03-11 15:17 - Dave Vieglais

Content was archived 2016-03-04. Required use of the CN client certificate and also enabling the MNStorage API in the node entry.

#5 - 2016-03-21 15:17 - Laura Moyers

- % Done changed from 30 to 100

- Status changed from In Progress to Closed

All tasks related to deprecating the CSAS MN are complete.