

## Member Nodes - Task #8124

MNDeployment # 3186 (Deprecated): USA-NPN (USA National Phenology Network)

### NPN MN deprecation activities

2017-07-07 13:54 - Laura Moyers

<b>Status:</b>	Closed	<b>Start date:</b>	2018-06-18
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Mark Servilla	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Story Points:</b>			
<b>Description</b>			
<p>NPN (urn:node:USANPN) has decided to close their DataONE MN. In future they plan to deposit their content at a USGS repository, likely the Science Data Catalog. SDC currently has a DataONE MN (urn:node:USGS_SDC), so any NPN content deposited there will be discoverable by DataONE users. The general plan for the deprecation of the NPN MN is to ensure that all content is replicated at KNB (urn:node:KNB) who has graciously agreed to host/own the content in future and that the authoritative MN is changed from NPN to KNB on all content. When the content is officially hosted at KNB, the NPN MN may be officially shut down.</p> <p>NPN's content varies from "trying things out" to valid, production content. Most (all?) of the "trying things out" content has been obsoleted by improved/corrected content. There was a successful replication policy in place for the content, some of which is already replicated at KNB. As of 7/7/17, there are seven data packages discoverable in production DataONE (<a href="https://search.dataone.org/index.html#profile/USANPN">https://search.dataone.org/index.html#profile/USANPN</a>).</p> <p>There are 38 NPN objects to be managed. Mark and Roger have done a thorough analysis of the content to determine each object's current disposition and what action(s) must be taken. Mark's analysis follows with a link to detailed, by-object information:</p> <hr/> <p>There are a total of 38 objects from USANPN that exist as replicated objects on the CN or on other MNs. Of these 38 objects, 19 are science metadata objects, 13 are data objects, and 6 are OREs.</p> <p>The desired outcome is to have all 38 objects replicated to KNB and have the "authoritative MN" attribute in the object's system metadata be set to "urn:node:KNB". Sixteen of the 38 objects have already been replicated on KNB through the normal DataONE replication process.</p> <p>The proposed approach to replicate the remaining 22 objects is to manually create the objects on the KNB member node and, once all 38 objects are replicated to KNB, set their "authoritative MN" attribute to KNB accordingly. Object creation for objects not currently replicated on KNB should use the corresponding file system object in the NPN source directory. In addition, all replica metadata in the system metadata for each object should be updated to indicate that the now defunct "urn:node:USANPN" member node does not contain a viable replica.</p> <p>Of the 38 objects, two science metadata objects have incorrect size/checksum values, which must be corrected by manually by modifying system metadata in both the KNB (for doi:10.5066/F7028PJS.1) and CN (for doi:10.5066/F7028PJS.1 and 10.5066/F7833Q2V.0) databases; any replicas should be marked for deletion from the non-KNB/CN replication target member node and then re-replicated to meet replication policy criteria.</p> <p>Object creation on KNB may have to be time-ordered based on existing obsolescence. One of either Jing Tao, Bryce Mecum, Chris Jones, or Matt Jones is recommended to perform object creation and Metacat database modifications on KNB.</p> <p>One of either Jing Tao, Rob Nahf, or Dave Vieglais is recommended to perform the Metacat database modifications on the CN. Please see the following Google Spreadsheet for detailed information about individual objects: <a href="https://docs.google.com/spreadsheets/d/1a5oRV6_OVc6O0g2HWfzQK7bWqLpPmsVC_4XvQpBI3Zo/edit?usp=sharing">https://docs.google.com/spreadsheets/d/1a5oRV6_OVc6O0g2HWfzQK7bWqLpPmsVC_4XvQpBI3Zo/edit?usp=sharing</a>.</p> <hr/> <p>For the two objects with size/checksum mis-matches, we originally thought we might be able to manually modify the size/checksum on either KNB or the CNs, but size and checksum are immutable. We need to examine other options for managing this content.</p> <p>The suggested order of operations is (at present):</p> <ul style="list-style-type: none"><li>- deal with the mis-matched objects</li><li>- add objects not currently on KNB to the KNB repository/MN</li><li>- for all objects, change the authoritativeMN from NPN to KNB</li><li>- for all "replicas", indicate that NPN is no longer viable (i.e. the "replica" hosted there is no longer available)</li></ul>			

When all NPN objects have been transferred to KNB, notify Lee Marsh at NPN, and "turn off" the NPN MN in production.

**Subtasks:**

Task # 8620: NPN: move to contributing repository

**Closed**

**History**

**#1 - 2018-01-22 18:59 - Amy Forrester**

**epad Notes Consolidation**

11/12/17: Process to replicate to EDI is now in progress; completing testing of GMN 2.4.0 in EDI prior to completion of deprecation

11/20: Preparation for replication to EDI now in process.

11/27: Process to replicate to EDI is now in progress; completing testing of GMN 2.4.0 in EDI prior to completion of deprecation

\* Need future meeting to discuss use cases of MN status and how to display them on Member Node dashboard - Dave will set up this call

12/18: Testing for content replication in other MNs

**#2 - 2018-02-21 20:42 - Mark Servilla**

It was suggested that existing USANPN content may perhaps already be replicated on another MN as duplicate content, but under the guise of a different identifier. If this were the case, then there should be multiple records in the CN Metacat systemmetadata table with identical checksum values, but with different identifiers and origin member nodes (assuming the same checksum algorithm was used; USANPN was not consistent with checksum algorithms and utilized both MD5 and SHA-1).

Each checksum was queried in the Metacat systemmetadata table to determine if any object duplication occurred. There were no multi-node duplicates, but 3 objects from USANPN were duplicates of themselves and identified with different identifiers:

1. USA\_NPN.3.1 == USANPN\_SPECIES\_LIST1.1
2. USA\_NPN.6.1 == USANPN\_PHENOPHASE\_LIST.1
3. USANPN\_2012.0.0 == USANPN\_2013.0.0

**#3 - 2018-04-09 18:00 - Mark Servilla**

31 of 38 have been successfully replicated to urn:node:EDI, along with resetting the authoritative MN to urn:node:EDI. The following objects would not allow a replica to be acknowledged in the CN context:

- USA\_NPN.3.1
- USA\_NPN.5.3
- USA\_NPN.6.1

In addition, there is still the issue of incorrect checksums and their related sibling objects to be addressed for the following:

- 10.5066/F7028PJS.0
- doi:10.5066/F7028PJS.1 (checksum error)
- doi:10.5066/F7028PJS.2
- 10.5066/F7833Q2V.0 (checksum error)

#### #4 - 2018-04-16 12:04 - Mark Servilla

New attempt to set replica for USA\_NPN.3.1 fails with exception:

```
[ INFO] 2018-04-16 12:01:18,131 [ProcessDaemonTask2] (SyncObjectTask:executeTransferObjectTask:293) Task-urn:node:EDI-USA_NPN.3.1 received
[ INFO] 2018-04-16 12:01:18,132 [ProcessDaemonTask2] (SyncObjectTask:executeTransferObjectTask:310) Task-urn:node:EDI-USA_NPN.3.1 submitted for execution
[ INFO] 2018-04-16 12:01:18,132 [SynchronizeTask18242] (V2TransferObjectTask:call:202) Task-urn:node:EDI-USA_NPN.3.1 - Locking task, attempt 1
[ INFO] 2018-04-16 12:01:18,178 [SynchronizeTask18242] (V2TransferObjectTask:call:207) Task-urn:node:EDI-USA_NPN.3.1 - Processing SyncObject
[ INFO] 2018-04-16 12:01:18,396 [SynchronizeTask18242] (V2TransferObjectTask:retrieveMNSystemMetadata:317) Task-urn:node:EDI-USA_NPN.3.1 - Retrieved SystemMetadata Identifier:USA_NPN.3.1 from node urn:node:EDI for ObjectInfo Identifier USA_NPN.3.1
[ INFO] 2018-04-16 12:01:19,817 [SynchronizeTask18242] (V2TransferObjectTask:createObject:730) Task-urn:node:EDI-USA_NPN.3.1 - Start CreateObject
[ INFO] 2018-04-16 12:01:21,133 [SynchronizeTask18242] (V2TransferObjectTask:call:234) Task-urn:node:EDI-USA_NPN.3.1 - Unlocked Pid.
[ERROR] 2018-04-16 12:01:21,133 [SynchronizeTask18242] (V2TransferObjectTask:call:269) Task-urn:node:EDI-USA_NPN.3.1 - UnrecoverableException: USA_NPN.3.1 cn.createObject failed: The identifier is already in use by an existing object. - InvalidRequest - The identifier is already in use by an existing object.
[ WARN] 2018-04-16 12:01:21,133 [SynchronizeTask18242] (SyncFailedTask:submitSynchronizationFailed:116) Task-urn:node:EDI-USA_NPN.3.1 - SynchronizationFailed: detail code: 6001 id:USA_NPN.3.1 nodeId:urn:node:CNUCSB1 description:Synchronization task of [PID::] USA_NPN.3.1 [::PID] failed. Cause: InvalidRequest: The identifier is already in use by an existing object.
[ INFO] 2018-04-16 12:01:21,288 [SynchronizeTask18242] (V2TransferObjectTask:call:294) Task-urn:node:EDI-USA_NPN.3.1 - exiting with callState: FAILED
[ INFO] 2018-04-16 12:02:15,324 [SynchronizationQuartzScheduler_Worker-4] (SyncQueueFacade:<init>:143) org.da.taone.cn.batch.synchronization.type.SyncQueueFacade@5404dc3a added 'urn:node:EDI' to its queue round-robin. size: 0
[ INFO] 2018-04-16 12:02:18,267 [ProcessDaemonTask2] (SyncObjectTask:reapFutures:372) Task-urn:node:EDI-USA_NPN.3.1 SyncObjectState: FAILED
```

#### #5 - 2018-06-18 19:00 - Amy Forrester

- Assignee changed from Dave Vieglais to Mark Servilla

#### #6 - 2018-06-18 19:08 - Amy Forrester

6/4/18: call

• Christopher Jones

- Mark Servilla
- Amy Forrester
- Jeanette Clark

Alyssa Rosemartin reached out to Chris et al. regarding updating their data in KNB. Internal discussion to share info

Chris to address:

\* Work needed to clean up replicas on KNB (maybe others)

\* communicate to NPN:

1. can still contribute new data to KNB, but KNB will be authoritative MN
2. Data in EDI will be static & EDI is authoritative MN of that data

**6/18: \*need to follow-up with Chris** status + where is the link break re: KNB not updating authoritative MN -- CN issue?

**#7 - 2018-08-17 14:12 - Amy Forrester**

- % Done changed from 0 to 100

- Status changed from New to Closed