

## Member Nodes - Task #8585

MNDeployment # 3228 (Operational): NCEI - National Centers for Environmental Information

### GMN not syncing obsoletedBy field correctly

2018-05-03 17:52 - Matthew Jones

<b>Status:</b>	Closed	<b>Start date:</b>	2018-05-03
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Roger Dahl	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Story Points:</b>			

#### Description

It seems the GMN node for NCEI has been having trouble with obsoletedBy fields and at a minimum needs to be upgraded. Upgrade the node to the newest GMN, and ensure that the objects are syncing and marked as obsoleted correctly. An email thread detailing the issues is included below.

Hi Ken,

I think we should have upgraded the GMN version for the NCEI Tier 1 instance at DataONE long ago. Sorry this didn't get taken care of. I will talk to Dave and Roger about upgrading ASAP and tracking down this issue,

Matt

On Thu, May 3, 2018 at 9:35 AM, Kenneth Casey - NOAA Federal <kenneth.casey@noaa.gov> wrote:

Matt, Xiaoyan -

Another 6 months has passed and the problem of showing > 50k records is still there. What do you guys think we should do about it? We have made more recent progress standing up the new tier 4 node but it as you know it still isn't operational.

Ken

On Fri, Oct 6, 2017 at 3:58 PM, Matt Jones <jones@nceas.ucsb.edu> wrote:

Hi Xiaoyan,

The new MN instance that you are working on developing would be based on the newest releases of GMN, and, if Roger's analysis turns out to be accurate, therefore would fix the synchronization issue that is being seen with the existing deployment. However, given that progress on the new node is proceeding fairly slowly, it may be prudent to upgrade the existing NCEI Tier 1 MN to the newest version of GMN and work to solve the synchronization problem now rather than wait for the new node to come to production status. In general I think it is a good idea to keep production repositories upgraded to the most recent version of the software so that these fixed issues get taken care of.

Matt

On Fri, Oct 6, 2017 at 10:47 AM, Sheekela Baker-Yeboah - NOAA Affiliate <sheekela.baker-yeboah@noaa.gov> wrote:

John/Ken and Matt can weigh in on that note.

Many thanks,  
Sheekela

-----  
-----  
Sheekela Baker-Yeboah, Ph.D. NOAA/NESDIS/NCWCP/SCOD 5830 University Research  
Ct., College Park, MD 20740 Phone: 301-683-3382  
Sheekela.Baker-Yeboah@noaa.gov

Assistant Research Scientist /Oceanographer  
University of Maryland Computer, Mathema  
tical, and Natural Sciences (CMNS) Earth System Science Interdisciplinary Center (ESSIC)  
College Park, MD 20740-3823 sbakerye@umd.edu

On Fri, Oct 6, 2017 at 2:37 PM, Xiaoyan Li - NOAA Affiliate <xiaoyan.li@noaa.gov> wrote:

Hi, Sheekela and Matt,

Yuanjie reported an mismatch issue for NCEI slender node, may it be fixed in the ARCTI  
C project's next step to create a connector between new NCEI GMN/OneSTOP?  
Please give us some advice.

Thanks!  
Xiaoyan

----- Forwarded message -----

From: Yuanjie Li <yuanjie.li@noaa.gov>  
Date: Fri, Oct 6, 2017 at 10:45 AM  
Subject: Re: NCEI MN  
To: Roger Dahl <dahl@unm.edu>  
Cc: "Moyers, Laura Burton (Laura)" <lmoyers1@utk.edu>, John Relph <john.relph@noaa.gov  
>, Mark Servilla <mark.servilla@gmail.com>, Monica Ihli <email@monicaihli.com>, Xiaoyan Li - NOAA  
Affiliate <xiaoyan.li@noaa.gov>, Kenneth Casey - NOAA Federal <kenneth.casey@noaa.gov>, Scott Cros  
s <scott.cross@noaa.gov>

Hi Roger,

On Fri, Oct 6, 2017 at 9:50 AM, Roger Dahl <dahl@unm.edu> wrote:

Hi Li,

The mismatch appears to be due to obsolescence status having failed to be properly  
synchronized from the NCEI member node to the coordinating node.

There is a total of 50,972 objects on the NCEI MN but many of these are obsoleted  
objects that should not show up in the search result. As a randomly selected example, here is the  
system metadata for the object with identifier {135C612E-0C57-4450-B052-95C4733A1E4A} on the NCEI  
MN:

<https://ncei-node.dataone.org/mn/v1/meta/%7B135C612E-0C57-4450-B052-95C4733A1E4A%7D>

D

While this is the system metadata for the same object on the CN:

<http://cn.dataone.org/cn/v2/meta/%7B135C612E-0C57-4450-B052-95C4733A1E4A%7D>

The system metadata on the MN contains an obsoletedBy field, which indicates that  
this object is obsoleted and should not be included in search results. Yet, the system metadata on  
the CN lacks the obsoletedBy field for this object, causing it to be included. From your expectat  
ion of seeing around 25,000 objects in the search result, this probably affects around half of the  
50,972 objects.

The version of GMN (the DataONE Generic Member Node) that runs on the NCEI MN is o  
utdated and is known to have an issue related to the synchronization of the obsoletedBy field. Wha

t I see on the NCEI MN does not completely fit with how I thought the known issue affected the obs oletedBy field, so it's possible that this is caused by something unrelated, but we can start out by handling this as if it's caused by the known issue and then investigate further if necessary. As the known issue is resolved in the current version of GMN, there is a good chance that your search result can be fixed by upgrading to the current version and triggering an extra synchronization of the existing objects.

Thanks for taking care of the issue! Please let us know if you need any information from us.

Related to upgrading GMN for NCEI, I'm wondering if you would be able to give us a brief overview of NCEI's plans for the new MN? As you may know, DataONE developed the NCEI MN that is currently in production and is hosting it as well. The main part of this work was creating what we call a Slender Node adapter that imports data from the CSW service to an instance of GMN. The NCEI MN currently in production is not a replication target. Do I understand correctly that the new MN will be both a replication target and a DataONE interface for CSW?

Xiaoyan is setting up the GMN software to run here at NCEI-MD, to enable the ocean archive system to be a Tier 4 member node. I am cc-ing this message to her.

Thanks,  
Li

Thank you,

Roger

On 10/05/2017 09:21 AM, Yuanjie Li wrote:

Thanks Laura!

Hi Roger,  
I got 50,484 records by searching for "NOAA NCEI Oceanograp...". The total number of records should be around 25000. Please let me know if you still use the CSW to get the metadata, and when was last time the metadata was harvested?

Thanks a lot,  
Li

## History

---

### #1 - 2018-05-07 16:02 - Amy Forrester

- Tracker changed from MNDeployment to Task

per roger 5/7/18: I'm working on NCEI upgrade and metadata fix. Hopefully will be done today.

### #2 - 2018-05-14 14:37 - Amy Forrester

- % Done changed from 0 to 100

- Status changed from New to Closed

per Roger 5/14/18