Member Nodes - Story #8683

MNDeployment # 7082 (Operational): USGS Science Data Catalog (SDC)

USGS SDC: redeploy as a v2 Slender Node with GMN

2018-08-22 16:25 - Amy Forrester

Status:	New	**Start date:**	2018-03-28
Priority:	Normal	**Due date:**	
Assignee:	John Evans	**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:			
Story Points:			

Description

Subtasks:

Task # 8526: duplicate data sets for USGS node + archiving **Closed**

History

#1 - 2018-08-22 17:40 - Amy Forrester

8/22/18: (Monica) Redeployment Notes:
- USGS is still working out their implementation of identifiers with their source repositories
- However, USGS will begin implementation of an OAI-PMH endpoint at this time in order to keep the project moving forward.
- USGS will also move on installing the GMN software .
- USGS is interested in deploying to sandbox first as a test environment to prototype their handling of identifiers, but this is to be done with the expectation that they are still developing their handling of identifiers.

#2 - 2018-09-25 14:12 - Monica Ihli

- Goal for this week: archive USGS existing data due to being almost all bad links and bogus identifiers.
- Notify them to take their server offline.
- Do we need to take them out of the node list in search interface? Because what will happen is someone could filter by clicking on their node but get zero results since everything is archived.

#3 - 2018-09-25 14:16 - Amy Forrester

- Description updated

#4 - 2018-09-25 14:44 - Amy Forrester

- Description updated

#5 - 2018-09-25 16:51 - Amy Forrester

UPDATE from Lisa Zolly

So, the plan with our Catalog is to use metadata to generate actual landing pages for each item, rather than just a pop-up modal for the metadata and links off to data residing in various repositories and catalogs. We will use the DOI to generate the unique landing page for each data product, and that DOI will be the unique identifier in our Catalog, and part of the URL to a given landing page.

This will happen over the course of FY19. Our first task is going to be assigning DOIs to the 7500+ items in our Catalog that are legacy (pre-2016) data releases and which don't currently have DOI assigned to them. This will have to be done in batches, and in close consultation with the centers who own the data, because we need to use the metadata to generate the DOIs, and then we will have to insert the DOIs into the metadata records, and pass those records back to the originating centers for them to re-host. This will ensure that our updated copies are in synch, so that they can

continue to manage them long-term for our weekly reharvesting. Of course, it's going to be a more complicated process, because each of these 7 centers hosting their own metadata that our Catalog harvests do their metadata management and their data management differently, and it's going to take us much of the fiscal year to get all of this done. The end goal by next summer is to have a Catalog in which every data product has a DOI that we can leverage as a unique identifier/unique landing page for the Catalog's index. Downstream catalogs such as data.gov and DataONE can then index the unique landing pages (which will link to the full metadata), and utilize more harvest-friendly versions of the metadata, instead of the full CSDGM.

For DataONE's purposes, our timeline is probably 9 months away at best, which is probably not the greatest news, given that it puts us close to the end of the current project; however, I think that the solution we're moving towards will reduce the number of issues we've had with harvesting and with the transform to EML, because the main record we pass will be lighter, and not subject to vagaries of imperfect transforms to EML or to ISO. The full metadata will - somehow...we're still working it out - be accessible for users who want the deeper details after the more basic bibliographic information has sufficiently piqued their interest.

We may be down at ORNL in early December, or late January, to have an in-person working meeting, so we will loop you in (and probably Monica, too) once we get it scheduled to come over to the lab and get the full picture. By that point, we will have details firmed up and will have the mechanics nailed down, even though it will take us most of the FY to enable and transition from the existing Catalog framework.

#6 - 2019-05-15 14:03 - Amy Forrester

- Assignee changed from Monica Ihli to John Evans

5/16/19: meeting scheduled - John, Roger, Aaron, Amy
per Lisa -

> mention the work we're trying to do SDC 3.0 to determine if we will have to work with the DataONE team again after we redesign the backend. If there's going to be significant levels of effort to get 2.0 back online now as a MN, and a similar level of effort once 3.0 deploys, both sides may want to evaluate availability of resources to do this twice. I'm not saying we won't commit resources to get us back online with DataONE short-term, but it might be good to evaluate, so that no one is caught off-guard when we deploy SDC at the end of this year {fingers crossed}.

ePad Meeting notes