Member Nodes - Story #8589

MNDeployment # 3230 (Planning): ARM - Atmospheric Radiation Measurement member node

ARM: Re-Discovery & Planning

2018-05-10 20:35 - Amy Forrester

Status:	New		**Start date:**	2018-05-10
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:				
Story Points:				

Description
5/10/2018: Meeting at ORNL - Giri Prakash expressed re-interest in becoming a MN. Following on USGS_SDC GMN implementation (~Fall 2018), Aaron Stokes can turn attention to ARM installation

Subtasks:	
Task # 8590: ARM: Requirements Analysis	**New**

History

#1 - 2018-05-10 20:53 - Monica Ihli

ARM provides streaming data. They cited their contribution to data.gov as an example of a workflow they are happy with, in which what they submit is really a metadata record representing the entire stream itself, which is periodically updated. The link provided in the metadata record sends users to a landing page for the stream as a whole, from which users select specific and other parameters for the specific interval of the stream they would like to retrieve.

For example, this:

https://catalog.data.gov/dataset/c-scanning-arm-precipitation-radar-csapr-vertical-scan

Leads to:

http://www.archive.arm.gov/discovery/#v/results/s/s::csaprvert

This approach accomplishes two things:

(1) There is one metadata record for each stream (approximately 11,000 streams to be exact), which prevents there from being a lot records as components which comprise the interval, and this is their preference. It is understood that data reproducability urges the ability to cite exactly the specific interval of data used in a study, and there is nothing to stop an end user from doing just that based on whatever interval they select. But because data intervals are dynamically generated by ARM, it does not suit ARM's business model to attempt to predict what those search parameters should be. It would be entirely arbitrary to break it into chunks just for inclusion into DataONE.

(2) By sending people to the landing page for the stream, ARM is able to comply with DOE imposed mandates in which they are required to collect certain information about those who download the data. Users are directed to create an account or login when they are ready to download. ARM is not permitted to replicate data or provide a link directly to the data in which users could bypass this requirement.

A discussion of implementation considerations:

- Would like their contractor Aaron to complete USGS re-implementation in OAI-PMH for USGS. With already having that experience, he would be in a good position to complete a similar implementation for ARM.
- The preferred approach would be that they provide a metadata record in ISO for each data stream as a whole, with the link in the metadata record leading to the steam landing page on ARM system.
- A stream-level identifier would serve as the SID in this case.
- Each month (or however often they wished) they could update the metadata record with a new version that includes an updated "end" for the time interval of data represented by the metadata record. Each version would have its own pid and be assigned the same stream-level identifier as SID.

#2 - 2018-05-10 21:44 - Rob Nahf

Impressive site. I saw that the user is able to generate a citation for the stream. Does that create a snapshot of the stream that ARM or the user may wish to preserve and expose to DataONE?

That is, could DataONE become the snapshot repository?