

Member Nodes - Task #8360

MNDeployment # 8254 (Planning): ABDS - Arctic Biodiversity Data Service (CAFF)

Story # 8358 (In Progress): Discovery & Planning (CAFF)

Feasibility Assessment

2018-02-12 16:34 - Amy Forrester

Status: Closed	Start date: 2018-02-12
Priority: Normal	Due date:
Assignee: Dave Vieglaiss	% Done: 100%
Category:	Estimated time: 0.00 hour
Target version:	
Story Points:	

Description
Initial determination of operational and technical feasibility. Addresses:

Operational Feasibility:

- Repository meets sustainability expectations.

Technical Feasibility:

- **What data will be exposed?**
- **Content mutability.**
- **Software needs:**
- DataONE software stack or implement new web service in existing applications (custom stack)?
- Full repository or Slender Node?
- metadata only or science data + replication?
- Repository's in-house preferences? (Java? Python?)
- **Consistency of metadata quality.**
- **Level of functionality (tier).**

Approval = Change Status > Planning

History

#1 - 2018-03-09 17:39 - Amy Forrester

- Assignee set to Dave Vieglaiss

#2 - 2018-03-09 17:43 - Amy Forrester

Meeting scheduled: March 28 @10:30am ET [Dave, Monica, Amy, Tom Barry]

** MEETING POSTPONED **

#3 - 2018-04-06 14:51 - Amy Forrester

- Description updated

#4 - 2018-04-06 15:03 - Amy Forrester

4/6/18 Meeting https://epad.dataone.org/pad/p/CAFF_and_DataONE

CAFF: Tom Barry (tom@caff.is)

DataONE: Amy, Dave, Monica

Actions:

* Tom to respond to several questions via email

- NodeID acronym
- which services are enabled in the Geonet server?
- Identifiers - change with edits or at all over time
- double identifiers

- Next connect with Design Plan

NOTE: Tom out of office May-June

#5 - 2018-04-06 15:19 - Amy Forrester

- Status changed from New to In Progress
- % Done changed from 0 to 30

#6 - 2018-04-06 16:00 - Monica Ihli

We expect to work with the geonetwork server as it exposes all the data.

Requirements to Consider:

- * strategy for avoiding duplicate content.
- * detecting metadata format (it can differ depending on legacy or more current)
- * The metadata format can change over time as legacy data is brought to current standards.
- * Does geonetwork provide reliable identifiers for content?

Seems everything originates from the geonetwork server. Things that go to GBIF get picked up by IPT. The content at the point of IPT is not associated with any kind of distinguishing identifier outside of its IPT endpoint, which appears to be arbitrary. Content harvested into GBIF via IPT are assigned a DOI by GBIF once the content is at GBIF. It does not appear that the DOI makes it back up the channel. This would mean being unable to easily recognize when something is already in GBIF. Need to confirm with their data manager how to best approach this.

Most likely will be Slender Node but hopefully data + metadata. We did not recall to bring up this matter during call on 4/6 but will promote replication at next opportunity.

As a turnkey solution geonetwork is bundled with capabilities for exposing content such as CSW (<http://geo.abds.is/geonetwork/srv/eng/csw?request=GetCapabilities&service=CSW&version=2.0.2>) (<http://www.opengeospatial.org/standards/cat>)

Follow-up points:

- * Example geonetwork resource has 2 identifiers
<http://geo.abds.is/geonetwork/srv/eng/catalog.search#/metadata/ee276820-caa1-44ad-93bc-2a3a99ed49d6>
- * Preference for endpoint? CSW?
- * Confirm with data manager if metadata or data identifiers change over time with edits.
- * Tom out of office May - June.

#7 - 2018-04-11 14:38 - Amy Forrester

- % Done changed from 30 to 100
- Status changed from In Progress to Closed

Move responses to Planning documentation