

Infrastructure - Story #8853

Make cn.resolve smarter

2019-11-15 16:46 - Jing Tao

Status:	New	Start date:	2019-11-15
Priority:	Normal	Due date:	
Assignee:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			

Description

In this case the cn.resolve() operation should be ignoring the node that is marked as offline, or at least placing it last in the list.

This should be a high priority fix, and should be fairly simple to implement since the information is available in the node document.

- Dave

On 2019-11-14, at 21:38, Matt Jones jones@nceas.ucsb.edu wrote:

FYI, thread from today with Ethan White on ebird replication, and the resolve() api in DataONE. Relates to our conversation today about making resolve() and MetacatUI downloads smarter.

Matt

Ethan White 5:06 PM
What's the right place to report data that if 404ing on DataONE?

Matt Jones 5:07 PM
support@dataone.org would work

5:08 PM
or let me know

5:08 PM
is it that same data set?

5:08 PM
the Ebird one?

Ethan White 5:09 PM
Yeah, which we had discovered had been reposted and spent a bunch of time gearing up to support again. We were in the middle of testing when it suddenly disappeared again.
http://dataone.ornith.cornell.edu/metacat/d1/mn/v2/object/EOD_CLO_2016.csv.gz

Matt Jones 5:10 PM
yeah. Cornell just gave us permission to replicate the data to other nodes. They haven't wanted us to do so in the past.

Ethan White 5:13 PM
Thanks. That's good news. So can we expect it to reappear at some point soonish?

Matt Jones 5:14 PM
Yeah, its been replicated. I'm checking to see if it is properly linked to the original.

5:15 PM
https://knb.ecoinformatics.org/view/EOD_CLO_2016.eml

new messages

Ethan White 5:16 PM
Thanks Matt. FYI that link I posted is the one being returned from a current search of DataONE.

Matt Jones 5:17 PM

Yeah. Because that's the 'authoritative' copy at cornell.

5:17 PM

but Cornell's node has been going up and down.

5:17 PM

our resolve service lists all copies of a data set

5:17 PM

so if one is down, you can get it from another location:

5:18 PM

```
$ curl -H "Accept: text/xml" https://cn.dataone.org/cn/v2/resolve/EOD_CLO_2016.eml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ns2:objectLocationList xmlns:ns2="http://ns.dataone.org/service/types/v1">
<identifier>EOD_CLO_2016.eml</identifier>
<objectLocation>
<nodeIdentifier>urn:node:CLOEBIRD</nodeIdentifier>
<baseURL>http://dataone.ornith.cornell.edu/metacat/d1/mn</baseURL>
<version>v1</version>
<version>v2</version>
<url>http://dataone.ornith.cornell.edu/metacat/d1/mn/v2/object/EOD_CLO_2016.eml</url>
</objectLocation>
<objectLocation>
<nodeIdentifier>urn:node:CN</nodeIdentifier>
<baseURL>https://cn.dataone.org/cn</baseURL>
<version>v1</version>
<version>v2</version>
<url>https://cn.dataone.org/cn/v2/object/EOD_CLO_2016.eml</url>
</objectLocation>
<objectLocation>
<nodeIdentifier>urn:node:KNB</nodeIdentifier>
<baseURL>https://knb.ecoinformatics.org/knb/d1/mn</baseURL>
<version>v1</version>
<version>v2</version>
<url>https://knb.ecoinformatics.org/knb/d1/mn/v2/object/EOD_CLO_2016.eml</url>
</objectLocation>
</ns2:objectLocationList>
```

Ethan White 5:19 PM

OK, thanks. That's why I thought the link in DataONE https://cn.dataone.org/cn/v2/resolve/EOD_CLO_2016.csv.gz would take me to a working version, but clearly I just don't understand the details. We'll just use the the one on KNB at least for the moment. Really appreciate your help as always.

Matt Jones 5:20 PM

No problem. I'd love to make this all work more seamlessly. (edited)

5:20 PM

So suggestions definitely welcome.

5:21 PM

I expect Cornell to take their node offline altogether — so the KNB will likely be the better location.

5:22 PM

Btw, the resolve link when executed in a browser just redirects to the first copy

Ethan White 5:23 PM

Yeah, Cornell's closed approach to things is a pretty big disappointment, especially on data like this that is generated by volunteers. We'll just go to the KNB version permanently.

Matt Jones 5:23 PM

whereas programatically you get the list of locations

5:23 PM

if you ask for XML

Ethan White 5:23 PM
That makes sense. Thanks.

Matt Jones 5:23 PM
and then you can choose to try one or more

History

#1 - 2019-11-15 16:50 - Matthew Jones

- *Description updated*