

## Infrastructure - Bug #2365

### formatId not accurately reported by getSystemMetadata

2012-02-22 15:02 - Dave Vieglais

|                         |                          |                        |            |
|-------------------------|--------------------------|------------------------|------------|
| <b>Status:</b>          | Closed                   | <b>Start date:</b>     | 2012-02-22 |
| <b>Priority:</b>        | High                     | <b>Due date:</b>       |            |
| <b>Assignee:</b>        | Ben Leinfelder           | <b>% Done:</b>         | 100%       |
| <b>Category:</b>        | Metacat                  | <b>Estimated time:</b> | 0.00 hour  |
| <b>Target version:</b>  | Sprint-2012.07-Block.1.4 | <b>Story Points:</b>   |            |
| <b>Milestone:</b>       | CCI-1.0.0                |                        |            |
| <b>Product Version:</b> | *                        |                        |            |

#### Description

On cn-dev.dataone.org, pids XYZ33256 and BAYXXX\_015ADCP015R00\_20051215.50.9 report having formatIds of eml://ecoinformatics.org/eml-2.0.1 in the SystemMetadata table and are listed as such when executing a CNRead.listObjects call.

However, when executing a CNRead.getSystemMetadata, the formatIds are represented as application/octet-stream.

The objects were created with the same EML document. I don't know if this is an issue, but it appears the xml\_relation table has the different docids (BAYXXX\_015ADCP015R00\_20051215.50 & autogen.2012022107103964186) pointed to the same object entry BAYXXX\_015ADCP015R00\_20051215.40

#### Original description:

It appears that the formatId appearing in system metadata is being altered at some point by CN processing. Perhaps during synchronization?

For example:

On the origin MN:

<https://gm-dev.test.dataone.org/mn/v1/meta/XYZ33256>

On the CN:

<http://cn-dev.dataone.org/cn/v1/meta/XYZ33256>

In this case the source formatId is eml://ecoinformatics.org/eml-2.0.1

After processing it is changed to: application/octet-stream

#### History

##### #1 - 2012-02-22 19:51 - Robert Waltz

- Subject changed from formatId apparently being replaced during synchronization to formatId not accurately reported by getSystemMetadata
- Category changed from d1\_synchronization to Metacat
- Assignee changed from Robert Waltz to Ben Leinfelder

##### #2 - 2012-02-22 20:36 - Ben Leinfelder

One immediate difference I see is that the Metacat listObjects implementation queries the backing DB tables directly for the System Metadata it returns whereas the getSystemMetadata method uses the HZ map. It is conceivable that something changed the values in the map and that this is not [yet?] reflected in the backing store (postgres tables). I thought we'd gone out of our way to immediately write any updates to the in-memory map directly to the DB tables.

I know it is very easy to directly edit the DB table values and not see this reflected in the map, but I thought the other way around was impossible.

So, what is CN processing doing to this record? Is it more than just this record that has the issue?

**#3 - 2012-02-22 21:25 - Ben Leinfelder**

*- Status changed from New to Closed*

Metacat was incorrectly performing a lookup for the formatId it had stored for the SM record. The look ups were not returning a format and so it defaulted to application/octet-stream  
Updated knb.war in cn-buildout - should be fixed on the next redeploy.