

Metacat - Bug #5523

setting authoritative MN and rightsHolder for KNB data on conversion

10/28/2011 02:16 PM - Matt Jones

Status:	Resolved	Start date:	10/28/2011
Priority:	Immediate	Due date:	
Assignee:	ben leinfelder	% Done:	0%
Category:	metacat	Estimated time:	0.00 hour
Target version:	2.0.0	Spent time:	0.00 hour
Bugzilla-Id:	5523		

Description

When EML and data in the NCEAS and LTER metacat is converted to use in the new KNB MN, we need to generate system metadata for each object.

For rightsHolder, I propose we use the current xml_documents.user_owner field.

For authoritativeNode, we should set that to the ID of either the KNB MN or the LTER Member Node (of course, we'll have to know what these node IDs are) as appropriate. This can probably be determine from the home server entry in our replication tables.

History

#1 - 12/08/2011 03:13 PM - ben leinfelder

We can certainly find out our local KNB MN Id since it will be configured, but I'm not sure how we would go about definitively finding out what the LTER MN is. Also, wont they be in the same boat for us? And then what do you suppose happens when we both have entries for each other's documents? Should the SystemMetadata already include entries for Replicas since they exist on the other server? And what are the replication policies supposed to be for the system metadata we generate on our newly minted MN?

#2 - 12/08/2011 04:15 PM - ben leinfelder

On further reflection, I think we might only want to be generating SystemMetadata for objects that were added directly to the KNB, not via replication. Otherwise we are essentially sharing everyone else's data. In terms of policy, would we be "allowed" to disperse other objects? (I'm thinking of Spain and Brazil which may or may not become their own MNs in the near future).

#3 - 12/12/2011 01:49 PM - ben leinfelder

More musing on the authoritative MN/replica policy for upgrading Metacat deployments:

~~If we only generate SystemMetadata for Metacat objects that are owned by the home server, then DataONE has no knowledge that we have those replicas from other Metacat instances.~~

~~When another formerly replicating Metacat comes online as a MN, their objects may be sent to our MN so that we can store them as replicas. If we have no existing SystemMetadata for those replicate objects, then we will end up storing 2 copies of the same object with different IDs-- one will be Metacat-only, and the other will have SystemMetadata (as a replica) and a normal Metacat docid.~~

~~I don't think we want to duplicate all objects that we had previously received with Metacat's legacy replication process. This means we should probably generate SystemMetadata information for objects that we don't directly "own" on the home Metacat server. But how should we identify the authoritative MN? We won't necessarily know which MN in the nodeList to use-- sure we can be smart about matching hostnames, but what about MNs that are not yet registered but that will be? We can't expect every old replication partner to be registered at the time we bring a DataONE-enabled Metacat MN online. Can we leave it blank? Does that make any sense even if it were possible?~~

#4 - 12/21/2011 01:53 PM - ben leinfelder

We decided to generate system metadata for every object housed in the KNB and use the KNB nodeId as the authoritative member node.

When a replication partner (like the LTER) becomes a MN, we will update the system metadata for every object we have that came from the LTER so that it uses LTER's nodeId as the authoritative membernode.

#5 - 03/27/2013 02:30 PM - Redmine Admin

Original Bugzilla ID was 5523