Infrastructure - Task #3109

PISCO/LTER revision history with SBC documents

2012-07-31 00:06 - Ben Leinfelder

Status:	In Progress	**Start date:**	2012-07-30
Priority:	Normal	**Due date:**	
Assignee:	Ben Leinfelder	**% Done:**	30%
Category:		**Estimated time:**	0.00 hour
Target version:			
Milestone:	None	**Story Points:**	
Product Version:	*		

Description

Margaret O'Brien has pointed out some unexpected documents ("duplicates") showing up in the ONEMercury index.
It seems that many of her SBC data packages began on the PISCO node at an early revision and then they decided to move ownership (in the legacy Metacat replication sense) from the PISCO Metacat to the LTER Metacat.
As I understand it, this was a manual SQL update that switched the serverid value in the xml_documents/revision table from PISCO to LTER (replication of the documents had already taken place).
If this were done properly, the following (obsolete) object should not be included in the index and should probably me marked with SystemMetadata.archived=true. Moreover, it probably should have been included in the serverid switch so that it was not considered part of PISCO's holdings. But it looks like PISCO generated a DOI for the 7.1 revision and then generated SystemMetadata for it.

On the CN:
https://cn.dataone.org/cn/v1/meta/doi:10.6085/AA/knb-lter-sbc.7.1

On the PISCO MN (note only SM for 7.1, not 7.2 which to me indicates that SystemMetadata was only generated for 7.1 because it was "owned" by the PISCO node whereas 7.2 is likely owned by LTER):
https://data.piscoweb.org/catalog/d1/mn/v1/meta/doi:10.6085/AA/knb-lter-sbc.7.1
https://data.piscoweb.org/catalog/metacat/knb-lter-sbc.7.1/default
https://data.piscoweb.org/catalog/metacat/knb-lter-sbc.7.2/default

You can see the revision graph on LTER for these documents starting with 7.2 (there is no rev 1 SystemMetadata, but note that it is in the Metacat catalog, likely as a replica from PISCO):
https://tropical.lternet.edu/knb/metacat/knb-lter-sbc.7.1/default
https://tropical.lternet.edu/knb/d1/mn/v1/meta/doi:10.6073/AA/knb-lter-sbc.7.2
https://tropical.lternet.edu/knb/d1/mn/v1/meta/doi:10.6073/AA/knb-lter-sbc.7.3
https://tropical.lternet.edu/knb/d1/mn/v1/meta/doi:10.6073/AA/knb-lter-sbc.7.4

So, we should craft an SQL query and update statement that correctly marks these 'archived' objects as appropriate. My hunch is that PISCO does not have 7.x in it's xml_documents table and therefore should not have contributed it as unarchived. We will have to coordinate with Mike and Mark to see exactly what each node has in it's Metacat tables. For this example docid, the KNB thinks that versions 1-4 all came from LTER. I'm curious to see what the PISCO node has.

Related issues:

Related to Infrastructure - Bug #6895: Complete obsolescence chain is being d...	**Rejected**	**2015-03-17**

History

#1 - 2012-07-31 01:34 - Ben Leinfelder

- Status changed from New to In Progress

Added a bit of script to find (and update) records on an MN that should be marked as archived but are not (and have not been obsoleted_by another pid).
https://code.ecoinformatics.org/code/metacat/trunk/src/system_metadata_revisions.sql

#2 - 2015-03-17 18:29 - Mark Servilla

- Related to Bug #6895: Complete obsolescence chain is being displayed for LTER content in ONEMercury added