

## Infrastructure - Task #3138

Story # 3136 (Closed): Inconsistent data in production 1.0.3 release

### Identify and correct missing xml\_documents or xml\_revisions table entries on ORC, UNM and UCSB

2012-08-15 16:10 - Robert Waltz

<b>Status:</b>	Closed	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Ben Leinfelder	<b>% Done:</b>	100%
<b>Category:</b>	Environment.Production	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Sprint-2012.29-Block.4.3	<b>Story Points:</b>	
<b>Milestone:</b>	CCI-1.0.4		
<b>Product Version:</b>	*		
<b>Description</b>			
<p>The identifiers table entries are missing corresponding entries in xml_documents or xml_revisions table. If there is an entry in systemMetadata table that has a corresponding entry in the identifiers table, then there should be an entry in either the xml_documents or xml_revisions table.</p> <p>Use the following query on the 3 production machines:</p> <pre>su postgres -c "psql -d metacat -c \"select count(guid) from identifier i \ where not exists (select docid, rev from xml_documents x where x.docid = i.docid and x.rev = i.rev \ union select docid, rev from xml_revisions r where r.docid = i.docid and r.rev = i.rev)\""</pre> <p>It will result in 14920 on ORC, 14901 on UNM and 18 on UCSB.</p> <p>The errors on UCSB are related to earlier releases. The problems causing these errors may have already been solved and the data may just need to be corrected on UCSB. No determination has been made about the problems on ORC and UNM, but most of the problems appear to be related to CDL Resource Maps not being replicated correctly.</p>			
<b>Related issues:</b>			
Related to Infrastructure - Task #3140: Missing UCSB xml_revisions		<b>Closed</b>	<b>2012-08-15</b>

## History

### #1 - 2012-08-16 17:26 - Ben Leinfelder

For the missing docs on ORC and UNM, it looks like Metacat replication is timing out when retrieving the list of docs from UCSB.

```
knb 20120816-17:11:20: [ERROR]: ReplicationHandler.update - Failed to get updated doc list for server cn-ucsb-1.dataone.org/knb/servlet/replication because Read timed out [ReplicationLogging]
```

We should up the configured timeout to 5 minutes since it is currently taking 1:02 and will only grow and grow:

```
replication.client.timeout=300000
```

### #2 - 2012-08-16 17:28 - Ben Leinfelder

- Status changed from New to In Progress

I also fixed a problem with metacat's force replication when a data object (ORE map) was inserted in Metacat using the DataONE API. It had been incorrectly using the force replication action with "create" when it should have been "update". This has essentially prevented us from replicating ORE documents on initial insert since day one. The fix will be included in Metacat 1.0.4

**#3 - 2012-08-17 19:03 - Ben Leinfelder**

- Status changed from In Progress to Closed

The ORC and UNM numbers have been reconciled. UCSB remains 18 documents shy -- in a different task.

**Files**

---

identifiersMissingUCSB-Documents.out	701 Bytes	2012-08-15	Robert Waltz
identifiersMissingUCSB-Documents.analysis.txt	15.8 KB	2012-08-15	Robert Waltz
identifiersMissingUNM-Documents.out	626 KB	2012-08-15	Robert Waltz
identifiersMissingORC-Documents.out	627 KB	2012-08-15	Robert Waltz