

Infrastructure - Task #864

Story # 912 (Closed): Fix known TODOs in the metacat code base

Metacat not downloading urls listed in xml\_catalog table

2010-09-17 21:18 - Robert Waltz

<b>Status:</b>	Closed	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Chad Berkley	<b>% Done:</b>	100%
<b>Category:</b>	Metacat	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Sprint-2010.46	<b>Story Points:</b>	
<b>Milestone:</b>			
<b>Product Version:</b>	*		

**Description**

My understanding is that schemas that have a system\_id with an URL should have their schema downloaded from the URL instead of looking for their schema in knb/schema folder.

However, and entry that is created from this insert statement:

```
INSERT INTO xml_catalog (entry_type, public_id, system_id) VALUES ('Schema', 'http://www.loc.gov/METS', 'https://repository.dataone.org/software/cicore/tags/D1_SCHEMA_0_4/dryadMetsAny.xsd')
```

results in the following error:

```
knb 20100917-14:12:13: [WARN]: XMLService.populateRegisteredSchemaList - Schema file:
/var/lib/tomcat6/webapps/knb/schema/dryadMetsAny.xsd is registered in the database but does not exist on the file system.
[edu.ucsb.nceas.metacat.service.XMLSchemaService]
```

History

- #1 - 2010-10-06 16:41 - Chad Berkley
- Parent task set to #912
- #2 - 2010-10-15 19:06 - Chad Berkley
- Status changed from New to In Progress
- #3 - 2010-10-15 21:40 - Chad Berkley
- Status changed from In Progress to Closed

I added code in Metacat to perform this task. I tested with the dryadMetsAny.xsd schema and it worked. If you put a url into the xml\_catalog for a schema, it will attempt to download the schema when a document is parsed. The schema is cached to disk then used for validation. See XMLSchemaService in metacat for implementation details.