Metacat - Bug #5273

docs with inline-data allow invalid xml into metacat

01/13/2011 09:37 AM - Chad Berkley

Status:	Resolved		**Start date:**	01/13/2011
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:	5273			

Description

If you insert a document with inline-data, the data is stripped out of the document before it is validated. However, when you do a GET on the document, it is read off of the disk. So if you insert a doc with inline-data that has invalid characters in it (like unescaped ampersands), metacat will not recognize that it is invalid, but when you try to get the document, you will get a parser error if you try to parse it.

We should be validating the document first before stripping inline-data out of it.

History

#1 - 01/14/2011 09:11 AM - ben leinfelder

is inline data not contained in CDATA? I thought you could put anything in CDATA and have it be ignored by parsers.

#2 - 01/14/2011 10:59 AM - Matt Jones

People certainly can use CDATA sections within their inline element, in which case escaping would be taken care of. But in this case, the data in the CDR document is not in a CDATA element, has reserved XML characters in it, and Metacat is not properly rejecting it as invalid.

#3 - 10/27/2011 06:58 PM - ben leinfelder

This file uses unescaped ampersand (&) in the inline data section.

#4 - 10/27/2011 07:00 PM - ben leinfelder

I tried this with the attached inline.xml file that has an invalid unescaped ampersand in the inline section -- Metacat rejected it as invalid.

If there is a specific CDR file that is causing this issue still, let's reopen and and figure out how it is slipping by. Otherwise, I believe this is not currently a problem in trunk.

#5 - 03/27/2013 02:29 PM - Redmine Admin

Original Bugzilla ID was 5273

Files

inline.xml		2.19 KB	10/28/2011	ben leinfelder