Infrastructure - Story #2759

Metacat should only parse XML documents against local schema files

2012-05-11 22:55 - Chris Jones

Status:	Rejected		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:	Chris Jones		**% Done:**	0%
Category:	Metacat		**Estimated time:**	0.00 hour
Target version:				
Story Points:				

Description

We've seen some instances of MN errors where the SAX parser "can't find a schema" document while parsing, such as:

knb 20120413-02:00:27: [ERROR]: DocumentImpl.write - Problem with parsing:
schema_reference.4: Failed to read schema document 'eml.xsd',
because
1) could not find the document;
2) the document could not be read;
3) the root element of the document is not xsd:schema. [edu.ucsb.nceas.metacat.DocumentImpl]
at edu.ucsb.nceas.metacat.DBSAXHandler.warning(DBSAXHandler.java:751)
at org.apache.xerces.util.ErrorHandlerWrapper.warning(Unknown Source)
at org.apache.xerces.impl.XMLErrorReporter.reportError(Unknown Source)
at org.apache.xerces.impl.XMLErrorReporter.reportError(Unknown Source)
at org.apache.xerces.impl.xs.traversers.XSDHandler.reportSchemaWarning(Unknown Source)
at org.apache.xerces.impl.xs.traversers.XSDHandler.getSchemaDocument1(Unknown Source)
at org.apache.xerces.impl.xs.traversers.XSDHandler.getSchemaDocument(Unknown Source)
at org.apache.xerces.impl.xs.traversers.XSDHandler.parseSchema(Unknown Source)
at org.apache.xerces.impl.xs.XMLSchemaLoader.loadSchema(Unknown Source)
at org.apache.xerces.impl.xs.XMLSchemaValidator.findSchemaGrammar(Unknown Source)
at org.apache.xerces.impl.xs.XMLSchemaValidator.handleStartElement(Unknown Source)
at org.apache.xerces.impl.xs.XMLSchemaValidator.startElement(Unknown Source)
at org.apache.xerces.impl.XMLNSDocumentScannerImpl.scanStartElement(Unknown Source)
at org.apache.xerces.impl.XMLNSDocumentScannerImpl$NSContentDispatcher.scanRootElementHook(Unknown Source)
at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl$FragmentContentDispatcher.dispatch(Unknown Source)
at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument(Unknown Source)
at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)

at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)

at org.apache.xerces.parsers.XMLParser.parse(Unknown Source)

at org.apache.xerces.parsers.AbstractSAXParser.parse(Unknown Source)

at edu.ucsb.nceas.metacat.DocumentImpl.write(DocumentImpl.java:2787)

at edu.ucsb.nceas.metacat.DocumentImpl.write(DocumentImpl.java:2580)

at edu.ucsb.nceas.metacat.DocumentImplWrapper.write(DocumentImplWrapper.java:63)

at edu.ucsb.nceas.metacat.MetacatHandler.handleInsertOrUpdateAction(MetacatHandler.java:1802)

at edu.ucsb.nceas.metacat.dataone.D1NodeService.insertOrUpdateDocument(D1NodeService.java:1130)

at edu.ucsb.nceas.metacat.dataone.D1NodeService.create(D1NodeService.java:393)

at edu.ucsb.nceas.metacat.dataone.MNodeService.replicate(MNodeService.java:530)

at edu.ucsb.nceas.metacat.restservice.MNResourceHandler$1.run(MNResourceHandler.java:662)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1110)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:603)

at java.lang.Thread.run(Thread.java:636)

Caused by: java.io.FileNotFoundException: /var/lib/tomcat6/eml.xsd (No such file or directory)
at java.io.FileInputStream.open(Native Method)

at java.io.FileInputStream.(FileInputStream.java:137)

at java.io.FileInputStream.(FileInputStream.java:96)
at sun.net.[www.protocol.file.FileURLConnection.connect(FileURLConnection.java:87)](www.protocol.file.FileURLConnection.connect)

at sun.net.[www.protocol.file.FileURLConnection.getInputStream(FileURLConnection.java:178)](www.protocol.file.FileURLConnection.getInputStream)

at org.apache.xerces.impl.XMLEntityManager.setupCurrentEntity(Unknown Source)

at org.apache.xerces.impl.XMLVersionDetector.determineDocVersion(Unknown Source)

at org.apache.xerces.impl.xs.opti.SchemaParsingConfig.parse(Unknown Source)

at org.apache.xerces.impl.xs.opti.SchemaParsingConfig.parse(Unknown Source)

at org.apache.xerces.impl.xs.opti.SchemaDOMParser.parse(Unknown Source)

Determine why DocumentImpl.write() is even entertaining an external schema file location other than those locally found at /schema. See XMLSchemaService.populateRegisteredSchemaList() and edu.ucsb.nceas.metacat.service.XMLSchema.setFileName().

History

#1 - 2012-10-05 15:15 - Chris Jones

- Status changed from New to Rejected

- Start date deleted (2012-05-11)

- translation missing: en.field_remaining_hours set to 0.0

Metacat is currently using the local schema file, but if a new schema types are found in science metadata, it will attempt to cache them, which is why we're having issues with 404'd schema locations.