

Infrastructure - Story #2759

Metacat should only parse XML documents against local schema files

2012-05-11 22:55 - Chris Jones

<div>Status:Rejected</div> <div>Priority:Normal</div> <div>Assignee:Chris Jones</div> <div>Category:Metacat</div> <div>Target version:</div> <div>Story Points:</div>	<div>Start date:</div> <div>Due date:</div> <div>% Done:0%</div> <div>Estimated time:0.00 hour</div>
<div>Description</div> <div>We've seen some instances of MN errors where the SAX parser "can't find a schema" document while parsing, such as:</div> <div>knb 20120413-02:00:27: [ERROR]: DocumentImpl.write - Problem with parsing: schema_reference.4: Failed to read schema document 'eml.xsd', because</div> <div>1) could not find the document;</div> <div>2) the document could not be read;</div> <div>3) the root element of the document is not <a href="#">xsd:schema</a>. [edu.ucsb.nceas.metacat.DocumentImpl]</div> <div>at edu.ucsb.nceas.metacat.DBSAXHandler.warning(DBSAXHandler.java:751)</div> <div>at org.apache.xerces.util.ErrorHandlerWrapper.warning(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLErrorReporter.reportError(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLErrorReporter.reportError(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.traversers.XSDHandler.reportSchemaWarning(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.traversers.XSDHandler.getSchemaDocument1(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.traversers.XSDHandler.getSchemaDocument(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.traversers.XSDHandler.parseSchema(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.XMLSchemaLoader.loadSchema(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.XMLSchemaValidator.findSchemaGrammar(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.XMLSchemaValidator.handleStartElement(Unknown Source)</div> <div>at org.apache.xerces.impl.xs.XMLSchemaValidator.startElement(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLNSDocumentScannerImpl.scanStartElement(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLNSDocumentScannerImpl\$NSContentDispatcher.scanRootElementHook(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl\$FragmentContentDispatcher.dispatch(Unknown Source)</div> <div>at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument(Unknown Source)</div> <div>at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)</div> <div>at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)</div> <div>at org.apache.xerces.parsers.XMLParser.parse(Unknown Source)</div> <div>at org.apache.xerces.parsers.AbstractSAXParser.parse(Unknown Source)</div> <div>at edu.ucsb.nceas.metacat.DocumentImpl.write(DocumentImpl.java:2787)</div> <div>at edu.ucsb.nceas.metacat.DocumentImpl.write(DocumentImpl.java:2580)</div> <div>at edu.ucsb.nceas.metacat.DocumentImplWrapper.write(DocumentImplWrapper.java:63)</div> <div>at edu.ucsb.nceas.metacat.MetacatHandler.handleInsertOrUpdateAction(MetacatHandler.java:1802)</div> <div>at edu.ucsb.nceas.metacat.dataone.D1NodeService.insertOrUpdateDocument(D1NodeService.java:1130)</div> <div>at edu.ucsb.nceas.metacat.dataone.D1NodeService.create(D1NodeService.java:393)</div> <div>at edu.ucsb.nceas.metacat.dataone.MNodeService.replicate(MNodeService.java:530)</div> <div>at edu.ucsb.nceas.metacat.restservice.MNResourceHandler\$1.run(MNResourceHandler.java:662)</div> <div>at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1110)</div>	

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