

CN REST - Bug #8733

Exception handling in ForesiteResourceMap needs improvement

2018-10-12 20:09 - Chris Jones

Status:	New	Start date:	2018-10-12
Priority:	Normal	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	d1_cn_index_processor	Estimated time:	0.00 hour
Target version:			
Story Points:			

Description

We're seeing some errors from the org.dataone.cn.indexer.resourcemap.ForesiteResourceMap constructor when indexing resource maps in Metacat 2.9.0, d1_cn_index_processor 2.3.4. A general Exception is being caught and re-thrown as an OREParserException on line 104:

```
metacat 20181012-10:53:02: [INFO]: done queueing doc index for pid urn:uuid:5a860679-159e-45a6-bcb6-beb67f4166c2 [edu.ucsb.nceas.metacat.MetacatHandler:handleReindexAction:2777]
metacat 20181012-10:53:02: [DEBUG]: ===== in the finally statement [edu.ucsb.nceas.metacat.MetacatHandler:handleReindexAction:2808]
metacat 20181012-10:53:02: [DEBUG]: ===== in the finally statement which out is not null [edu.ucsb.nceas.metacat.MetacatHandler:handleReindexAction:2810]
org.dspace.OREParserException: java.lang.NullPointerException
    at org.dataone.cn.indexer.resourcemap.ForesiteResourceMap.<init>(ForesiteResourceMap.java:104)
    at org.dataone.cn.indexer.resourcemap.ResourceMapFactory.buildResourceMap(ResourceMapFactory.java:45)
    at edu.ucsb.nceas.metacat.index.resourcemap.ResourceMapSubprocessor.processResourceMap(ResourceMapSubprocessor.java:97)
    at edu.ucsb.nceas.metacat.index.resourcemap.ResourceMapSubprocessor.processDocument(ResourceMapSubprocessor.java:86)
    at edu.ucsb.nceas.metacat.index.SolrIndex.process(SolrIndex.java:236)
    at edu.ucsb.nceas.metacat.index.SolrIndex.insert(SolrIndex.java:426)
    at edu.ucsb.nceas.metacat.index.SolrIndex.update(SolrIndex.java:637)
    at edu.ucsb.nceas.metacat.index.SolrIndex.update(SolrIndex.java:592)
    at edu.ucsb.nceas.metacat.index.SystemMetadataEventListerner.entryUpdated(SystemMetadataEventListerner.java:146)
    at edu.ucsb.nceas.metacat.index.SystemMetadataEventListerner.entryAdded(SystemMetadataEventListerner.java:119)
    at com.hazelcast.client.impl.EntryListenerManager.notifyListeners(EntryListenerManager.java:148)
    at com.hazelcast.client.impl.EntryListenerManager.notifyListeners(EntryListenerManager.java:130)
    at com.hazelcast.client.impl.ListenerManager.customRun(ListenerManager.java:88)
    at com.hazelcast.client.ClientRunnable.run(ClientRunnable.java:30)
    at java.lang.Thread.run(Thread.java:748)
```

The resource map that is failing to index is

<https://knb.ecoinformatics.org/knb/d1/mn/v2/object/urn%3Auuid%3A5a860679-159e-45a6-bcb6-beb67f4166c2>

This resource map parses fine using com.hp.hpl.jena.rdf.model.Model.read() outside of the index processor code.

This ForesiteResourceMap constructor is taking an org.w3c.dom.Document as input, converting it to an in-memory ByteArrayOutputStream, and converting that to an in-memory String, to finally get a ReaderInputStream to be passed to _init(). We may need to address the handling inefficiency as a secondary improvement to the code, if feasible.

We also need to not catch a general Exception on line 117, but rather catch the more specific exception that is thrown, which will help with troubleshooting where the exception is coming from. Debug logging statements would help too.

The exception is a NullPointerException, so we ultimately need to track down where this is coming from.

It doesn't look like this code has changed in the 2.3.6 version of the library, so an upgrade wouldn't make a difference. I did look on cn-orc-1 to see if we are having similar issues outside of a Metacat MN context, and see errors like these:

```
cn-index-processor-daemon.log.9-[ERROR] 2018-10-12 06:42:44,977 (ErrorHandlerFactory$ErrorLogger:logError:84) {E201} rdf:resource not allowed as attribute here.
cn-index-processor-daemon.log.9:[ERROR] 2018-10-12 06:42:44,977 (ForesiteResourceMap:_init:163) Unable to parse ORE document:
cn-index-processor-daemon.log.9-org.apache.jena.riot.RiotException: {E201} rdf:resource not allowed as attribute here.
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.system.ErrorHandlerFactory$ErrorHandler.error(ErrorHandlerFactory.java:128)
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.lang.LangRDFXML$ErrorHandlerBridge.error(LangRDFXML.java:241)
cn-index-processor-daemon.log.9-xErrorHandler.java:38)      at com.hp.hpl.jena.rdf.arp.impl.ARPSaxErrorHandler.error(ARPSaxErrorHandler.java:38)
cn-index-processor-daemon.log.9-      at com.hp.hpl.jena.rdf.arp.impl.XMLHandler.warning(XMLHandler.java:206)
cn-index-processor-daemon.log.9-      at com.hp.hpl.jena.rdf.arp.impl.XMLHandler.warning(XMLHandler.java:183)
cn-index-processor-daemon.log.9-      at com.hp.hpl.jena.rdf.arp.impl.XMLHandler.warning(XMLHandler.java:178)
cn-index-processor-daemon.log.9-pport.java:199)      at com.hp.hpl.jena.rdf.arp.impl.ParserSupport.warning(ParserSupport.java:199)
cn-index-processor-daemon.log.9-Lexer.java:187)      at com.hp.hpl.jena.rdf.arp.impl.AttributeLexer.error(AttributeLexer.java:187)
cn-index-processor-daemon.log.9-ava:80)      at com.hp.hpl.jena.rdf.arp.impl.QNameLexer.lookup(QNameLexer.java:80)
cn-index-processor-daemon.log.9-(AttributeLexer.java:71)      at com.hp.hpl.jena.rdf.arp.impl.AttributeLexer.processSpecials(AttributeLexer.java:71)
cn-index-processor-daemon.log.9-(WantDescription.java:60)      at com.hp.hpl.jena.rdf.arp.states.WantDescription.startElement(WantDescription.java:60)
cn-index-processor-daemon.log.9-tartElement(AbsWantLiteralValueOrDescription.java:84)      at com.hp.hpl.jena.rdf.arp.states.AbsWantLiteralValueOrDescription.startElement(AbsWantLiteralValueOrDescription.java:84)
cn-index-processor-daemon.log.9-n.startElement(WantLiteralValueOrDescription.java:56)      at com.hp.hpl.jena.rdf.arp.impl.XMLHandler.startElement(XMLHandler.java:56)
cn-index-processor-daemon.log.9-dler.java:121)      at com.hp.hpl.jena.rdf.arp.impl.XMLHandler.startElement(XMLHandler.java:121)
cn-index-processor-daemon.log.9-known Source)      at org.apache.xerces.parsers.AbstractSAXParser.startElement(Unknown Source)
cn-index-processor-daemon.log.9-ement(Unknown Source)      at org.apache.xerces.parsers.AbstractXMLDocumentParser.emptyElement(Unknown Source)
cn-index-processor-daemon.log.9-t(Unknown Source)      at org.apache.xerces.impl.XMLNamespaceBinder.handleStartElement(Unknown Source)
cn-index-processor-daemon.log.9-own Source)      at org.apache.xerces.impl.XMLNamespaceBinder.emptyElement(Unknown Source)
cn-index-processor-daemon.log.9-tartElement(Unknown Source)      at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanStartElement(Unknown Source)
cn-index-processor-daemon.log.9-entContentDispatcher.dispatch(Unknown Source)      at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl$FragmentContentDispatcher.dispatch(Unknown Source)
cn-index-processor-daemon.log.9-document(Unknown Source)      at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument(Unknown Source)
cn-index-processor-daemon.log.9-urce)      at org.apache.xerces.parsers.DTDConfiguration.parse(Unknown Source)
cn-index-processor-daemon.log.9-urce)      at org.apache.xerces.parsers.DTDConfiguration.parse(Unknown Source)
cn-index-processor-daemon.log.9-      at org.apache.xerces.parsers.XMLParser.parse(Unknown Source)
cn-index-processor-daemon.log.9-source)      at org.apache.xerces.parsers.AbstractSAXParser.parse(Unknown Source)
cn-index-processor-daemon.log.9-r.java:151)      at com.hp.hpl.jena.rdf.arp.impl.RDFXMLParser.parse(RDFXMLParser.java:151)
cn-index-processor-daemon.log.9-      at com.hp.hpl.jena.rdf.arp.ARPParser.load(ARP.java:119)
cn-index-processor-daemon.log.9-142)      at org.apache.jena.riot.lang.LangRDFXML.parse(LangRDFXML.java:142)
cn-index-processor-daemon.log.9-$1.read(RDFParserRegistry.java:142)      at org.apache.jena.riot.RDFParserRegistry$ReaderRIOTFactoryImp.read(RDFParserRegistry.java:142)
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.RDFDataMgr.process(RDFDataMgr.java:818
)
```

```
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.RDFDataMgr.read(RDFDataMgr.java:258)
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.RDFDataMgr.read(RDFDataMgr.java:244)
cn-index-processor-daemon.log.9-      at org.apache.jena.riot.adapters.RDFReaderRIOT.read(RDFReaderR
IOT.java:69)
cn-index-processor-daemon.log.9-      at com.hp.hpl.jena.rdf.model.impl.ModelCom.read(ModelCom.java:
274)
cn-index-processor-daemon.log.9-      at org.dspace.foresite.jena.JenaOREParser.parseToModel(JenaORE
Parser.java:119)
cn-index-processor-daemon.log.9-      at org.dspace.foresite.jena.JenaOREParser.parse(JenaOREParser.
java:69)
cn-index-processor-daemon.log.9-      at org.dataone.ore.ResourceMapFactory.parseResourceMap(Resourc
eMapFactory.java:444)
cn-index-processor-daemon.log.9:      at org.dataone.cn.indexer.resourcemap.ForesiteResourceMap._ini
t(ForesiteResourceMap.java:160)
cn-index-processor-daemon.log.9:      at org.dataone.cn.indexer.resourcemap.ForesiteResourceMap.<ini
t>(ForesiteResourceMap.java:73)
cn-index-processor-daemon.log.9-      at org.dataone.cn.indexer.resourcemap.ResourceMapFactory.build
ResourceMap(ResourceMapFactory.java:41)
cn-index-processor-daemon.log.9-      at org.dataone.cn.index.processor.IndexTaskProcessor.getNextIn
dexTask(IndexTaskProcessor.java:641)
cn-index-processor-daemon.log.9-      at org.dataone.cn.index.processor.IndexTaskProcessor.processFa
iledIndexTaskQueue(IndexTaskProcessor.java:229)
cn-index-processor-daemon.log.9-      at org.dataone.cn.index.processor.IndexTaskProcessor.processIn
dexTaskQueue(IndexTaskProcessor.java:191)
cn-index-processor-daemon.log.9-      at org.dataone.cn.index.processor.IndexTaskProcessorJob.execut
e(IndexTaskProcessorJob.java:57)
```

This error is also related to `_init`, but isn't an NPE, so `_init()` might be a good place to look first.