

Metacat - Task #6040

Metacat-index does not handle <references>

07/25/2013 05:01 PM - ben leinfelder

Status:	New	Start date:	07/25/2013
Priority:	High	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	index	Estimated time:	0.00 hour
Target version:	2.3.0	Spent time:	0.00 hour
Description			
I indexed a document from EVOS that uses a reference for a creator rather than the details of the person:			
<pre><creator><references>1359152217358</references></creator></pre>			
But in the index it shows up as " " instead of following the reference back the the id where it was declared:			
<pre><associatedParty id="1359152217358">...</pre>			
http://evos.nceas.ucsb.edu/evos/metacat/df35c.9.14/default			

History

#1 - 07/25/2013 05:12 PM - ben leinfelder

Here is a bit of the bean definition used by indexing to pick out the content from EML

```
<bean id="eml.origin" class="org.dataone.cn.indexer.parser.CommonRootSolrField"
  p:multivalue="true"
  p:root-ref="originRoot">
  <constructor-arg name="name" value="origin" />
</bean>

<bean id="originRoot" class="org.dataone.cn.indexer.parser.utility.RootElement"
  p:name="origin"
  p:xPath="//dataset/creator"
  p:template="[individualName]||[organizationName]">
  <property name="leafs"><list><ref bean="organizationNameLeaf"/></list></property>
  <property name="subRoots"><list><ref bean="individualNameRoot" /></list></property>
</bean>
```

#2 - 08/01/2013 02:05 PM - ben leinfelder

- Target version changed from 2.1.0 to 2.1.1

#3 - 08/23/2013 02:42 PM - ben leinfelder

- Target version changed from 2.1.1 to 2.2.0

#4 - 09/05/2013 04:31 PM - ben leinfelder

- Target version changed from 2.2.0 to 2.2.1

#5 - 10/03/2013 11:07 AM - ben leinfelder

- Priority changed from Normal to High

- Target version changed from 2.2.1 to 2.2.0

Apparently this is fixed in cn-index-processor v1.2.0 -- so we will need to pull in this newer dependency in metacat-index and adjust the code accordingly.

#6 - 10/03/2013 11:45 AM - ben leinfelder

- Target version changed from 2.2.0 to 2.2.1

#7 - 10/03/2013 12:01 PM - ben leinfelder

- Parent task set to #6114

This is included in the 1.2.0 d1 index release. It will not include || but instead will use blanks. Not a very great "solution" but better.

#8 - 10/03/2013 12:17 PM - Matt Jones

Spaces aren't really sufficient as a solution, and there are a lot of references fields in EML. We probably need to contribute a fix for this if Skye is not going to fix it for DataONE.

#9 - 10/10/2013 09:36 AM - Jing Tao

Skye said that the sax parser is used to parse those information. This change may require to use DOM parser. It is a big change.

#10 - 10/10/2013 09:39 AM - ben leinfelder

Even with a SAX parser, the implementation could keep track of all elements with "id" attributes and anytime a "references" element is encountered, substitute with that node. The tricky part would be when we encounter a references element before the actual element that declares the id -- would have to track the references that are unfulfilled and fill them in when we actually get to the id elements.

#11 - 10/21/2013 02:23 PM - ben leinfelder

- Parent task deleted (#6114)