

## Metacat - Bug #7176

### Metacat-index RDF/XML subprocessor not populating prov\_hasDerivations field

03/23/2017 03:09 PM - Peter Slaughter

<b>Status:</b>	Closed	<b>Start date:</b>	03/23/2017
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	index	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	2.8.2	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>			
<b>Description</b>			
<p>The package <a href="https://dev.nceas.ucsb.edu/#view/urn:uuid:c7cda366-5658-4350-ba5a-8d2b84829f5d">https://dev.nceas.ucsb.edu/#view/urn:uuid:c7cda366-5658-4350-ba5a-8d2b84829f5d</a> has one prov relationship 'urn:uuid:94cb9677-be83-4873-aa7c-6691e32229a3' <a href="http://www.w3.org/ns/prov#wasDerivedFrom">http://www.w3.org/ns/prov#wasDerivedFrom</a> urn:uuid:146239cd-2f41-4312-8f90-75c8cad09a48'</p> <p>From this prov relationship, the Solr index 'prov_hasDerivations' field for urn:uuid:146239cd-2f41-4312-8f90-75c8cad09a48 should be set to urn:uuid:94cb9677-be83-4873-aa7c-6691e32229a3. See <a href="https://dev.nceas.ucsb.edu/knb/d1/mn/v1/query/solr/?q=id:%22urn:uuid:146239cd-2f41-4312-8f90-75c8cad09a48%22">https://dev.nceas.ucsb.edu/knb/d1/mn/v1/query/solr/?q=id:%22urn:uuid:146239cd-2f41-4312-8f90-75c8cad09a48%22</a></p> <p>However, the prov_wasDerived from field (the reciprocal relationship) is set for the derivation: <a href="https://dev.nceas.ucsb.edu/knb/d1/mn/v1/query/solr/?q=id:%22urn:uuid:94cb9677-be83-4873-aa7c-6691e32229a3%22">https://dev.nceas.ucsb.edu/knb/d1/mn/v1/query/solr/?q=id:%22urn:uuid:94cb9677-be83-4873-aa7c-6691e32229a3%22</a></p> <p>The problem may be related to the 'prov_hasDerivations' SPARQL query in metacat-index/src/main/resources/application-context-prov-base.xml: &lt;bean id="prov20150115.hasDerivations" class="org.dataone.cn.indexer.annotation.SparqlField"&gt; ... SELECT (str(?pidValue) as ?pid) (str(?derivedDataPidValue) as ?prov_hasDerivations) FROM &lt;\$GRAPH_NAME&gt; WHERE { ?derived_data prov:wasDerivedFrom ?source_data . ?source_data cito:documentedBy ?source_metadata . ?source_metadata dcterms:identifier ?pidValue . ?derived_data dcterms:identifier ?derivedDataPidValue . } Not sure why the 'source_metadata' is included in this query. Also, this query is not the reciprocal of the 'prov_wasDerivedFrom' query:</p> <pre>&lt;bean id="prov20150115.wasDerivedFrom" class="org.dataone.cn.indexer.annotation.SparqlField"&gt; ; ... SELECT (str(?pidValue) as ?pid) (str(?wasDerivedFromValue) as ?prov_wasDerivedFrom ) FROM &lt;\$GRAPH_NAME&gt; WHERE { ?derived_data prov:wasDerivedFrom ?primary_data . ?derived_data dcterms:identifier ?pidValue . ?primary_data dcterms:identifier ?wasDerivedFromValue . }</pre> <p>Note that this will also be a problem for the CN DataONE d1_cn_index_processor component.</p> <p>The resource map for this package has been included.</p>			

#### History

##### #1 - 04/13/2017 03:22 PM - Jing Tao

- Target version set to 2.9.0

##### #2 - 04/20/2017 12:56 PM - Jing Tao

- Target version changed from 2.9.0 to 2.8.2

**#3 - 04/26/2017 03:45 PM - Jing Tao**

*- Status changed from New to Closed*

The query was modified and it worked.

## **Files**

---

resmap-no-hasDerivations.xml	8.37 KB	03/23/2017	Peter Slaughter
------------------------------	---------	------------	-----------------