

## Kepler - Bug #4462

### Component search in remote repositories not working

10/15/2009 12:36 PM - Aaron Aaron

|   |             |                        |                   |
|---|-------------|------------------------|-------------------|
| <b>Status:</b>  | Resolved    | <b>Start date:</b>     | 10/15/2009        |
| <b>Priority:</b>  | Immediate   | <b>Due date:</b>       |                   |
| <b>Assignee:</b>  | Sean Riddle | <b>% Done:</b>         | 0%                |
| <b>Category:</b>  | core        | <b>Estimated time:</b> | 0.00 hour         |
| <b>Target version:</b>  | 2.0.0       | <b>Spent time:</b>     | 0.00 hour         |
| <b>Bugzilla-Id:</b>   | 4462        |                        |                   |
| <b>Description</b>  |             |                        |                   |
| Selecting the checkboxes next to remote repositories from the source button of the component library tab does not include any new components in the search results. |             |                        |                   |
| <b>Related issues:</b>  |             |                        |                   |
| Blocked by Kepler - Bug #2493: actor repository tracking bug  |             | <b>Resolved</b>        | <b>07/18/2006</b> |
| Blocks Kepler - Bug #4781: Use the KarXmlGenerator to create the xml metadata...  |             | <b>Resolved</b>        | <b>02/08/2010</b> |
| Blocks Kepler - Bug #4798: Create parser for KAR XML specification  |             | <b>Resolved</b>        | <b>02/16/2010</b> |
| Blocked by Kepler - Bug #4517: Verify that the search capabilities work   |             | <b>Resolved</b>        | <b>10/29/2009</b> |
| Blocks Kepler - Bug #4857: Search on ecogrid server does not return proper re...  |             | <b>Resolved</b>        | <b>03/01/2010</b> |

### History

#### #1 - 10/15/2009 12:41 PM - Aaron Aaron

Adding to this bug the feature enhancement of being able to search in multiple remote repositories for components...

#### #2 - 10/16/2009 04:07 PM - Aaron Aaron

The problem here was that org.kepler.objectmanager.repository.EcogridRepositoryLibrarySearcher was set as the \_repositorySearcher configuration parameter but the class had been moved to org.kepler.gui. So it was not getting instantiated at startup. I moved the class to the right package and so no it gets instantiated.

But there are errors coming back from the Ecogrid which are very similar to errors we saw at some other point during our wrp testing session...

```
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "61.05.21" in statement [SET TABLE T1939104900 SOURCE "urn.Isid.localdata.db73f738.0.0;fs="]
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "99.05.01" in statement [SET TABLE T1939127964 SOURCE "urn.Isid.localdata.ca5b9640.0.0;fs="]
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "87.11.01" in statement [SET TABLE T1939106822 SOURCE "urn.Isid.localdata.bedf2026.0.0;fs="]
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "59.09.01" in statement [SET TABLE T1939103939 SOURCE "urn.Isid.localdata.7258e813.0.0;fs="]
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "70.06.01" in statement [SET TABLE T1939105861 SOURCE "urn.Isid.localdata.fd9d1a92.0.0;fs="]
ERROR (org.ecoinformatics.seek.dataquery.DBTablesGenerator:generateDBTextTable:481) The error in generateDBTable is bad TEXT table source file - line number: 1 java.lang.NumberFormatException: For input string: "42.11.19" in statement [SET TABLE T1939102978 SOURCE "urn.Isid.localdata.c920ce0f.0.0;fs="]
ER
```

#### #3 - 01/29/2010 11:24 AM - Sean Riddle

Trying to search the kepler dev repository yielded the following errors:

```
[run] trying to clone tpc02-water-flow-base but you can't clone this object for some reason: Problem cloning '
output'
[run] ERROR (org.ecoinformatics.seek.datasources.EcogridDataCacheItem:getDataItemFromEcoGrid:326) Ecogrid
DataCacheItem - error connecting to Ecogrid
[run] AxisFault
[run] faultCode: {http://schemas.xmlsoap.org/soap/envelope/}Server.userException
[run] faultSubcode:
```

```

[run]   faultString: java.rmi.RemoteException: Error reading document: judithk.628
[run]   faultActor:
[run]   faultNode:
[run]   faultDetail:
[run]     {http://xml.apache.org/axis/}hostname:127.0.0.1
[run]
[run] java.rmi.RemoteException: Error reading document: judithk.628
[run]   at org.apache.axis.message.SOAPFaultBuilder.createFault (SOAPFaultBuilder.java:222)
[run]   at org.apache.axis.message.SOAPFaultBuilder.endElement (SOAPFaultBuilder.java:129)
[run]   at org.apache.axis.encoding.DeserializationContext.endElement (DeserializationContext.java:1087
)
[run]   at org.apache.xerces.parsers.AbstractSAXParser.endElement (Unknown Source)
[run]   at org.apache.xerces.impl.XMLNSDocumentScannerImpl.scanEndElement (Unknown Source)
[run]   at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl$FragmentContentDispatcher.dispatch (Un
known Source)
[run]   at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument (Unknown Source)
[run]   at org.apache.xerces.parsers.XML11Configuration.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.XML11Configuration.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.XMLParser.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.AbstractSAXParser.parse (Unknown Source)
[run]   at javax.xml.parsers.SAXParser.parse (SAXParser.java:375)
[run]   at org.apache.axis.encoding.DeserializationContext.parse (DeserializationContext.java:227)
[run]   at org.apache.axis.SOAPPart.getAsSOAPEnvelope (SOAPPart.java:696)
[run]   at org.apache.axis.Message.getSOAPEnvelope (Message.java:435)
[run]   at org.apache.axis.handlers.soap.MustUnderstandChecker.invoke (MustUnderstandChecker.java:62)
[run]   at org.apache.axis.client.AxisClient.invoke (AxisClient.java:206)
[run]   at org.apache.axis.client.Call.invokeEngine (Call.java:2784)
[run]   at org.apache.axis.client.Call.invoke (Call.java:2767)
[run]   at org.apache.axis.client.Call.invoke (Call.java:2443)
[run]   at org.apache.axis.client.Call.invoke (Call.java:2366)
[run]   at org.apache.axis.client.Call.invoke (Call.java:1812)
[run]   at org.ecoinformatics.ecogrid.queryservice.stub.QueryServiceStub.get (Unknown Source)
[run]   at org.ecoinformatics.ecogrid.queryservice.QueryServiceGetToStreamClient.get (Unknown Source)
[run]   at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getDataItemFromEcoGrid (EcogridDataC
acheItem.java:316)
[run]   at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getContentFromSource (EcogridDataCac
heItem.java:159)
[run]   at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.downloadDataFromSource (EcogridDataC
acheItem.java:94)
[run]   at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.doWork (EcogridDataCacheItem.java:85
)
[run]   at org.kepler.objectmanager.cache.DataCacheObject.run (DataCacheObject.java:422)
[run]   at java.lang.Thread.run (Thread.java:613)
[run] ERROR (org.ecoinformatics.seek.datasource.EcogridDataCacheItem:getDataItemFromEcoGrid:326) Ecogrid
DataCacheItem - error connecting to Ecogrid
[run] AxisFault
[run]   faultCode: {http://schemas.xmlsoap.org/soap/envelope/}Server.userException
[run]   faultSubcode:
[run]   faultString: java.rmi.RemoteException: /var/metacat/data/judithk.894.1 (Too many open files)
[run]   faultActor:
[run]   faultNode:
[run]   faultDetail:
[run]     {http://xml.apache.org/axis/}hostname:127.0.0.1
[run]
[run] java.rmi.RemoteException: /var/metacat/data/judithk.894.1 (Too many open files)
[run]   at org.apache.axis.message.SOAPFaultBuilder.createFault (SOAPFaultBuilder.java:222)
[run]   at org.apache.axis.message.SOAPFaultBuilder.endElement (SOAPFaultBuilder.java:129)
[run]   at org.apache.axis.encoding.DeserializationContext.endElement (DeserializationContext.java:1087
)
[run]   at org.apache.xerces.parsers.AbstractSAXParser.endElement (Unknown Source)
[run]   at org.apache.xerces.impl.XMLNSDocumentScannerImpl.scanEndElement (Unknown Source)
[run]   at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl$FragmentContentDispatcher.dispatch (Un
known Source)
[run]   at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument (Unknown Source)
[run]   at org.apache.xerces.parsers.XML11Configuration.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.XML11Configuration.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.XMLParser.parse (Unknown Source)
[run]   at org.apache.xerces.parsers.AbstractSAXParser.parse (Unknown Source)
[run]   at javax.xml.parsers.SAXParser.parse (SAXParser.java:375)
[run]   at org.apache.axis.encoding.DeserializationContext.parse (DeserializationContext.java:227)
[run]   at org.apache.axis.SOAPPart.getAsSOAPEnvelope (SOAPPart.java:696)
[run]   at org.apache.axis.Message.getSOAPEnvelope (Message.java:435)
[run]   at org.apache.axis.handlers.soap.MustUnderstandChecker.invoke (MustUnderstandChecker.java:62)
[run]   at org.apache.axis.client.AxisClient.invoke (AxisClient.java:206)
[run]   at org.apache.axis.client.Call.invokeEngine (Call.java:2784)

```

```

[run] at org.apache.axis.client.Call.invoke(Call.java:2767)
[run] at org.apache.axis.client.Call.invoke(Call.java:2443)
[run] at org.apache.axis.client.Call.invoke(Call.java:2366)
[run] at org.apache.axis.client.Call.invoke(Call.java:1812)
[run] at org.ecoinformatics.ecogrid.queryservice.stub.QueryServiceStub.get(Unknown Source)
[run] at org.ecoinformatics.ecogrid.queryservice.QueryServiceGetToStreamClient.get(Unknown Source)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getDataItemFromEcoGrid(EcogridDataC
acheItem.java:316)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getContentFromSource(EcogridDataCac
heItem.java:159)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.downloadDataFromSource(EcogridDataC
acheItem.java:94)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.doWork(EcogridDataCacheItem.java:85
)
[run] at org.kepler.objectmanager.cache.DataCacheObject.run(DataCacheObject.java:422)
[run] at java.lang.Thread.run(Thread.java:613)
[run] ERROR (org.ecoinformatics.seek.datasource.EcogridDataCacheItem:getDataItemFromEcoGrid:326) Ecogrid
DataCacheItem - error connecting to Ecogrid
[run] AxisFault
[run] faultCode: {http://schemas.xmlsoap.org/soap/envelope/}Server.userException
[run] faultSubcode:
[run] faultString: java.rmi.RemoteException: /var/metacat/data/judithk.630.1 (Too many open files)
[run] faultActor:
[run] faultNode:
[run] faultDetail:
[run] {http://xml.apache.org/axis/}hostname:127.0.0.1
[run]
[run] java.rmi.RemoteException: /var/metacat/data/judithk.630.1 (Too many open files)
[run] at org.apache.axis.message.SOAPFaultBuilder.createFault(SOAPFaultBuilder.java:222)
[run] at org.apache.axis.message.SOAPFaultBuilder.endElement(SOAPFaultBuilder.java:129)
[run] at org.apache.axis.encoding.DeserializationContext.endElement(DeserializationContext.java:1087
)
[run] at org.apache.xerces.parsers.AbstractSAXParser.endElement(Unknown Source)
[run] at org.apache.xerces.impl.XMLNSDocumentScannerImpl.scanEndElement(Unknown Source)
[run] at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl$FragmentContentDispatcher.dispatch(Un
known Source)
[run] at org.apache.xerces.impl.XMLDocumentFragmentScannerImpl.scanDocument(Unknown Source)
[run] at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)
[run] at org.apache.xerces.parsers.XML11Configuration.parse(Unknown Source)
[run] at org.apache.xerces.parsers.XMLParser.parse(Unknown Source)
[run] at org.apache.xerces.parsers.AbstractSAXParser.parse(Unknown Source)
[run] at javax.xml.parsers.SAXParser.parse(SAXParser.java:375)
[run] at org.apache.axis.encoding.DeserializationContext.parse(DeserializationContext.java:227)
[run] at org.apache.axis.SOAPPart.getAsSOAPEnvelope(SOAPPart.java:696)
[run] at org.apache.axis.Message.getSOAPEnvelope(Message.java:435)
[run] at org.apache.axis.handlers.soap.MustUnderstandChecker.invoke(MustUnderstandChecker.java:62)
[run] at org.apache.axis.client.AxisClient.invoke(AxisClient.java:206)
[run] at org.apache.axis.client.Call.invokeEngine(Call.java:2784)
[run] at org.apache.axis.client.Call.invoke(Call.java:2767)
[run] at org.apache.axis.client.Call.invoke(Call.java:2443)
[run] at org.apache.axis.client.Call.invoke(Call.java:2366)
[run] at org.apache.axis.client.Call.invoke(Call.java:1812)
[run] at org.ecoinformatics.ecogrid.queryservice.stub.QueryServiceStub.get(Unknown Source)
[run] at org.ecoinformatics.ecogrid.queryservice.QueryServiceGetToStreamClient.get(Unknown Source)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getDataItemFromEcoGrid(EcogridDataC
acheItem.java:316)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.getContentFromSource(EcogridDataCac
heItem.java:159)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.downloadDataFromSource(EcogridDataC
acheItem.java:94)
[run] at org.ecoinformatics.seek.datasource.EcogridDataCacheItem.doWork(EcogridDataCacheItem.java:85
)
[run] at org.kepler.objectmanager.cache.DataCacheObject.run(DataCacheObject.java:422)
[run] at java.lang.Thread.run(Thread.java:613)
[remainder clipped]

```

#### #4 - 01/29/2010 04:35 PM - ben leinfelder

looks like each and every search result is being fully parsed, which then kicks off the data download for EML actors in the compositeActor...which then overwhelms Metacat..

#### #5 - 02/05/2010 09:00 AM - ben leinfelder

kepler-dev has been wiped clean.

This means we shouldn't be at risk of hosing up the KNB server when issuing searches for components in kepler pointing to kepler-dev.

The search result processing needs to avoid running each result through the MOML parser. I know Aaron was able to do this for local repositories, and I'm hopeful some of that work can be applied to processing remote results.

**#6 - 02/05/2010 09:11 AM - Oliver Soong**

I think the local repositories might still be running the MOML parser. At the very least, they seem to be downloading and caching all the data needed by the workflows.

**#7 - 02/06/2010 10:28 AM - ben leinfelder**

if synchronizing the local repositories is still running the moml parser, then that means you'll be prompted for authentication credentials if you are using protected data in your workflows. is that the case?

**#8 - 02/06/2010 10:29 AM - ben leinfelder**

In regard to replicating problematic components on Kepler-dev, I sent this email to Sean:

-----  
If I was correct about what they contained, a "safe" version would be to search for a dataset in Kepler ("Datos Meteorologicos" is a good clean test data set) and you just drag that onto the canvas.  
Save that "workflow" to the Kepler-dev repository you'll have a testbed that hopefully won't cripple KNB when you search Kepler-dev for that workflow component. wilrwind  
To clarify - I don't think there are any "damaged metacat records" - but there are certainly "problematic moml" files that reference lots of data on the KNB and bring it to a grinding halt when Kepler parses them in a whirlwind of new Threads.

**#9 - 02/09/2010 03:34 PM - ben leinfelder**

Need to verify with Aaron, but I believe:  
components are now being uploaded using a new XML metadata format (a wrapped version of the entity MOML that has a bunch of other KAR manifest information in it),  
We need to change the search. Possibly quite a lot.  
Most crucial is the document type ("namespace" in ecogrid) being searched. Right now it is looking for 'entity' doc types whereas we will have to change it to search for 'kar' doctypes. Then I believe we'll have to actually download the actual kar file that contains the component and process that (rather than, say a plain old XML MOML file like it is now).

**#10 - 02/09/2010 03:36 PM - ben leinfelder**

if you go on Keper-dev's "dev skin" and search using this squery, you can see the "kar" metadata is returned.  
(URL: <http://kepler-dev.nceas.ucsb.edu/kepler/style/skins/dev/querymetacat.html>)

```
<pathquery version='1.2'>
<returndoctype>kar</returndoctype>
<returnfield>/entity/@name</returnfield>
<returnfield>entity/property[@name='KeplerDocumentation']/property[@name='author']/configure</returnfield>
<returnfield>entity/property[@name='KeplerDocumentation']/property[@name='description']/configure</returnfield>
<returnfield>entity/property[@name='KeplerDocumentation']/property[@name='createDate']/configure</returnfield>
<returnfield>entity/property[@name='KeplerDocumentation']/property[@name='workflowId']/configure</returnfield>
<returnfield>entity/property[@name='karLSID']/@value</returnfield>
<returnfield>entity/property[@name='entityId']/@value</returnfield>
<querygroup operator='INTERSECT'>
<queryterm casesensitive='false' searchmode='contains'>
<value>%</value>
</queryterm>
</querygroup>
</pathquery>
```

**#11 - 02/09/2010 03:41 PM - ben leinfelder**

then there's a sample of this "kar metadata" file that aaron uploaded here:  
<http://kepler-dev.nceas.ucsb.edu/kepler/metacat/7885.93.1/xml>

we'll want to search for /kar/mainAttributes/lsid to actually get the KAR file's lsid so we can download it as a component [say, when we drag the search result to the canvas].

**#12 - 03/19/2010 11:27 AM - Jing Tao**

sean:

After searching the remote repository, the search result shows the docid, such as 8128:69:1 rather than the kar name.

**#13 - 03/19/2010 11:30 AM - Sean Riddle**

Yes, I'm aware of that odd naming convention. The problem is, the KARs themselves don't have any other name on the repository, as far as I know. The only name or ID given to the KAR file itself, as opposed to something contained within it, is the LSID. It's far from perfect, and I encourage suggestions of what to use instead. Aaron, you might know a good way to generate a useful name.

**#14 - 03/19/2010 11:59 AM - Aaron Aaron**

ummmm, the filename gets lost in metacat somewhere? You should just display the filename of the KAR file in the search results? If the file name gets lost for some reason going through Metacat then we could save the name as a Main attribute next to the Isid in the MANIFEST.MF

**#15 - 03/19/2010 12:03 PM - Aaron Aaron**

Also Sean, I don't think it makes sense to have the ontology hierarchy inside of the KAR search results. Can you return kar results in addition to ontology results, in a similar manner as to how the local components are displayed?

**#16 - 03/19/2010 01:17 PM - Sean Riddle**

I sent out an email about this a week ago. Ben and Chad thought this (semantic types under a KAR file) would be clearest. I can take another look at it, though.

**#17 - 03/19/2010 01:21 PM - Aaron Aaron**

Sorry I wasn't following that thread... My input would be that if the search results are going to be displayed in the Component Library, then they should be organized in a similar fashion. If a different organization is desired, then the results (and perhaps the search function itself) should be completely separated from the local component search and results.

**#18 - 03/19/2010 01:28 PM - Sean Riddle**

Personally, Aaron, I agree with you. I'll look into reorganizing the tree. Ben, Chad or others, let me know if you are really married to the old representation we discussed.

**#19 - 04/16/2010 11:34 AM - Chad Berkley**

Sean says this is completed. Closing. Reopen if problems are found.

**#20 - 03/27/2013 02:26 PM - Redmine Admin**

Original Bugzilla ID was 4462