

## Kepler - Bug #1836

### data search and access problems via ecogrid

12/16/2004 07:45 AM - Dan Higgins

<b>Status:</b>	Resolved	<b>Start date:</b>	12/16/2004
<b>Priority:</b>	Immediate	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	data access	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	1.0.0rc1	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	1836		

#### Description

In the recent BEAM ENM meeting, ~20 people were searching kepler/ecogrid for a datafile (on metacat). It worked OK for some, but seemed to take a long time. For others, nothing was returned (time out?); others got different numbers of hits returned! (The difference appeared to be hits from DIGIR appeared for some and didn't for others.)

Overall, performance was extremely varied and generally poor. (Maybe because there were a lot of people hitting DIGIR and Metacat at the same time.)

#### History

##### #1 - 12/16/2004 07:50 AM - Rod Spears

One of the reasons DiGIR is slow is because we ALWAYS hit all 10 or providers and have to wait for all the resultsets to come back before it can return any results.

This "should" be improved once we have metadata in the regisrty and enable the users to be more selective about where they want to search.

If they don't want to be selective and want to get all the data from everywhere then it will take a while. This has particaly to do with setting the user's expectations about the task they are performing. We know what it is doing behind the scenes so we have a different set of expectations.

They have no idea the amount of data that is being searched and how it is collected and combined and returned.

##### #2 - 12/16/2004 07:51 AM - Rod Spears

Also, as we enable more providers the slowness with only get worse.

##### #3 - 12/17/2004 12:54 PM - Laura Downey

Giving some further input on this to help the debug prograss. When some of the users couldn't get any search results returned on their machines, they moved to a neighbor's machine (that had previously been successful in getting some search results) and performed the search there and some users were able to get results.

The same problems ocured when the search term was "test" or "Datos Meteorologicos." With the latter, people tried various pieces of the search term. Some used the entire search term, others used just "datos" or "Datos" and still were unable to get any search results.

Still other users tried re-booting Kepler and at least one user was successful in getting search results after doing that.

A few users tried to limit the search to just metacat and were still unsuccessful in getting any search results returned.

Regardless of whether users are searching across a large data set or several large data sets, their expectation was some type of results, not a time out (or the progress indicator just stopping with no results returned). Obviously the moving progress indicator is what lets users know that things are still working, and it is a nice feature. However, if we anticipate even more

slowness in the future, we will need to design some additional feedback, such as letting the user know the system is searching data set 1, then data set 2 etc.

**#4 - 11/02/2005 11:48 AM - Matt Jones**

Initial fixes introduced using the Factory method to fix some of these issues. Need to verify that they work, and adapt to the decision to move to plain web services in the next EcoGrid release. Then test, and close this bug. Higgins agreed to develop a test case for this demonstrating data access failures.

**#5 - 11/02/2005 12:06 PM - Dan Higgins**

This bug is related to Bug # 2174 in SEEK Bugzilla which describes details of problems with multiple datasources.

A test for this is the workflows/eco/IPCC\_Base\_Layers.xml workflow with tries to load 10 IPCC datasource files. This reliably 'hangs' on the last of the datasources (which remains 'red').

**#6 - 03/01/2006 11:05 AM - Jing Tao**

After the ecogrid switch to pure axis, the issues seems get better. In the workshop at Jan. 2006, about 30 people kept searching and getting data from ecogrid and it worked well.

I used 3 machine to do a stress testing in ecogrid and it worked well.

But the stress testing by the cluster in NCEAS (which has 14 machines), sometimes failed - some search didn't get result set back.

**#7 - 06/14/2006 11:32 AM - Dan Higgins**

Although there are some performance issues, the ecogrid connection now seems much more robust. I am thus removing this general bug as FIXED. Specific other problems should be added as needed. (Dan Higgins)

**#8 - 01/18/2008 10:35 AM - Jing Tao**

In 2008 workshop in New Mexico, the ecogrid search worked fine. Dan and I think we can close this bug.

**#9 - 03/27/2013 02:18 PM - Redmine Admin**

Original Bugzilla ID was 1836