Kepler - Bug #3655

create a Workflow Run Manager with ability to export Publication Ready Archives

11/12/2008 05:28 PM - Derik Barseghian

Status:	New	**Start date:**	11/12/2008
Priority:	Normal	**Due date:**	
Assignee:	Derik Barseghian	**% Done:**	0%
Category:	workflow run manager	**Estimated time:**	0.00 hour
Target version:	3.X.Y	**Spent time:**	0.00 hour
Bugzilla-Id:	3655		

Description

Users would like to be able to create 'publication ready archives' as described here:
http://reap.ecoinformatics.org/Wiki.jsp?page=UsecasePublicationReadyArchive

To this end we are creating a "Provenance Browser" within Kepler from which a user may, for example, select a given workflow execution and export it to a Pub Ready Archive.

Other features will also be added, like the ability to delete workflow runs in provenance, open workflows found in provenance, and search. A workflow-run annotation mechanism is also desirable to help search--a user could tag a workflow run for easy identification and searching against later on.

The Provenance Browser will use a (non-gui) Provenance Client API to interact with provenance stores.

Related issues:

Blocks Kepler - Bug #3656: create a Provenance Client	**Resolved**	**11/12/2008**
Blocks Kepler - Bug #4131: Workflow Run Manager - Open in New Window allows o...	**New**	**06/08/2009**
Blocks Kepler - Bug #4132: Workflow Run Manager - Open in New Window is just ...	**New**	**06/08/2009**
Blocks Kepler - Bug #4133: Workflow Run Manager - column widths should be siz...	**Resolved**	**06/08/2009**
Blocks Kepler - Bug #4134: Workflow Run Manager - double clicking on search c...	**Resolved**	**06/08/2009**
Blocks Kepler - Bug #4135: Workflow Run Manager - default sorting (by date) s...	**Resolved**	**06/08/2009**
Blocks Kepler - Bug #4136: Workflow Run Manager - date and duration search sh...	**Resolved**	**06/08/2009**
Blocks Kepler - Bug #4137: Workflow Run Manager - resizing width of column wh...	**New**	**06/08/2009**
Blocks Kepler - Bug #4138: Workflow Run Manager - search cell edit focus isn'...	**New**	**06/08/2009**
Blocks Kepler - Bug #4140: Workflow Run Manager - delete runs functionality n...	**Resolved**	**06/09/2009**
Blocks Kepler - Bug #4144: Workflow Run Manager - moving columns / new window...	**Resolved**	**06/09/2009**
Blocks Kepler - Bug #4327: Workflow Run Manager - ability to import run archives	**New**	**08/21/2009**

History

#1 - 11/12/2008 07:38 PM - Timothy McPhillips

We have a terminology issue here. The term 'Provenance Browser' has been applied to the graphical user interface contributed to Kepler by the pPOD project (and stored in the provenance_apps module in the Kepler repository). The term is used to describe this system in the publication describing it [Bowers et al 2008 LNCS 5272, 70-78], the Kepler newsletter, and 8 bugs (bug 3546, bug 3555, bug 3557, bug 3558, bug 3560, bug 3566, bug 3568, and bug 3552). In particular, bug 3546 proposes that this provenance browser (optionally) display the trace of each workflow run in Kepler automatically, and bug 3568 requests that the Kepler Provenance Recorder be enhanced to support display of detailed intra-run provenance in this provenance browser for any recorded run (not just COMAD workflows).

I bring this up because the current functionality of the existing Provenance Browser is distinct from that proposed here, in that it enables one to browse the detailed provenance of data products from **within** a run of a workflow, whereas the proposed features have to do with managing the records of multiple runs in a project. It strikes me that what is proposed here could be described as a 'project workspace view', or a 'workflow run browser'.

We need both of these sets of capabilities in Kepler.

Note also that pPOD developed a simple project workspace browser for Kepler as well, and its functionality overlaps with the features requested in this bug. I propose that the REAP and pPOD run browser widgets be merged in Kepler, and that the current provenance browser be retained as a distinct graphical user interface element, and integrated with the Kepler GUI.

#2 - 11/12/2008 10:01 PM - Matt Jones

I agree with Tim -- this would be more aptly called a 'Workflow Run Browser' to differentiate it from the Provenance Browser contributed by pPOD.

I've revised the summary of the RFE to reflect this change, but we could still call it something else as discussion ensues.

#3 - 11/20/2008 06:25 PM - Derik Barseghian

Some details on progress on this bug may be found here:
https://dev.kepler-project.org/developers/interest-groups/provenance-interest-group/interfaces-to-provenance-for-reap

The Workflow Run Manager mockup will be revamped in the near future...

#4 - 02/25/2009 04:42 PM - Derik Barseghian

Eric would like to see webplots of the new data. I've renamed the reap01 (Baskett) datasource (which was still named after my office) and inserted reap02 (McL). The dataturbine workflows need to be updated according, and workflows that create the data for web plots need to be modified/created and run.

#5 - 02/25/2009 05:03 PM - Derik Barseghian

Ignore comment #4, it was meant for a different bug.

#6 - 03/02/2009 06:17 PM - Derik Barseghian

Current todos:

0) The WRM is broken again, it needs to be made to work with the new kepler layout, and within the wrp module.
1) WRM needs to fetch and open associated workflow into Kepler Editor when a row is clicked (or double-clicked, or a context menu is clicked?)
2) Provenance store configuration -- should there be a menu? Where is default hsql location? Some sensible population should occur on [initial] Kepler launch.
3) Workflow tagging interface needs to be implemented, tags saved into provenance and made retrievable with queryable interface.
4) Workflow metadata Author needs to be saved (is this a tag?)
5) Ability to assign unique ids (lsids?) to workflow executions, publication ready archives, and so forth needs to be completed.
6) Publication Ready Archive spec needs to be completed, and corresponding Export and Upload actions implemented.
7) Import of runs found in Pub Ready Archives into a given provenance store needs to be implemented
8) Repository needs to allow storing, searching, downloading Pub Ready Archives.
9) Search interface and gui search areas of WRM need to be implemented to search for runs.
10) 'Delete selected runs' method(s) need to be written
11) Date column needs to be made user friendly.

#7 - 03/09/2009 09:35 AM - ben leinfelder

I've been "playing" around with provenance now, too. Seems like we'll want a more application-wide provenance store configuration that will be used across workflows. Otherwise we might be creating/referencing a bunch of different ones - and then generating reports will be even trickier because we'll need to keep track of which store the token values were saved in.
Would a centralized provenance store be configured as part of the WRM or should it be more closely integrated with the core provenance stuff?
I remember there was talk a while ago about changing how the Provenance Recorder was added to the workflow (that it would actually be "added" to every workflow by default)...What's the current thinking on that now?

#8 - 03/09/2009 11:05 AM - Derik Barseghian

Dan's created a properties file for default prov store location. I've done the same for the WRM. Both by default are local hsql. I believe Dan's planning to make an easy way to turn provenance on and off, e.g. maybe with a button in the toolbar. I believe he's also making a config menu in Kepler, wherein you set your prov store type and location. I may overload that dialog with a spot for the WRM prov store. By default they'll both be local hsql, and it would be untypical to set the WRM and the prov recorder to different locations (I'm still debating if it's a necessary option to include - the ability to record to one, and view another at the same time). You could refer to the prov store values in this menu and presumably be ok.

#9 - 03/09/2009 11:19 AM - ben leinfelder

I was also thinking that we could extend the current DBConnectionFactory to allow for more hsql DBs to be served from the running HSQL that Kepler now uses. Right now it's only using a single db - we could configure it such that another (still file-based) hsql DB instance is available.
It's not necessary at the moment....until people get keen on running multiple Kepler instances that point to the same standalone-file-based-hsqldb (and we end up with the same problem we just had with the cache db).
Changing DBConnection to serve multiple DBs does not address the provenance configuration - it would just allow a provenance recorder to *point* to a db that is served by the hsql "server"

#10 - 04/06/2009 03:59 PM - Derik Barseghian

Some updates to the above todo list:

2) Provenance store configuration -- should there be a menu? Where is default
hsql location? Some sensible population should occur on [initial] Kepler
launch.

-- Provenance module now has a config dialog accessible from a toolbar button. During the REAP call today it was recommended I don't implement a similar dialog for the WRM for now, instead just use the default values.

3) Workflow tagging interface needs to be implemented, tags saved into

provenance and made retrievable with queryable interface.

-- Sean and Shawn are working on this task now, "workflow header panel" we've been calling it.

4) Workflow metadata Author needs to be saved (is this a tag?)

-- Subtask for Sean and Shawn.

5) Ability to assign unique ids (lsids?) to workflow executions, publication
ready archives, and so forth needs to be completed.

-- Aaron, Chad and Matt are discussing this.

6) Publication Ready Archive spec needs to be completed, and corresponding
Export and Upload actions implemented.

-- Aaron's working on this.

9) Search interface and gui search areas of WRM need to be implemented to
search for runs.

-- Partially complete. During the REAP call today, an alternate search interface was recommended that will be easier and faster to implement. I'll
remove the groupable, editable, movable header stuff that I have in place now and try this approach.

11) Date column needs to be made user friendly.

-- Nice formatting in place, but still messing with it a bit.

#11 - 04/10/2009 05:40 PM - Derik Barseghian

After some further thinking and discussing of the search interface, I've decided to continue to pursue the integrated table-header approach. I've made
some more headway -- cell editors for the date and duration cells have been created. If new columns are added, this approach can be followed.
These aren't yet tied into the search mechanism, I'll be doing that soon. There are a few gui bugs I want to take care of as well, including search not
working when columns are reordered.

#12 - 04/22/2009 05:37 PM - Derik Barseghian

Workflow, Date and User columns have been tied to search and work pretty well now, though single-date-operator-less search has yet to be
implemented.

Duration and execid need to be implemented. I plan to punt on execid search, since it might be better to use a column of Aaron's ids instead. I'll also
of course have to wait for the tagging system before Tags search.

So the plan for now is to finish Date and Duration search, and then fix column reordering/header bug, sorting not working bug, and a few other
fundamental problems (not populated by default, gui messed up in a few ways). After that, context menu items - delete, open, export, upload. I can
implement Delete on my own. Open will hopefully be able to use Ben's module communication code. Export and Upload will hopefully utilize Aaron's
lsid code.

#13 - 04/23/2009 03:54 PM - Derik Barseghian

Duration search has been mostly implemented and is now usable. Like date, it still needs single-duration operator-less search.

#14 - 04/29/2009 05:09 PM - Derik Barseghian

You can rearrange columns now and they'll stay in sync with the search header.
Column specific default text now works after rearrange.

I've introduced or discovered a new bug. On the mac it only happens for the duration column, but on linux it seems to happen for all columns. If you
move the e.g. duration column, then edit the search cell with something that fails the verify test, the cell gets stuck in editor, the cursor no longer
blinking.

#15 - 05/11/2009 12:51 PM - Derik Barseghian

The two bugs related to search cells getting stuck in edit mode have been fixed.
A gui bug with row highlighting has been fixed.

#16 - 05/14/2009 12:23 PM - Derik Barseghian

Sorting has been fixed.

#17 - 05/14/2009 02:15 PM - Derik Barseghian

Filter button has been eliminated and table is populated by default, bringing WRM more in line with the mockups. 'Filter' is still actually just search, for
now. Search takes place whenever editing stops on a header cell, this should improved to only occur if the value has changed.

#18 - 05/19/2009 02:24 PM - Derik Barseghian

'Open Workflows in New Windows' context menu now works. I cobbled this together though, and need to clean it up and make sure it's ok.

New rows now appear as runs complete. This should be improved:
- if two windows are open and connected to the same prov store, I think both should auto-show the new run, instead of only the window that ran the workflow.

Also in need of improvement, how I'm checking for changes to provenance configuration (updating the WRM to display a different store's runs). It would be better if I were able to listen to ProvenanceConfigurer for changes, instead of sniffing them out later.

Misc bug I've found:
1) if I apple-trackpad-click (to do a right-click, instead of using a mouse right button) on only one row, and select Open Workflows in New Windows, it won't do the action. It will if I select more than one row.

#19 - 05/20/2009 06:53 PM - Derik Barseghian

priorities:
- use and show kepler-lsid
- work with aaron, dan, et al to determine what goes in pub ready archive.
- ability to publish pub ready archive to repository
- open workflow in current window

more todos:
- run manager needs to be able to run without gui, to be able to do Export and/or Upload to Repo options at least.
- provenance On by default
- WRM pane gets stuck, with not much visible
- for durations of 0, instead show "<1s"
- row highlighting, Red for failure, Green for running
- switch to control-click instead of apple-click for right click
- enter search cell editor only when double clicking in this area, not anywhere in header. this will fix that an unwanted column sort always occurs when entering cell editor.

#20 - 06/05/2009 02:33 PM - Derik Barseghian

Locally I've created a WorkflowRunManager class and associated Event and EventListener classes. With these I have the 'not updating across windows on execution finished' problem fixed. This needs more work though. LSIDs are being used much more. Run metadata momls are being put in kars alongside the run-workflow when you do an Export Archive. Continuing work to make things like export work without WRM gui...

#21 - 06/08/2009 12:39 PM - Derik Barseghian

Change referred to in #20 have been checked in.
This bug's become huge, going to create separate bugs that block this one for remaining tasks for clarity.

#22 - 07/16/2009 10:03 PM - Derik Barseghian

I believe I have the trackpad / single button mouse wrm jtable context menu bug fixed, locally. Will check in the fix next time I have Internet.

#23 - 03/27/2013 02:24 PM - Redmine Admin

Original Bugzilla ID was 3655