

## Arvados - Feature #14322

### [CWL] Accept collection uuid in input

10/10/2018 02:21 PM - Peter Amstutz

<b>Status:</b> Resolved	<b>Start date:</b> 03/13/2019
<b>Priority:</b> Normal	<b>Due date:</b>
<b>Assigned To:</b> Peter Amstutz	<b>% Done:</b> 100%
<b>Category:</b>	<b>Estimated time:</b> 0.00 hour
<b>Target version:</b> 2019-03-27 Sprint	
<b>Description</b>	
Arvados-cwl-runner should allow users to provide uuids in input documents. For example:	
<pre>{   "file1": {     "class": "File",     "location": "keep:zzzz-4zz18-zzzzzzzzzzzzzz/file1.txt"   } }</pre>	
Arvados-cwl-runner should replace the value in "location" with the portable data hash and record the UUID in the "arv:collection" field. This applies to both File and Directory objects. Implementation note: this should probably happen somewhere in <code>upload_workflow_deps()</code> which is responsible for updating File references for uploaded files.	
<pre>{   "file1": {     "class": "File",     "location": "keep:abc+123/file1.txt",     "arv:collection": "zzzz-4zz18-zzzzzzzzzzzzzz/file1.txt"   } }</pre>	
Users may provide both "location" by PDH and "arv:collection" in the input. If both portable data hash and arv:collection are present, the portable data hash will take precedence. Print a warning (or error???) if the collection uuid is not readable or the does not match the PDH that was provided.	
When constructing container requests, if "arv:collection" is known for a File or Directory object, include it in the mount object alongside the portable data hash.	
Implementation note: code paths that test <code>startswith("keep:")</code> will need to be updated to distinguish keep identifiers by UUID or PDH. (Alternately, we could use a different uri scheme for uuids).	
<b>Subtasks:</b>	
Task # 14893: Review 14322-cwl-uuid-input	<b>Resolved</b>
<b>Related issues:</b>	
Related to Arvados - Feature #11442: [CWL] Resolve symbolic names to collections	<b>New</b>
Related to Arvados - Feature #14323: [API] Accept container mounts that speci...	<b>Resolved</b> 11/07/2018

### Associated revisions

Revision 90bb5de4 - 03/18/2019 06:43 PM - Peter Amstutz

Merge branch '14322-cwl-uuid-input' refs #14322

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <[pamstutz@veritasgenetics.com](mailto:pamstutz@veritasgenetics.com)>

### History

#1 - 10/10/2018 02:21 PM - Peter Amstutz

- Status changed from New to In Progress

#2 - 10/10/2018 02:21 PM - Peter Amstutz

- Subject changed from [CWL] Accept collection uuid to [CWL] Accept collection uuid

- Status changed from In Progress to New

**#3 - 10/10/2018 02:24 PM - Peter Amstutz**

- Description updated

**#4 - 10/10/2018 02:25 PM - Peter Amstutz**

- Subject changed from [CWL] Accept collection uuid to [CWL] Accept collection uuid in input

**#5 - 10/10/2018 02:30 PM - Tom Clegg**

- Related to Feature #11442: [CWL] Resolve symbolic names to collections added

**#6 - 10/10/2018 03:03 PM - Tom Clegg**

- Related to Feature #14323: [API] Accept container mounts that specify both uuid and portable\_data\_hash added

**#7 - 10/10/2018 03:40 PM - Peter Amstutz**

- Description updated

**#8 - 10/10/2018 04:21 PM - Peter Amstutz**

- Description updated

**#9 - 10/10/2018 04:22 PM - Tom Morris**

- Target version changed from To Be Groomed to Arvados Future Sprints

- Story points set to 2.0

**#10 - 02/27/2019 04:32 PM - Tom Morris**

- Target version changed from Arvados Future Sprints to 2019-03-13 Sprint

**#11 - 02/27/2019 04:41 PM - Peter Amstutz**

- Assigned To set to Peter Amstutz

**#12 - 03/01/2019 07:33 PM - Tom Morris**

- Release set to 15

**#13 - 03/08/2019 09:30 PM - Peter Amstutz**

- Status changed from New to In Progress

**#14 - 03/13/2019 01:07 PM - Peter Amstutz**

- Target version changed from 2019-03-13 Sprint to 2019-03-27 Sprint

**#15 - 03/13/2019 01:21 PM - Peter Amstutz**

14322-cwl-uuid-input 3e9cb56544f3acecf6aa2bf967263600abf0c584

- Accept 'location: keep:[zzzz-4zz18-zzzzzzzzzzzzzzzzzzz](#)' and convert it to portable data hash
- Record uuid in arv:collectionUUID field
- Input can also provide both 'location: keep:PDH' and 'collectionUUID: zzz', this will check that the collectionUUID has the expected PDH.

<https://ci.curoverse.com/view/Developer/job/developer-run-tests/1116/>

Looking at it this morning I realized I should probably update the documentation as well.

**#16 - 03/13/2019 07:29 PM - Peter Amstutz**

14322-cwl-uuid-input @ [b65b36691117322a34170f28ae1997073f2829f0](#)

<https://ci.curoverse.com/view/Developer/job/developer-run-tests/1121/>

- Accept 'location: keep:[zzzz-4zz18-zzzzzzzzzzzzzzzzzzz](#)' and convert it to portable data hash
- Record uuid in arv:collectionUUID field
- Input can also provide both 'location: keep:PDH' and 'collectionUUID: zzz', this will check that the collectionUUID has the expected PDH.
- Update docs
- CollectionFsAccess accepts keep:uuid URIs
- Workbench updated to use collectionUUID field

#17 - 03/14/2019 08:49 PM - Eric Biagiotti

- Looks like the python 2 versions of test\_submit\_mismatched\_uuid\_inputs and test\_submit\_unknown\_uuid\_inputs have unicode problems.
- Not sure why the following needs a while loop. Why doesn't arvranner.api.collections().list() return all the collections we need? [https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/sdk/cwl/arvados\\_cwl/runner.py#L176](https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/sdk/cwl/arvados_cwl/runner.py#L176). Maybe add a comment here?
- Were the conformance and integration tests run?
- Anything we want to add to workbench 2, like we did for workbench?

#18 - 03/15/2019 02:56 PM - Peter Amstutz

Eric Biagiotti wrote:

- Looks like the python 2 versions of test\_submit\_mismatched\_uuid\_inputs and test\_submit\_unknown\_uuid\_inputs have unicode problems.

Fixed.

- Not sure why the following needs a while loop. Why doesn't arvranner.api.collections().list() return all the collections we need? [https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/sdk/cwl/arvados\\_cwl/runner.py#L176](https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/sdk/cwl/arvados_cwl/runner.py#L176). Maybe add a comment here?

For a large number of fetch\_uuids, API server may limit response size, so we need to keep fetching from API server has nothing more to give us.

- Were the conformance and integration tests run?

<https://ci.curoverse.com/view/CWL/job/arvados-cwl-conformance-tests/60/>

- Anything we want to add to workbench 2, like we did for workbench?

Workbench2 lacks the same display of workflow input/output annotated with links to collections. I don't think workflow running in workbench2 understands uuids, either. Added [#14322](#)

14322-cwl-uuid-input @ [45974ce224baf26d0a4c445dd1e9322193f1f64f](https://ci.curoverse.com/view/Developer/job/developer-run-tests/1130/)

<https://ci.curoverse.com/view/Developer/job/developer-run-tests/1130/>

#19 - 03/15/2019 07:35 PM - Eric Biagiotti

<https://ci.curoverse.com/view/CWL/job/arvados-cwl-conformance-tests/60/>

Python 3 conformance test 186 timed out. Not sure if this is related. Maybe a flaky test?

Workbench2 lacks the same display of workflow input/output annotated with links to collections. I don't think workflow running in workbench2 understands uuids, either. Added [#14322](#)

I think you meant to link to 14974 here.

Below is the output from running the example we discussed. Sorry for the confusing virtualenv name, but a-c-r and the arvados-python-client should be running the latest.

```
(arvmount-test-env) eric@ubuntu:~/arvbox/arvbox/arvados/doc/user/cwl/bwa-mem$ /home/eric/arvmount-test-env/bin/arvados-cwl-runner
--create-workflow bwa-mem.cwl bwa-mem-template.yml
2019-03-15 15:27:19 cwltool INFO: /home/eric/arvmount-test-env/bin/arvados-cwl-runner 1.3.1.20190315153329,
arvados-python-client 1.3.1.20190313174948, cwltool 1.0.20181217162649
2019-03-15 15:27:19 cwltool INFO: Resolved 'bwa-mem.cwl' to 'file:///home/eric/.arvbox/arvbox/arvados/doc/user/cwl/bwa-mem/bwa-mem.cwl'
Traceback (most recent call last):
  File "/home/eric/arvmount-test-env/bin/arvados-cwl-runner", line 4, in <module>
    __import__('pkg_resources').run_script('arvados-cwl-runner==1.3.1.20190315153329', 'arvados-cwl-runner')
  File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/pkg_resources/__init__.py", line 658, in
run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/pkg_resources/__init__.py", line 1445,
in run_script
    exec(script_code, namespace, namespace)
  File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/arvados_cwl_runner-1.3.1.20190315153329
```

```
-py2.7.egg/
  EGG-INFO/scripts/arvados-cwl-runner", line 10, in <module>

File "build/bdist.linux-x86_64/egg/arvados_cwl/__init__.py", line 327, in main
File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/cwltool-1.0.20181217162649-py2.7.egg/cw
ltool/main.py", line 785, in main
  secret_store=runtimeContext.secret_store)
File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/cwltool-1.0.20181217162649-py2.7.egg/cw
ltool/main.py", line 368, in init_job_order
  visit_class(job_order_object, ("File",), functools.partial(add_sizes, make_fs_access(input_basedir)))
File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/cwltool-1.0.20181217162649-py2.7.egg/cw
ltool/utlils.py", line 214, in visit_class
  visit_class(rec[d], cls, op)
File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/cwltool-1.0.20181217162649-py2.7.egg/cw
ltool/utlils.py", line 212, in visit_class
  op(rec)
File "/home/eric/arvmount-test-env/local/lib/python2.7/site-packages/cwltool-1.0.20181217162649-py2.7.egg/cw
ltool/process.py", line 349, in add_sizes
  obj["size"] = fsaccess.size(obj["location"])
File "build/bdist.linux-x86_64/egg/arvados_cwl/fsaccess.py", line 160, in size
File "build/bdist.linux-x86_64/egg/arvados_cwl/fsaccess.py", line 103, in get_collection
File "build/bdist.linux-x86_64/egg/arvados_cwl/fsaccess.py", line 78, in get
IOError: [Errno 2] Could not access collection '2463fa9efeb75e099685528b3b9071e0+438': Not Found
```

**#20 - 03/18/2019 02:31 PM - Eric Biagiotti**

I was able to run the bwa-mem example with the different input files from the CLI and see the results on workbench. I was also able to run arvados-cwl-runner --create-workflow bwa-mem.cwl bwa-mem-template.yml and observe the correct reference file being populated in the container request, and under the status tab, the cwl.input.json text is correctly populated with working links for location and <http://arvados.org/cwl#collectionUUID>

Last thing, I would update the following comment:

[https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/apps/workbench/app/helpers/application\\_helper.rb#L681](https://dev.arvados.org/projects/arvados/repository/revisions/14322-cwl-uuid-input/entry/apps/workbench/app/helpers/application_helper.rb#L681).

Other than that, LGTM.

**#21 - 03/27/2019 01:11 PM - Peter Amstutz**

- Status changed from *In Progress* to *Resolved*