

Arvados - Bug #16139

arvados-cwl-runner 2.0.0 secondaryFiles issue with CWL v1.0

02/07/2020 04:45 PM - Bryan Cosca

Status:	Resolved	Start date:	02/10/2020
Priority:	Normal	Due date:	
Assigned To:	Peter Amstutz	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	2020-02-12 Sprint		

Description

```
bcosca2@lightning-dev1:~/issues/15996/cwl$ arvados-cwl-runner --submit --no-wait --project-uuid=su921-j7d0g-4i75jw7iivgg3w7 --api=containers --debug gatk-haplotypcaller.cwl testyaml/gatk-haplotypcaller.yml
```

```
/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/__init__.py:17: CWLToolDeprecationWarning:
```

```
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020.
```

```
Please upgrade your Python as the Python 2.7 version of cwltool won't be maintained after that date.
```

```
    """ , category=CWLToolDeprecationWarning)
INFO /usr/bin/arvados-cwl-runner 2.0.0, arvados-python-client 2.0.0, cwltool 1.0.20190831161204
INFO Resolved 'gatk-haplotypcaller.cwl' to 'file:///data-sdc/bcosca2/issues/15996/cwl/gatk-haplotypcaller.cwl'
DEBUG Creating collection reader for 1e488d8a8de69c5e07b447c8e1ae059b+755
DEBUG Creating collection reader for a3af04432df3d71d22f2fe8be549ba96+5974
INFO Using empty collection d41d8cd98f00b204e9800998ecf8427e+0
INFO Using empty collection d41d8cd98f00b204e9800998ecf8427e+0
INFO Using collection cache size 256 MiB
ERROR Unhandled error:
  string indices must be integers, not str
Traceback (most recent call last):
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/main.py", line 833, in main
    logger=_logger)
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/arvados_cwl/executor.py", line 660, in arv_executor
    runnerjob = next(jobiter)
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/future/builtins/newnext.py", line 59, in newnext
    return iterator.next()
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/arvados_cwl/runner.py", line 649, in job
    self._init_job(job_order, runtimeContext)
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/process.py", line 718, in _init_job
    discover_secondaryFiles=getdefault(runtime_context.toplevel, False))
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/builder.py", line 276, in bind_input
    bindings.extend(self.bind_input(f, datum[f["name"]], lead_pos=lead_pos, tail_pos=f["name"], discover_secondaryFiles=discover_secondaryFiles))
  File "/usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/builder.py", line 317, in bind_input
    if "${" in sf["pattern"] or "${" in sf["pattern"]:
TypeError: string indices must be integers, not str
```

My CWL is good though...

```
bcosca2@lightning-dev1:~/issues/15996/cwl$ arvados-cwl-runner --validate gatk-haplotypcaller.cwl /usr/share/python2.7/dist/python-arvados-cwl-runner/lib/python2.7/site-packages/cwltool/__init__.py:17: CWLToolDeprecationWarning:
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020.
```

Please upgrade your Python as the Python 2.7 version of cwltool won't be maintained after that date.

```
""" , category=CWLToolDeprecationWarning)
INFO /usr/bin/arvados-cwl-runner 2.0.0, arvados-python-client 2.0.0, cwltool 1.0.20190831161204
INFO Resolved 'gatk-haplotypecaller.cwl' to 'file:///data-sdc/bcosca2/issues/15996/cwl/gatk-haplotypecaller.cwl'
gatk-haplotypecaller.cwl is valid CWL.
```

Subtasks:

Task # 16143: Review 16139-cwl-secondaryfiles-fix

Resolved

Associated revisions

Revision 6272cb13 - 02/07/2020 09:11 PM - Peter Amstutz

16139: Fix secondaryFile errors when running --submit --no-wait

When submitting, we want to preserve the original CWL version of the document.

However, when it creates a RunnerContainer (a cwltool.Process) it examines the input interface and expects it to be in the 1.1 data model. But if we preserve/reload the original document, it is still in the 1.0 data model. This causes

The code was deliberately overriding the CWL version in metadata to make this work. This results in the problem reported on this ticket.

The fix is to maintain both the updated and preserved documents, and use them appropriately where they are expected.

refs #16139

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

Revision 04332347 - 02/10/2020 03:14 PM - Peter Amstutz

Merge branch '16139-cwl-secondaryfiles-fix'

fixes #16139

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

History

#1 - 02/07/2020 04:57 PM - Bryan Cosca

- File *gatk-haplotypecaller.yml* added
- File *gatk-haplotypecaller.cwl* added

#2 - 02/07/2020 05:51 PM - Bryan Cosca

I changed the cwlVersion to v1.1 and it submitted successfully.

#3 - 02/07/2020 05:54 PM - Bryan Cosca

- Subject changed from *arvados-cwl-runner 2.0.0 secondaryFiles issue* to *arvados-cwl-runner 2.0.0 secondaryFiles issue with CWL v1.0*

#4 - 02/07/2020 09:15 PM - Peter Amstutz

- Target version set to *2020-02-12 Sprint*
- Assigned To set to *Peter Amstutz*
- Status changed from *New* to *In Progress*

#5 - 02/09/2020 11:11 PM - Peter Amstutz

16139-cwl-secondaryfiles-fix @ [02dbc92e0a0df291a97edcfcc419319fdb5f750](#)
[developer-run-tests: #1721](#) [icon?job=developer-run-tests&build=1721](#)

#6 - 02/10/2020 03:00 PM - Lucas Di Pentima

This LGTM, thanks!

#7 - 02/10/2020 04:10 PM - Peter Amstutz

- % Done changed from 0 to 100

- Status changed from In Progress to Resolved

Applied in changeset [arvados|04332347f3cca9e3aad62c7e67fe981e818b918c](#).

#8 - 02/10/2020 09:29 PM - Peter Amstutz

- Release set to 22

Files

gatk-haplotypecaller.cwl	596 Bytes	02/07/2020	Bryan Cosca
gatk-haplotypecaller.yml	169 Bytes	02/07/2020	Bryan Cosca