Arvados - Bug #14723

[CWL] File that should appear multiple times in output collection only appears once

01/11/2019 04:37 PM - Peter Amstutz

Status:	Resolved		**Start date:**	04/11/2019
Priority:	Normal		**Due date:**	
Assigned To:	Eric Biagiotti		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2019-04-24 Sprint			

Description

A user has reported a situation where, in producing CWL output, there are directories that should contain the same file, however the file only appears in one directory. The theory is that the source files get mapped to exactly one destination in the output collection (picked arbitrarily) instead of copied to every location in the destination collection.

Specifically, make_output_collection uses NoFollowPathMapper() to map each file to exactly one destination. It probably ought to work the other way, mapping each destination to an input file.

Subtasks:

Task # 15037: Review 14723-cwl-multiple-file-targets   **Resolved**

Associated revisions

Revision 454ee2b8 - 04/19/2019 02:59 PM - Eric Biagiotti

Merge branch 'master' into 14723-cwl-multiple-file-targets

refs #14723

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

Revision 556da32e - 04/19/2019 08:32 PM - Eric Biagiotti

Merge branch '14723-cwl-multiple-file-targets'

refs #14723

Arvados-DCO-1.1-Signed-off-by: Eric Biagiotti <ebiagiotti@veritasgenetics.com>

History

#1 - 01/11/2019 04:37 PM - Peter Amstutz

- Status changed from New to In Progress

#2 - 01/11/2019 04:37 PM - Peter Amstutz

- Status changed from In Progress to New

#3 - 01/11/2019 07:06 PM - Peter Amstutz

- Description updated

#4 - 01/11/2019 09:21 PM - Peter Amstutz

- Description updated

#5 - 01/11/2019 09:22 PM - Peter Amstutz

- Description updated

#6 - 02/19/2019 03:27 PM - Peter Amstutz

- Target version set to To Be Groomed

#7 - 03/27/2019 04:04 PM - Tom Morris

- Assigned To set to Eric Biagiotti

- Target version changed from To Be Groomed to 2019-04-10 Sprint

#8 - 04/04/2019 06:48 PM - Eric Biagiotti

- Status changed from New to In Progress

#9 - 04/05/2019 09:02 PM - Eric Biagiotti

Am able to reproduce with the following workflow:

```
class: ExpressionTool
cwlVersion: v1.0

inputs:
  inputfile: File
outputs:
  out: Directory[]
requirements:
  InlineJavascriptRequirement: {}
expression: |
  ${
    var dirs = [];
    dirs.push({"class": "Directory",
              "basename": "testdir1",
              "listing": [inputs.inputfile]});
    dirs.push({"class": "Directory",
              "basename": "testdir2",
              "listing": [inputs.inputfile]});
    return {"out": dirs};
  }
```

and input file:

```
inputfile:
  class: File
  path: test.txt
```


#10 - 04/10/2019 01:21 PM - Eric Biagiotti

- Target version changed from 2019-04-10 Sprint to 2019-04-24 Sprint

#11 - 04/10/2019 09:08 PM - Eric Biagiotti

The 14723-cwl-multiple-file-targets branch at 810fc3e7aab72aa5ae951141955b56f6d4e47cfd now maps files from their target destination instead. The exception is for literals, which store file contents instead of a resolved location.

I also added a test for this and found a bug where an ExpressionTool with a file array output of literals with the same name would not use name conflict resolution. I fixed this and also added a test.

Unit Tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1185/
C/I Tests: https://ci.curoverse.com/view/CWL/job/arvados-cwl-conformance-tests/89/

#12 - 04/11/2019 04:04 PM - Eric Biagiotti

Re-run of service/api tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests-services-api/1210/

#13 - 04/11/2019 08:39 PM - Peter Amstutz

Please add a comment about how StagingPathMapper, instead of internally mapping source → target, maps targets → source.

Also how mapper() is overridden to maintain the behavior of mapping source → target

Also how literals are a special case, they are still mapped by source (identifier) → target

```
            if v.resolved == src or (v.type == "CreateFile" and k == src):
                return v
```

Let's make this logic exclusive:

```
if (v.type != "CreateFile" and v.resolved == src) or (v.type == "CreateFile" and k == src):
```

The code blocks at arvcontainer.py:150 and executor.py:444 are very similar if not identical in function, I wonder if they could be refactored to use the same logic.

Unrelated to this immediate bug, but I noticed this in executor.py:464

```
        except IOError as e:
            logger.warning("While preparing output collection: %s", e)
```

Seems like that should be logger.error() and/or re-raise the exception.

#14 - 04/19/2019 07:53 PM - Eric Biagiotti

Peter Amstutz wrote:

> Please add a comment about how StagingPathMapper, instead of internally mapping source → target, maps targets → source.
>
> Also how mapper() is overridden to maintain the behavior of mapping source → target
>
> Also how literals are a special case, they are still mapped by source (identifier) → target

Done

> [...]
>
> Let's make this logic exclusive:
>
> [...]

Fixed

> The code blocks at arvcontainer.py:150 and executor.py:444 are very similar if not identical in function, I wonder if they could be refactored to use the same logic.

I don't see a huge benefit to refactoring this at the moment, at least as a part of this bug. Maybe we can add a ticket for this?

> Unrelated to this immediate bug, but I noticed this in executor.py:464
>
> [...]
>
> Seems like that should be logger.error() and/or re-raise the exception.

Fixed

Latest at [454ee2b8f0385c542b6f1165a3baf2820425e1a3](#).
Unit Tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1195/
C/I Tests: Run and passed on my machine.

#15 - 04/19/2019 08:24 PM - Peter Amstutz

Eric Biagiotti wrote:

> Latest at [454ee2b8f0385c542b6f1165a3baf2820425e1a3](#).
> Unit Tests: https://ci.curoverse.com/view/Developer/job/developer-run-tests/1195/
> C/I Tests: Run and passed on my machine.

LGTM

#16 - 04/19/2019 08:35 PM - Eric Biagiotti

- Status changed from In Progress to Resolved

#17 - 05/21/2019 10:27 PM - Tom Morris

- Release set to 15