

## Arvados - Bug #15358

### [cwl] CWL conformance test formattest2 fails with C locale

06/12/2019 06:11 PM - Tom Morris

<b>Status:</b>	Resolved	<b>Start date:</b>	07/03/2019
<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>	2019-07-17 Sprint		
<b>Description</b>			
When LANG=C instead of a UTF-8 locale like en_US.UTF-8, the CWL conformance test, v1.0/formattest2.cwl, fails with an encoding error trying to read EDAM.owl which contains UTF-8 characters, but doesn't have an XML encoding declaration in its prolog.			
<pre>Test 65 failed: /home/ci/arvados-cwl-runner-with-checksum.sh --outdir=/tmp/tmp7oGCB5 --quiet v1.0/formattest2.cwl v1.0/formattest2-job.json Test format checking against ontology using subclassOf. Returned non-zero URI prefix 'edam' of 'edam:format_1929' not recognized, are you missing a \$namespaces section? Could not load extension schema keep:29dc87213e125b67355699e8953d3820+62/EDAM.owl: 'ascii' codec can't decode byte 0xc3 in position 3352: ordinal not in range(128) ERROR Workflow execution failed: Expected value of 'input' to have format http://edamontology.org/format_2330 but File has an incompatible format: {   "format": "http://edamontology.org/format_1929",   "basename": "ref.fasta",   "nameroot": "ref",   "nameext": ".fasta",   "location": "keep:23b1d68b203d6c75f314fe9804f50c0e+59/ref.fasta",   "class": "File",   "size": 12010 } ERROR Workflow error, try again with --debug for more information: Workflow did not return a result.</pre>			
the offending XML snippet is:			
<pre>&lt;dc:creator&gt;Matúš Kalaš&lt;/dc:creator&gt;</pre>			
<b>Subtasks:</b>			
Task # 15393: Review 15358-fetch-text-encoding			<b>Resolved</b>
<b>Related issues:</b>			
Related to Arvados - Bug #15655: [CWL] encoding error when printing error log...			<b>Resolved</b> 10/09/2019

#### Associated revisions

Revision 88af3b04 - 07/08/2019 03:24 PM - Peter Amstutz

Merge branch '15358-fetch-text-encoding' closes #15358

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

#### History

#1 - 06/12/2019 06:11 PM - Tom Morris

- Subject changed from [cwl] CWL conformance test fails with C locale to [cwl] CWL conformance test formattest2 fails with C locale

#3 - 06/12/2019 06:13 PM - Peter Amstutz

This might just be an upstream conformance test fix to the XML file to declare the correct encoding (assuming the Python XML loader handles it).

#4 - 06/12/2019 06:16 PM - Tom Morris

- Target version set to 2019-07-03 Sprint

**#5 - 06/13/2019 09:48 PM - Tom Morris**

Peter Amstutz wrote:

This might just be an upstream conformance test fix to the XML file to declare the correct encoding (assuming the Python XML loader handles it).

That may mask the bug, but since the default encoding for XML files is supposed to be UTF-8, it should work without an explicit declaration.

**#6 - 06/19/2019 03:34 PM - Peter Amstutz**

- Assigned To set to Peter Amstutz

**#7 - 06/21/2019 01:15 PM - Peter Amstutz**

This might be a default encoding problem when reading from keep, not sure.

**#8 - 06/21/2019 05:36 PM - Peter Amstutz**

- Status changed from New to In Progress

**#9 - 07/03/2019 01:33 PM - Peter Amstutz**

- Target version changed from 2019-07-03 Sprint to 2019-07-17 Sprint

**#10 - 07/03/2019 05:37 PM - Peter Amstutz**

to reproduce:

```
$ export LANG=C
$ arvados-cwl-runner formattest2.cwl formattest2-job.json
INFO /home/peter/work/scripts/venv/bin/arvados-cwl-runner 1.4.0.20190627185953, arvados-python-client 1.4.0.20190627173408, cwltool 1.0.20190607183319
INFO Resolved 'formattest2.cwl' to 'file:///home/peter/work/common-workflow-language/v1.0/v1.0/formattest2.cwl'
INFO Upload local files: "ref.fasta"
INFO Using collection 23b1d68b203d6c75f314fe9804f50c0e+59 (4xphq-4zz18-gdz4bibpfb0e5ko)
INFO Upload local files: "EDAM.owl"
INFO Using collection 29dc87213e125b67355699e8953d3820+62 (4xphq-4zz18-qweb7yfd9dbmqbir)
Could not load extension schema keep:29dc87213e125b67355699e8953d3820+62/EDAM.owl: 'ascii' codec can't decode byte 0xc3 in position 3352: ordinal not in range(128)
...
```

**#11 - 07/03/2019 06:07 PM - Peter Amstutz**

15358-fetch-text-encoding @ [927d62b545e90676bf4729b6c1ebee56d51eacbe](https://ci.curoverse.com/view/Developer/job/developer-run-tests/1366/)

Add encoding option to CollectionFsAccess.open() and use in fetch\_text()

<https://ci.curoverse.com/view/Developer/job/developer-run-tests/1366/>

**#12 - 07/03/2019 07:05 PM - Peter Amstutz**

Running CWL tests here:

<https://ci.curoverse.com/view/CWL/job/arvados-cwl-conformance-tests/177/>

**#13 - 07/08/2019 05:35 PM - Peter Amstutz**

- Status changed from In Progress to Resolved

Applied in changeset [arvados|88af3b04a254dbf32224c5fd90b7abe9be693501](https://ci.curoverse.com/view/Developer/job/developer-run-tests/1366/).

**#14 - 09/26/2019 02:31 PM - Peter Amstutz**

- Related to Bug #15655: [CWL] encoding error when printing error log tail added

**#15 - 01/22/2020 02:16 PM - Peter Amstutz**

- Release set to 22