

art - Feature #3774

Improved ability to generate output file names.

04/26/2013 11:16 AM - Herbert Greenlee

Status:	Closed	Start date:	04/26/2013
Priority:	Normal	Due date:	
Assignee:	Christopher Green	% Done:	100%
Category:	I/O	Estimated time:	0.00 hour
Target version:	1.10.00	Spent time:	16.00 hours
Scope:	Internal	SSI Package:	art
Experiment:	MicroBooNE		
Description			
Currently, RootOutput will generate output files with a fixed name, or fixed name plus sequence number.			
I'd like to suggest that RootOutput be given more flexibility to generate file names according to a template, where template parameters might include the following.			
<ol style="list-style-type: none">1. Literal strings (as now).2. Sequence number (as now, but not necessarily only at the end of the file name).3. Name of most recently opened input file (or parts of name, like base name or extension).4. Pattern substitution on most recently opened input file name (or parts), like s/pat/sub/.5. Formatted time stamps.6. Environment variables (a particular use case I have in mind is \$CLUSTER and \$PROCESS).			
Related issues:			
Related to art - Feature #4357: Implement on-close file configurable file ren...			Closed

Associated revisions

Revision e3bef940 - 05/09/2014 04:19 PM - Christopher Green

Implement issue #3774.

History

#1 - 08/09/2013 03:37 PM - Christopher Green

- Status changed from New to Feedback

Something very similar to this feature has been implemented as part of issue [#4357](#). Please review this issue and advise if other specifiers would need to be recognized and interpolated and if so, what they are. Please note that even, "sequence no." is now not well specified in the light of this feature. Do you mean "nth output file in this stream" or, "nth output file that would otherwise have the same name?"

#2 - 03/24/2014 09:34 AM - Herbert Greenlee

Nth file in output stream

#3 - 04/09/2014 11:18 AM - Christopher Green

- Target version set to 1.10.00

#4 - 05/08/2014 10:38 AM - Christopher Green

- Assignee set to Christopher Green

- Scope set to Internal

- Experiment MicroBooNE added

Current status:

1. Literal strings (as now).

Done (of course).

2. Sequence number (as now, but not necessarily only at the end of the file name).

Will be implemented as %#.

3. Name of most recently opened input file (or parts of name, like base name or extension).

Will be implemented as %ifb (basename, no extension), ife (extension), ifn (basename with extension), ifp (full path), ifd (fully-resolved path, no filename). No input file (e.g. EmptyEvent) will result in a substitution of - in all cases.

4. Pattern substitution on most recently opened input file name (or parts), like s/pat/sub/.

%ifs/in/out/[ig]% (note trailing %). Careful escaping will be essential. Literal % characters (escaped or otherwise) will truncate the expression, and are therefore illegal.

5. Formatted time stamps.

Already available as %to and %tc (formatted to ISO 8061).

6. Environment variables (a particular use case I have in mind is \$CLUSTER and \$PROCESS).

\$PROCESS and \$CLUSTER are only applicable to condor and not to (say) PBS for arranged or opportunistic operation at OSG sites. I would much prefer that this remain the purview of the experiment to pre-process the .fcl file in their workflow. We can discuss this further at a stakeholder meeting if you wish. More generally, environment variable substitution in .fcl files gets raised every now and again, and gets argued down every time by the stakeholders. Truly, pre-processing of configuration files by a workflow script is a tried and tested method of doing this kind of thing, and is completely flexible and under the control of the experiment(er) rather than requiring individual feature additions by the art team.

Please let us know what you think about these proposals.

#5 - 05/09/2014 04:28 PM - Christopher Green

- Status changed from *Feedback* to *Resolved*

- % Done changed from 0 to 100

- SSI Package art added

- SSI Package deleted ()

Implemented and tested with [e3bef94](#).

Change to input file name substitution regex specification vs previous proposal: format is:

```
%ifs%<match>%<sub>%<flags>%
```

Things to note:

- <match> and <sub> are expected in ECMAScript regex format (effectively Perl regex).
- Escape carefully.
- % characters (escaped or otherwise) are not allowed except as required delimiters and bounds shown above.
- Refer to capture groups with the Perl notation \${#}.
- Named capture groups are not supported.

Per stakeholder discussion, (6) has not been implemented.

#6 - 05/09/2014 04:43 PM - Christopher Green

- Status changed from *Resolved* to *Assigned*

#7 - 05/09/2014 04:44 PM - Christopher Green

- Status changed from *Assigned* to *Resolved*

#8 - 06/12/2014 01:28 PM - Christopher Green

- Status changed from *Resolved* to *Closed*