art - Feature #3744

Missing sam metadata

04/23/2013 11:23 AM - Herbert Greenlee

Status:	Closed		**Start date:**	
Priority:	Normal		**Due date:**	09/30/2013
Assignee:	Christopher Green		**% Done:**	100%
Category:	Metadata		**Estimated time:**	32.00 hours
Target version:	1.10.00		**Spent time:**	23.00 hours
Scope:	Internal		**SSI Package:**	art
Experiment:	-			

Description

Between them, FileCatalogMetadata and RootOutput module should set all predefined sam metadata fields that they are capable to set. The full list of predefined sam metadata fields can be found on the sam web redmine wiki here:

https://cdcvs.fnal.gov/redmine/projects/sam-web/wiki/Metadata_format

The fields that are currently not being set, and which were not mentioned in the requirements document, but could and should be set, are as follows:

a) group
b) event_count
c) first_event
d) last_event
e) start_time
f) end_time
g) runs
h) parents

Group is a static parameter, that could be set by fcl parameter in FileCatalogMetadata. The rest of these are per-file that need to be set in RootOutput. Note that sam run information for each run consists of a 2-tuple consisting of an integer run number and a string run type. Generally speaking, the run number and run type should be inherited from the input run number and run type (if any).

Related issues:

Related to art - Feature #3743: FileCatalogMetadata service doesn't set file ...	**Closed**	**09/30/2013**
Related to art - Feature #3745: Art should not require metadata that sam cons...	**Closed**	**09/30/2013**
Related to art - Feature #3746: Art should support generating per-file user-s...	**Closed**	**09/30/2013**

Associated revisions

Revision d41c5366 - 06/04/2014 11:25 AM - Christopher Green

Remaining items for issue #3744.

History

#1 - 08/09/2013 03:23 PM - Christopher Green

- Due date set to 09/30/2013

- Status changed from New to Accepted

- Target version set to 1.09.00

- Start date deleted (04/23/2013)

- Estimated time set to 32.00 h

- Scope set to Internal

- Experiment - added

- SSI Package art added

Accepted to within changes made at the SAM requirements meeting in May.

#2 - 08/21/2013 11:51 AM - Christopher Green

According to notes from the meeting on 2013-06-18:

    a) group

This should be user-provided via FHiCL.

    b) event_count
    c) first_event
    d) last_event
    e) start_time
    f) end_time
    g) runs

Only if the user specify run-type. Should coordinate with Robert Illingworth on whether this should be a triple (run, subrun, type) or omit the subrun.

    h) parents

Immediate input files only. This should be on a per output label / per output file basis.

Additionally, file_type should be optional, with "unknown" being filled in by art if not otherwise specified.

#3 - 02/17/2014 12:27 PM - Christopher Green

- Target version changed from 1.09.00 to 521

#4 - 04/09/2014 11:18 AM - Christopher Green

- Target version changed from 521 to 1.10.00

#5 - 05/09/2014 04:43 PM - Christopher Green

- Status changed from Accepted to Assigned

#6 - 05/09/2014 04:45 PM - Christopher Green

- Assignee set to Christopher Green

#7 - 05/27/2014 11:30 AM - Christopher Green

- Status changed from Assigned to Feedback

- % Done changed from 0 to 60

The end time ((f) in your list) **could** be inserted into the file, but this would have to be done before the metadata were written in to the file, and would obviously be off by a second or two. This would also involve changing the semantics of %tc in the output file renaming to match. Is this acceptable?

#8 - 06/04/2014 11:58 AM - Christopher Green

- % Done changed from 60 to 100

Resolved as follows with [35c18e5](#):

- New options --sam-group (services.FileCatalogMetadata.group, metadata item group), --sam-run-type (services.FileCatalogMetadata.runType, used in metadata item runs).
- New automatically-generated metadata item event_count.
- New metadata item runs, automatically generated if metadata item run_type specified, either with command-line, FHiCL or manually. Text here is canonicalized (practically, this means wrapped in double quotes.).
- New metadata item parents, automatically generated if appropriate. Text is canonicalized.
- New metadata item start_time. Example format: 2014-06-04T16:14:11.

Important notes:

- first_event and last_event have **not** been implemented. This is due to the fact that SAM accepts them as a single number only, which would lose the run and subrun information.
- Text items have been canonicalized in tuple contexts, otherwise the tuple could be un-parseable in certain circumstances. Please advise if they should be canonicalized in other contexts. Ditto for keys.
- The runs item lists all run / subrun combinations written as art SubRun records in the file, regardless of whether events from those subruns have been written to the file.
- The end_time item has **not** been written, as the file has not yet been closed. Guidance was not received on changing the definition of "end time" to before the writing of the metadata. This would also change the definition of the time used for the %tc file name substitution.

#9 - 06/04/2014 02:22 PM - Christopher Green

- Status changed from Feedback to Resolved

#10 - 06/05/2014 04:41 PM - Christopher Green

- Status changed from Resolved to Assigned

Per discussions in the stakeholders meeting:

- first_event and last_event will be output as [run, subrun, event]
- end_time will be set to the time at which the metadata are to be written to the SQLite DB. This will not affect the time used in the substitution of %tc in file names, which will continue to be the file closed time.
- The new behavior of sam_metadata_dumper will be to output the data in JSON format: all strings, including keys, will be canonicalized (double-quoted, escaped). This puts the data into a form where tools are readily available to handle double-quoted text items.

#11 - 06/10/2014 09:16 AM - Christopher Green

- Status changed from Assigned to Resolved

1. end_time is now written to the metadata is the time directly before it was filled. The meaning of end_time for output file name substitution is unchanged (immediately after file close).
2. first_event and last_event are written as tuples: [r, sr, e].
3. The default behavior of sam_metadata_dumper is to write a JSON document. Sample output:

```
{
  "test/Integration/TestMetadata_plugin_t.d/out.root": {
    "process_name": "DEVEL",
    "key1": "value1",
    "key2": "value2",
    "file_format": "artroot",
    "file_format_era": "ART_2011a",
    "file_format_version": 5,
    "start_time": "2014-06-10T13:59:22",
    "end_time": "2014-06-10T13:59:22",
    "event_count": 1,
    "first_event": [ 1, 0, 1 ],
    "last_event": [ 1, 0, 1 ]
  },
  "test/Integration/SAM_metadata.d/out.root": {
    "applicationFamily": "Ethel",
    "applicationVersion": "v0.00.01a",
    "file_type": "MC",
    "run_type": "MCChallenge",
    "group": "MyGang",
    "process_name": "SAMMetadataW",
    "testMetadata": "success!",
    "dataTier": "The one with the thickest frosting",
    "streamName": "o1",
    "file_format": "artroot",
    "file_format_era": "ART_2011a",
    "file_format_version": 5,
    "start_time": "2014-06-10T13:59:22",
    "end_time": "2014-06-10T13:59:22",
    "runs": [ [ 1, 0, "MCChallenge" ] ],
    "event_count": 1,
    "first_event": [ 1, 0, 1 ],
    "last_event": [ 1, 0, 1 ]
  }
}
```

This document has passed JSON verification at http://jsonformatter.curiousconcept.com/ and should be readable with any parser. See http://json.org/ for details. The original human-readable output may be obtained by using the -H, --hr or --human-readable options.

#12 - 06/10/2014 09:44 AM - Christopher Green

Resolved with commits cb6a622, fe1b8fe and 94a71a2.

#13 - 06/12/2014 01:28 PM - Christopher Green

- Status changed from Resolved to Closed