A Proposal for LBNE Metadata in SAM

Jonathan Insler, Tom Junk, Qizhong Li

(Oct. 28, 2013)

(updated Nov. 20, 2014)

Metadata are a set of parameters that describes data files. These data files can be real data events collected by the experiment or they can be Monte Carlo generated or simulated events.

The metadata serve two purposes:

- To be used to search data files with required information or parameters;
- To be used by the data catalog system to manage these data files.

It is important to design a set of metadata parameters to be used to describe lbne data files. We use these metadata to store data files onto tapes in FNAL ENSTORE system through SAM. SAM is a data handling system for storing and retrieving files and the files' metadata for an experiment. We use lbne SAM as our data catalog system, to store and retrieving files by their metadata. Usually users do not directly interact with the lbne SAM database, but through a http server, SAMWEB server. Using SAMWEB commands, users can query files by using the metadata, or can define a dataset with desired parameters in the metadata.

The SAM metadata includes two categories:

1) Predefined metadata parameters;
2) LBNE specific metadata parameters.

The following table contains the SAM predefined metadata parameters (can be found at SAM team's wiki web page: https://cdcvs.fnal.gov/redmine/projects/sam-web/wiki/Metadata_format).

file_name	String	Required	The name of the file
file_type	String	Required	The type of the file
file_size	Integer	Required	The size of the file (in bytes)
crc	Integer or String	Optional	The file checksum
file_format	String	Optional	The format of the file (defaults to " unknown")
content_status	String	Optional	The file contents status (defaults to "good")
Group	String	Optional	The SAM work group the file is associated with
data_tier	String	Optional	The data tier
event_count	Integer	Optional	The number of events in the file
first_event	Integer	Optional	The first event number in the file
last_event	Integer	Optional	The last event number in the file
start_time	Date	Optional	The start time of the file
end_time	Date	Optional	The end time of the file
file_partition	Integer	Optional	The file partition number for the file
application		Optional	The application used to create the file
process_id	Integer	Optional	The SAM process ID for the job that created this file. (Only meaningful for files which are produced by a SAM processing job.)
data_stream	String	Optional	The stream of the file
runs	A list of pairs of { run_number, run_type }	Optional	The run numbers associated with this file
lum_block_ranges	A list of pairs of { min, max }	Optional	Pairs of luminosity block numbers for this file
parents	A list of files	Optional	The parents of this file
params		Optional	Generic metadata parameters
category.name*		Optional	Experiment specific

*The "category.name" here should be "lbne_MC.name" or "lbne_data.name".

We propose the following values for the basic parameters:

Metadata parameter	Values
file_type	mc physics-data test-data calib-data fastmc unknown
file_format	root tar text unkown
data_tier	raw hit-reconstructed full-reconstructed generated simulated
group	Lbne
runs	{ number, cosmic } { number, test } { number, physics } { number, calibration } { number, special }

The following table contains the proposed lbne experiment specific metadata parameters for data files:

Metadata Parameter	Type	Values	Description
lbne_data.name	String		Sample name
lbne_data.run_mode	String	continuous main DAQ mode triggered main DAQ mode immediate triggered mode triggered window mode wide--window burst mode burst mode scope mode	Run Mode
lbne_data.detector_type	String	35t Other	Detector type

The following table contains the proposed lbne experiment specific metadata parameters for MC files:

Metadata Parameter	Type	Values	Description
lbne_MC.name	String		Sample name
lbne_MC.beam_energy	String		Beam energy
lbne_MC.detector_type	String	ND 10kt 34kt 35t other	Detector type
lbne_MC.geometry_version	String		Geometry version
lbne_MC.trigger-list-version	String		Trigger list and version #
lbne_MC.generators	String	GENIE CRY GENIE_CRY	Generators
Lbne_MC.overlay	String		overlaid or not
lbne_MC.neutrino_flavors	String	nu_e nu_mu nu_tau anti-nu_e anti-nu_mu anti-nu_tau	ν_e, ν_μ, ν_τ, $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$
lbne_MC.oscillationP	String	oscillated, unoscillated, delta_CP,	Oscillation parameters
lbne_MC.mass_hierarchy	String	Normal mass hierarchy, Inverted mass hierarchy	mass hierarchy
lbne_MC.beam_flux_ID	Integer	1 (ν_e flux, anuflux) 2 ($\bar{\nu}_e$ flux, anuflux) 3 ($\nu_\mu$flux, anuflux) 4 ($\bar{\nu}_\mu$ flux, anuflux)	

		-1 (ν_e flux, nuflux) -2 ($\bar{\nu}_e$ flux, nuflux) -3 ($\nu_\mu$flux, nuflux) -4 ($\bar{\nu}_\mu$ flux, nuflux)	
lbne_MC.miscellaneous	String		log files, …

The metadata format for declaring files must be a valid JSON (http://www.json.org) object.

Here is an example JSON file for metadata of an lbne MC file:

genieNtp_20130423_lbne_lbneCD1_anuflux_numuflux_nue_g280_Ar40_5000_0.root.json


```
{
 "file_name": "genieNtp_20130423_lbne_lbneCD1_anuflux_numuflux_nue_g280_Ar40_5000_0.
root",
 "file_size": 81451838 ,
 "file_type": "mc",
 "file_format": "root",
 "data_tier": "simulated",
 "group": "lbne",
 "application": {
  "family": "art",
  "name": "lar",
  "version": "dev2013.07.19"
 },
 "runs": [
  [ 1, "physics" ]
 ],
```

"event_count": 100,

"first_event": 1,

"last_event": 100,

"start_time": "2013-04-23T10:15:10",

"end_time": "2013-04-23T10:15:10",

"lbne_MC.name": "genieNtp_20130423_lbne_lbneCD1_anuflux_numubarflux_nuebar_g280_Ar40",

"lbne_MC.detector_type": "10kt",

"lbne_MC.generators": "GENIE",

"lbne_MC.neutrino_flavors": "nu_e",

"lbne_MC.beam_flux_ID": "3"

}

And an example JSON file for metadata of lbne raw data file from 35t vertical slice test:

{

"file_name": "v35t_r0000101_s0001_test_raw.root",

"file_size": 540026,

"event_count": 70,

"first_event": 1,

"last_event": 70,

"file_type": "test-data",

"file_format": "root",

```
  "data_tier": "raw",
  "group": "lbne",
  "application": {
   "family": "art",
   "name": "daqag",
   "version": "v00_00_01"
  },
  "runs": [
   [ 101, "test" ]
  ],
  "start_time": "2014-11-20T20:19:59",
  "end_time": "2014-11-20T20:20:04",
  "lbne_data.name": "v35t_testdata_201411",
  "lbne_data.detector_type": "35t",
  "lbne_data.run_mode": "immediate triggered mode"
}
```