P106 (Pass6) Reprocessing for L&EO data

P106 Reprocessing

status: **Complete**
last update: 30 June 2010

This page is a record of the configuration for the P106 reprocessing project. It targets 200 runs of L&EO data for processing with a new ~~alignment~~ calibration and other improvements to investigate issues with the Pass7 classification.

The name "P106" derives from the word "processing" and the initial file version to be used for the output data products, e.g., r0123456789_v106_merit.root.

- P106-LEO-MERIT - this task reads DIGI, runs the full reconstruction code in Gleam and produces reprocessed RECON + MERIT + CAL + GCR + FT1

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v106.

XROOT location and file naming

Location template:

```
/glast/Data/Flight/Reprocess/<reprocessName>/<dataType>
```

Locations for P106-LEO:

```
/glast/Data/Flight/Reprocess/P106-LEO/merit
```

And so forth for the other file types produced.

File naming:

Data Type	Naming template
MERIT, RECON, GCR, CAL	r<run#>_<version>_<dataType>.root
FT1	gll_ph_r<run#>_<version>.fit

Example:

```
/glast/Data/Flight/Reprocess/P106-LEO/merit/r0239557414_v106_merit.root
/glast/Data/Flight/Reprocess/P106-LEO/ft1/gll_ph_r0239559565_v106.fit
```

DataCatalog location and naming

Logical directory and group template:

```
Data/Flight/Reprocess/<reprocessName>:<dataType>
```

Note that the <dataType> field (following the colon) is a DataCatalog 'group' name.

Logical directories for P106-LEO:

```
Data/Flight/Reprocess/P106-LEO:MERIT
Data/Flight/Reprocess/P106-LEO:FT1
```

And so forth...

In the DataCatalog, all file names are of the form r<run#>.

Naming examples:

```
Data/Flight/Reprocess/P106-LEO:MERIT r0239557414
Data/Flight/Reprocess/P106-LEO:FT1 r0239557414
```

Data Sample

The currently defined data sample for P110 and P110-LEO reprocessing includes:

	P106-LEO
First run	237783738 (2008-07-15 03:02:18 UTC)
Last run	244401823 (2008-09-29 17:23:43 UTC)
Total runs	200
Total MERIT events	~488M
Total FT1 events	~23M

Note that the L&EO data represent a **discontiguous** set of runs.

==

Bookkeeping

1. (This page): Define ingredients of reprocessing (processing code/configuration changes)
2. Processing History database: http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P106-LEO
    a. List of all reprocessings
    b. List of all data runs reprocessed
    c. Pointers to all input data files (-> dataCatalog)
    d. Pointers to associated task processes (-> Pipeline II status)
3. Data Catalog database: http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp
    a. Lists of and pointers to all output data files
    b. Meta data associated with each output data product

P106-LEO-MERIT

Status chronology

- 29 June 2010 task continues with little impact on xroot/nfs servers (at least as ganglia is concerned). But wait. The processClump step ran without undue stress on xroot, but the mergeClumps was not so lucky. Both of the newest wains (60,61), which are 12-CPU machines, both got overloaded and lost contact with the world, as shown in the Ganglia plot.

  By 21:00 task complete.
- 28 June 2010 14:15 - begin full task...slowly
- 27 June 2010 - update task with new calibration flavor, add back in 200th run, and fire off test stream (4)
- 13 May 2010 - four test streams submitted...formally successful. Anders has given cursory glance at first few events of first stream (recon/merit only) and the log file, saying all looks okay. However, it has been discovered that the alignment issue thought to be an issue with the L&EO data is a non-issue. So put this task ON HOLD.
- 11 May 2010 - Prepare task

Configuration

The configuration for this reprocessing is nearly the same as for the ordinary science data with three exceptions: the run list was provided by Anders (and consists of a discontiguous set of runs); fewer data products are produced; and, the algorithm for finding FT2 files was modified to accommodate these earlier data (in fact, Warren produced a new set of 1-second FT2 files specifically for this reprocessing project). Note that the original list of runs counted 200, but a single run proved troublesome, 238781852, and was removed from the list, leaving 199 runs to reprocess.

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P106-LEO-MERIT
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp

GlastRelease	v15r47p12gr10
Input Data Selection	??????? "standard" from https://confluence.slac.stanford.edu/display/SCIGRPS/LAT+Dataset+Definitions along with "&& (RunQuality != "Bad" || is_null (RunQuality)"
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P106-LEO-MERIT/config/runFile.txt
ScienceTools	v9r15p5
EvtClassDefs	00-14-00
Evt Classifier	Pass6_Reprocessing_Classifier
photonFilter	PASS6_FSW_CUTS = (FswGamState==0||FswGamState==3) && (CTBCORE>0) && (CTBBestEnergyProb>0) && (CTBBestEnergy>10) && (CTBBestEnergyRatio<5) && (CTBClassLevel>0)
IRFs	P6_V3_DIFFUSE
jobOpts	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P106-LEO-MERIT/config/doRecon.txt
Output Data Products	MERIT, RECON, CAL, GCR, FT1
Special Calib	``` CalCalibSvc.FlavorAsym = "calAsym_test"; \\ CalibDataSvc.CalibFlavorList = {"vanilla","calAsym_test"}; ```

Other configurations.

ROOT version	v5.20.00-gl5
Skimmer version	v7r3p3-gl2

Time

processClump (~23k evts) - about ~45 CPU min (hequ), and ~65 CPU min on fell

mergeClumps (full run) - about 62 CPU min (hequ), mostly due to gtdiffrsp

All 200 runs were fully reprocessed within 30 hours elapsed time. A longer task will not necessarily scale from these numbers, as this elapsed time includes all phases of the task: ramp-up, steady-state, xroot troubles, ramp-down, clean-up. In addition, for the first part of this reprocessing the heavy-lifting jobs were run in the long batch queue which at the time had a per user limit of 1000 jobs (this has subsequently been changed to 'no limit'). The latter part of the task was run in the xlong queue which had no per user limit - but a global limit of 3000 jobs.

Space

(excerpted from the dataCatalog)

Name	Files	Events	Size	Created (UTC)
CAL	200	467,566,271	1.5 TB	13-May-2010 18:29:41
FT1	200	23,467,183	2.0 GB	13-May-2010 18:29:40
GCR	200	488,288,751	10.1 GB	13-May-2010 18:29:42
MERIT	200	488,288,751	368.6 GB	13-May-2010 18:29:42
RECON	200	488,268,806	6.4 TB	13-May-2010 18:29:41

Total xroot disk space (exclusive of /glast/Scratch) occupied by this task = 8.3 TB