P107 Reprocessing Notes

P107 Reprocessing

status: **Complete**
last update: 5 November 2010

This reprocessing will create GCR (Galactic Cosmic Ray) root files which were inadvertantly not produced from a period of Nov 2009 through July 2010.

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v116.

XROOT location and file naming

Location template:

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

Locations for P107:

```
/glast/Data/Flight/Reprocess/P107/gcr
```

File naming:

Data Type	aka	Send to FSSC	Naming template
GCR		No	

Example:

```
/glast/Data/Flight/Reprocess/P107gcr/.root
```

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, and file names are of the form r<run#>.

Naming example:

```
/Data/Flight/Reprocess/P107:GCR r0239557414
```

Data Sample

The currently defined data sample for P116 reprocessing includes:

First run	277039746 (MET), 2009-10-12 11:29:06 (UTC)
Last run	302464405 (MET), 2010-08-02 17:53:25 (UTC)
Total runs	4460
Total DIGI/RECON events	9,916,507,358

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P107-GCR
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp
GlastRelease	v15r47p12gr13
ROOT version	v5.26.00a-gl1
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/P107-GCR/config/runFile.txt
jobOpts	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-MERIT/config/reGCR.txt
Output Data Products	GCR

Processing log

- 11/5/2010 - Production complete
- 11/2/2010 - Validation passed, restart production.
- 11/1/2010 - process additional 79 test/validation runs (bring total # runs to 100). Use trickleStream, limiting number of simultaneous processClump jobs to 500. This put a significant strain on wain038, essentially 100% CPU utilization for ~1 hour, and ~350 MB/s I/O. Also affected were wains 037 and 039, but not to quite the same extent.
- 10/27/2010 - commission new form of task in which GCR files produced in clumps and then merged (by tskim); process first 21 test/validation runs.
- 10/20/2010 - discover two facts: first 20 test runs were limited to 100k events, and Gleam has huge memory leak and crashes after about 400k-600k events have been reprocessed. Report to Heather, et al.
- 10/19/2010 - Configure task and create first 20 test/validation runs. Successful.