

**GAMMA-RAY LARGE AREA
SPACE TELESCOPE
(GLAST)**

**LAT EVENT SUMMARY
DATABASE
REQUIREMENTS
DOCUMENT**

DRAFT VERSION 1.2

MARCH 14, 2002

GAMMA-RAY LARGE AREA SPACE TELESCOPE
(GLAST)

LAT EVENT SUMMARY DATABASE
REQUIREMENTS

MARCH 14, 2002

NASA Goddard Space Flight Center

Greenbelt, Maryland

LAT EVENT SUMMARY DATABASE REQUIREMENTS DOCUMENT

Prepared by:

_____ Date _____

_____ Date _____

Concurrence:

_____ Date _____

_____ Date _____

Approval:

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

DOCUMENT TITLE: GLAST Science DATABASE Requirements Document**DOCUMENT DATE:** March 14, 2002

ISSUE	DATE	PAGES AFFECTED	DESCRIPTION
DRAFT	03/14/02		

TABLE OF CONTENTS

1	Purpose.....	1
2	Acronyms.....	2
3	Glossary.....	3
4	Applicable Documents	4
5	Requirements.....	5
5.1	Description.....	5
5.1.1	Event Summary Databases	5
5.2	Event Summary Database Requirements	5
5.2.1	General Database Requirements.....	5
5.2.2	HEASARC Compatibility Requirements	7
5.2.3	Cost for the photon summary database	Error! Bookmark not defined.
5.2.4	Cost for the comprehensive event summary database.....	Error! Bookmark not defined.
5.2.5	Performance for the photon summary database.....	7
5.2.6	Performance for the comprehensive event database.....	8

1 PURPOSE

This document defines the requirements that must be met by the GLAST LAT event Databases for serving GLAST data to the GLAST collaboration and the scientific community in general. This document specifies the requirements that these databases must satisfy.

This document first provides a list of acronyms (§2) and a glossary of key terms and concepts (§3). The relevant documents from which these requirements are derived are listed in §4. Finally, the actual requirements are presented (§5).

2 ACRONYMS

API	Application Program Interface
DEC	Declination
DBMS	Database Management System
GBM	GLAST Burst Monitor
GIOC	GBM IOC
GRB	Gamma-Ray Burst
GSFC	Goddard Space Flight Center
GI	Guest Investigator
HEASARC	High Energy Astrophysics Science Archive Research Center
IOC	Instrument Operations Center
LAT	Large Area Telescope
LHEA	Laboratory for High Energy Astrophysics
LIOC	LAT IOC
NASA	National Aeronautics and Space Administration
ODBC	Object Database Connectivity
OGIP	Office of Guest Investigator Programs
PI	Principal Investigator
RA	Right Ascension
SAA	South Atlantic Anomaly
SSC	Science Support Center
SWG	Science Working Group
TBD	To Be Determined
TBR	To Be Reviewed

3 GLOSSARY

Level 0 data—The raw data from the spacecraft after the telemetry packets have been time-ordered and repeated packets have been removed. The data streams from the spacecraft and the instruments are separated. The processing that produces Level 0 data is called Level 0 processing.

Level 1 data—Level zero instrument data, which has been processed into discrete particle events. The event tracks have been reconstructed and parameters from the tracks are extracted for use in scientific analysis. The processing that produces Level 1 data is called Level 1 processing.

4 APPLICABLE DOCUMENTS

- “GLAST Large Area Telescope Flight Investigation: An Astro-Particle Physics Partnership Exploring the High-Energy Universe,” P. Michelson, PI.
- GLAST SSC Functional Requirements Document, 433-RQMT-0002
- “Gamma Ray Burst Monitor,” C. Meegan, PI.
- GLAST DPWG (Data Products Working Group) Report, Draft 2/15/02, S. Digel
http://glast.gsfc.nasa.gov/ssc/Report_DPWG.pdf
- HEASARC FITS File Recommendations”
(http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html) By the HEASARC FITS Working Group
- LAT IOC (Instrument Operations Center) System Specification
- GBM IOC System Specification
- GLAST Level I Requirements Document
- Operations Concept Document, 433-OPS-0001
- MOC Functional Requirements Document, 433-RQMT-0001
- Science Requirements Document, 433-SRD-0001
- Project Data Management Plan (PDMP), 433-PLAN-0009
- LAT Science Analysis Software Requirements Document, LAT-SS-20.0
- LAT IOC Performance Specification - Level II(B) Specification, LAT-SS-15.1
- LAT Science Analysis Software Management Plan, LAT-MD-360.1

5 REQUIREMENTS

5.1 Description

The main publicly available scientific data from the GLAST LAT instrument will come in the form of Event Summary files generated by the LIOC. In order to make these easily useable (e.g., spatially searchable) they will be stored in a database. This document specifies the requirements on such a database.

5.1.1 Event Summary Databases

This is the generic term for databases that contain particle and photon events. The event summary data is expected to be about 90% charged particles and 10% photons. Since it is expected that the scientific community will be mostly interested in the photons, it is likely that the events will be separated by kind into different databases.

5.1.1.1 Summary Photon Database

This contains an abbreviated list of parameters (energy, time, direction, spacecraft position, etc. as specified in the DPWG report) for gamma rays detected by the LAT. These summaries will be delivered by the LIOC to the SSC. It is expected that this database will be the most heavily used database by the external scientific community.

5.1.1.2 Comprehensive summary event database

This is similar to the photon summary database, but includes all particle event summaries, (neutral and charged). Again, these summary files are described in the DPWG report as products delivered by the LIOC.

5.2 Event Summary Database Requirements

These requirements apply to both the summary photon and comprehensive summary event databases except in the area of performance and cost, where separate sections exist for the two different databases. The performance requirements for the photon database are much more stringent than for the comprehensive event summary database. Note that this does not preclude having a single database to serve both needs if an architecture could be found to meet the performance and cost requirements of both databases simultaneously (TBR).

5.2.1 General Database Requirements

These requirements are independent of the database architecture (including whether or not a DBMS is used).

5.2.1.1 Searchability Requirements

This section lists all of the requirements on the types of searches that the database must be able to perform be searched.

- 5.2.1.1.1 It must be searchable on fields that are integers, reals, dates, and times.
 - 5.2.1.1.2 Times must be searchable to microsecond precision.
 - 5.2.1.1.3 Must be searchable by 2-dimensional location on a sphere.
 - 5.2.1.1.4 Must be able to include data quality as a search criterion.
- 5.2.1.2 There must be an API that allows convenient access to the data in one or more programming languages commonly used in the GLAST LAT project (at least one of the following: C++, JDBC, perl DBI)
- 5.2.1.3 The database must be remotely accessible.

The database must be able to be remotely accessed by computer programs.

5.2.1.4 The database must be able to be Mirrored

There must be a procedure to send data updates to more than one copy of the database. These updates must be able to delete, add, and modify data in the database. These updates can be sent to mirrored copies of the database so that all databases can be exactly the same within 6 hours.

- 5.2.1.4.1 Must have a tool to verify that mirror sites are concurrent with master database.
- 5.2.1.5 It is preferable that real numbers in the database and in any indices use the native IEEE floating-point representation.
- 5.2.1.6 A credible path for upgrading software and/or hardware for the database must be specified. This implies the ability to run the database on standard computers (not specially configured to suit a single DBMS feature) and not writing software which accesses features available in only one single brand of DBMS. The consequence of this for the database is that it must be able to achieve acceptable performance through a relatively standard API. For SQL, this is SQL 99.
- 5.2.1.7 If a DBMS is used, it will be an ODBC compliant database (TBR).
- 5.2.1.8 The command language must be a dialect of SQL, close enough to the ANSI standard so as not to conflict with 5.2.1.6.

5.2.2 HEASARC Compatibility Requirements

- 5.2.2.1 The database must be able to be turned over to HEASARC staff to run after the mission ends.
 - 5.2.2.1.1 The database must not require a excessive amounts of effort (>1.0 FTEs) to maintain and operate.
 - 5.2.2.1.2 The database must not have large (> \$20000) yearly maintenance fees.

The HEASARC may place additional requirements on hardware and software they are willing to maintain. (TBD)

5.2.3 Performance for the photon summary database

5.2.3.1 Service time for public data requests

When the database is queried for photons, the resulting data must be available within a reasonable length of time.

- 5.2.3.1.1 Requests for data from standard searches (which we take here as containing < 300 Mb must retrieve all data in <30 minutes time from when searching begins, per years worth of data searched.
 - 5.2.3.1.1.1 The circular standard search consists of finding all photons coming from within a spherical radius of 15 degrees from some location on the sky sphere.
 - 5.2.3.1.1.2 The rectangular standard search consists of finding all photons coming from a spherical square whose area is equal to the standard search circle (a 15 degree radius circle is roughly 707 square degrees).
- 5.2.3.1.2 A subselection of data from a standard search, where filtering is done on all additional searchable parameters must be served within 45 minutes from the start of the search processing per years worth of data searched.
- 5.2.3.1.3 Requests for data > 2 times the number of years Gbytes should be serviced within 3 days. This requirement is meant to allow large requests to be processed during off-peak hours (which will be determined empirically) to prevent bogging down the database.

5.2.3.2 Number of service requests per day.

Must allow for the estimated number of daily requests of data. We make the specification in terms of

- 5.2.3.2.1 Must be able to do 60 standard search requests (for 1 years worth of data) in a day.
- 5.2.3.2.2 Multiple concurrent searches (at least 2) must be allowed without undue delay. Undue delay is defined as more than 50% (TBR) more time than processing the same search requests serially.
- 5.2.3.3 Ingest speed.

Data must be available for searching in the database 10 minutes after the start of the ingest process for a newly delivered photon summaries file. The photon summaries file is a filtered version (containing only photons) of the standard Level 1 event file delivered by the LAT IOC to the SSC (assuming the file contains no more than 5 hours worth of data).

5.2.3.4 Modification of data in the database

The following processes must be able to be carried out without undue (< 30 minutes) interruption to the access to the full database. If the database is broken into subsets, undue delay means either $< 30 / (\text{number of subsets})$ minutes or < 10 minutes, whichever constraint is longer.

- 5.2.3.4.1 Must allow ingest of new data at the anticipated data rates (about 300,000 events per day).
- 5.2.3.4.2 A reprocessed 5 hour time photon summary file must be able to be inserted into the database without undue interruption. (It must take less than 60 minutes time to make the new version of the data available for that interval.)
- 5.2.3.4.3 A complete re-indexing of the database after a crash must require less than 3 days time per year's worth of data.

5.2.4 Performance for the comprehensive event database

The comprehensive event database, while not the primary database for researchers outside of GLAST, still contains much important data useful for scientific analysis and must be accessible to external investigators, but does not need the same level of performance.

5.2.4.1 Service time for public data requests

When a query is made for data using some search criteria, that data must be available within a reasonable amount of time. Must be able to handle 1 standard search per day.

- 5.2.4.1.1 Requests for data from standard searches must retrieve all data in < 10 hours time per years worth of data searched.

- 5.2.4.1.1.1 The circular standard search consists of finding all photons coming from within a spherical radius of 15 degrees from some location on the sky sphere.
- 5.2.4.1.1.2 The rectangular standard search consists of finding all photons coming from a spherical rectangle of area equal to the standard search circle (707 square degrees).
- 5.2.4.1.2 Requests for standard searches where filtering is done on all parameters will be served within 15 hours.
- 5.2.4.1.3 Requests for data > 20 Gigabytes should be serviced within 7 days.

Requests for large amounts of data will require special treatment to avoid bogging down the database.

5.2.4.2 Number of service requests per day.

Must allow for 1 at least 1 request of < 3 Gb of data by per day.

- 5.2.4.2.1 Must allow for TBD requests of TBD Mb of data by TBD selection criteria.
- 5.2.4.2.2 Multiple concurrent selections (at least 2) must be allowed and processed without undue delay (< 100% more time than the time required to do the same requests serially).

5.2.4.3 Ingest speed.

Data must be available for searching in the database 100 minutes after the start of the ingest process for a newly delivered event summaries file. The file referred to here is the standard Level 1 product delivered by the LAT IOC to the SSC (assuming the file contains no more than 5 hours worth of data).

5.2.4.4 Modifying data in the database

The following processes must be able to be carried out without undue interruption to the access to the full database.

- 5.2.4.4.1 Must allow ingest of new data at the anticipated data rates (about 3 million events per day)
- 5.2.4.4.2 Reprocessing an existing 5 hour time interval of data must be able to be done without undue interruption (< 200 minutes)
- 5.2.4.4.3 A complete re-indexing of the database after a rebuild must require less than 1 weeks time per years worth of data. (TBR)

