Created:June 01, 1997
Modified:August 12, 1998

Gemini Engineering Archive Report

Engineering Archive Requirements

Phil Puxley, Dayle Kotturi, Jim Oschmann, Bret Goodrich, Daniel Durand, David Schade, David Bohlender and Séverin Gaudet

SPE-C-G0073/02

This document outlines the requirements for an engineering archive.

1.0 Introduction

This document outlines the requirements for the GEMINI telescopes engineering archive (GEA). The function of this archive is to collect data from all of the GEMINI systems and to allow their efficient retrieval for the purpose of analysing the system behaviour and thereby to optimize the scientific output of the GEMINI telescopes. This archive system is particularly important during the commissioning phase, but is an essential component during the entire lifetime of the observatories.

To avoid duplication, the GEA may contain pointers to data in existing databases such as the DHS permanent store and Observing Database. The DHS permanent store contains all image data, image processing histories and recipes, and completed science programs. The Observing Database contains planned and active science program information.

The GEA will contain a superset of the science data which, during Gemini operations, will be available via the science archive.

2.0 Users

Users that might access the engineering archive in order to fulfil their duties are:

1. Instrument and telescope engineers: To check and analyse the behaviour of the instruments, the telescope, the entire telescope configuration and the environment.
2. System support associates: To analyse performance and/or compare with earlier experiences.
3. Resident scientists: To obtain scientist contact information, to analyse overall science throughput, weather, etc.
4. Director: To monitor site usage, efficiency, global usage, observatory statistics, etc.

In general, the EA is not intended to be accessed directly by the Principal Investigators (PIs), although some queries might be possible, via the Gemini scientific staff. External access to archived data will generally be through the science archive.

3.0 Gemini Engineering Archive User Requirements

3.1 Gemini Engineering Archive Data

This section lists and defines the categories of data to be available from the GEA. All data must be timestamped, or have a time associated with it, to denote the period of validity.

3.1.1 SCIENCE data

- pixel data and header meta data, includes calibration pixel+meta data.

3.1.2 WEATHER SERVER data

- sky images including weather maps, various temperature, pressure, humidity, windspeed sensors, seeing monitor

3.1.3 ENGINEERING data

- EPICS records which characterise the state of the system, includes *one-off* monitoring or logging activities

3.1.4 AUXILIARY data

- miscellaneous aspects of the system configuration or information, including software versions, filter and grating curves, narrative observing log, Phase I proposal info, Observing tool ASCII science program definition

3.2 GEA queries

There are many, maybe infinite, possibilities for archive queries. Fundamentally the queries should use the timestamp to associate data in different categories and from different sources.

Here are some example queries to give some idea of the scope:

(Qa) Retrieve all values of a specified engineering parameter between times time_1 and time_2.

(Qb) Retrieve the names of all of the engineering parameters which were archived between times time_1 and time_2.

(Qc) Retrieve all science data, all-sky weather images, seeing monitor parameters and observing log between times time_1 and time_2. [This example is also appropriate for the one-off transfer of commissioning data into the science archive, once the latter is implemented, but conceivably might also be the mechanism for the general transfer of data from GEA to the science archive or for distribution of data to users]

(Qd) Retrieve all raw science data and mirror temperatures involving the near IR imager within times time_1 and time_2.

(Qe) Retrieve the current NIRS grating transmission curves.

(Qf) Retrieve all observing log information entered by Puxley within the last 6 months.

(Qg) Retrieve all science data for object name *M82*.

As a crude benchmark, a query of the database such as *Qa* or *Qc* should take not more than 5 minutes for a search time interval of 6 months.

Commonly, the next stage in the use of the retrieved GEA data will be to analyse, say, the engineering parameters or elements of science meta data, with popular third-party software (e.g. Microsoft EXCEL).

3.3 Data ingestion considerations

For each data category we give examples of the content, dataset size and likely data rate.

3.3.1 Science data

This category contains all the raw and processed pixel data and associatedheaders stored by the DHS. The time of validity is contained within the header e.g. as the shutter open/closed or exposure start/stop times. It is assumed that these data will have been stored by the DHS in its *permanent store*.

Raw pixel data rates and volumes by instrument, estimated by Simons (TN-PS-G0048) range from 1 (MIRI) to 40 (GMOS) Gbit/night. Allowing another factor of 30% to include processed data products, a median data rate might be 30 Gbit/night.

The header meta data for each file (i.e. one per OBSERVE command) may typically consist of order 100 FITS keywords and values.

It is required that the science pixel data for at least the last 6 months be available on-line. A catalogue containing all of the header data ever acquired should be on-line. It is acceptable to mount removable media to retrieve old pixel data (older than 6 months) identified from query of the header catalogue.

3.3.2 Weather server data

These are included with the engineering data below.

3.3.3 Engineering data

The intent is to specify a set of engineering parameters that characterise the system and which will always be archived (unless they are remove from the list, of course). The GEA should allow this list to be readily extended if it becomes apparent that other parameters are needed, for example by selecting from a list of all archived parameters.

There will also be occasions when it is neccessary to monitor or log specific sets of engineering parameters, or WFS images, perhaps at high rates (say, 20Hz-1kHz). Typical scenarios are:

- monitor a list of PVs for changes,
- read a list of PVs at a specified rate,
- monitor/read a list of PVs for 10 seconds,

- monitor/read a list of PVs from time_1 to time_2.

The GEA should support the permanent storage of these datasets (but it would be acceptable to rely on another system for buffering the data) and their retrieval as described in the query section above.

The following table gives examples of the engineering parameters to be archived, as an illustration of the scope, their data type and typical rates.

Parameter or source	Number of items	Data type	Storage rate
temperatures	500[a]	number	once per minute
windsensors	5	number	once per minute
humidity	3	number	once per 10 minutes
encoder average uncertainty	10	number	once per 10 sec
WFS performance parameters	10	number	once per 10 sec
WFS images + header (meta)[b]	2	image (1kx1k pixels)	once per hour
WFS images +header (meta)	6	image (80x80 pixels)	once per hour
miscellaneous systems monitoring parameters	500	number	once per minute

a. Expected to decrease to, perhaps, 100 items by operational handover.

b. It would be acceptable for the image data to be stored in the same way as the science data e.g. via the DHS in the permanent store, with retrieval via a pointer from the GEA.

3.3.4 Auxiliary data

This is a catch-all category for everything else. In general the archiving should be triggered by the event e.g. change to NIRI filter complement, acquisition of a new GMOS grating, generation of a narrative observing log entry etc. This category includes the Phase I and Phase II science application and program information.

The instrument *static* information (grating curves, filter profiles etc) will not change frequently. The instrument complement (probably ascii files denoting which filter or grating is in which slot, which instrument is on which port etc) might change after each engineering run on that instrument. Likewise for adoption of a new software version. The data rates are therefore expected to be small.