

Handling Chunks of Image Data in the Gemini Data Handling System

Jennifer Dunn, Shannon Jaeger, Norm Hill, Séverin Gaudet, Steve Cockayne

*National Research Council Canada/Herzberg Institute of Astrophysics,
5071 West Saanich Rd., Victoria, BC, Canada V8X 4M6*

Abstract. The Data Handling System (DHS), developed for the Gemini project by the Canadian Astronomy Data Centre, accepts observational data from multiple sources. There may be data descriptions, image data in chunks (to be composed into a single image), and multiple images for one dataset, all being received at different times. All information for a single dataset is written to one Flexible Image Transport System (FITS) file. The methods developed for handling and assembling data chunks are presented.

1. System Overview

The role of the DHS in the Gemini Control System is to manage the flow of data from the instruments on the Gemini Telescopes. DHS is data-driven and is composed of a collection of servers that work together to act on any data it receives from any Gemini principal system. See Hill et al. (1999a) for more information on the DHS clients and servers.

The Data Server is responsible for handling data put and get requests and is responsible for managing the data. This paper focuses on the handling and assembling of datasets by the DHS Data Server.

1.1. Communication

All communication with the Data Server is provided by the DHS library. The library provides mechanisms to clients for establishing connections to the Data Server, creating and decoding dataset structures and sending and receiving dataset structures from the DHS Data Server over the Gemini Data LAN. The library is built on the DRAMA Interprocess Message Passing (IMP) and Self-defining Data System (SDS) packages, developed by the Anglo-Australian Observatory. See Hill et al. (1999b) for more details on DHS communications.

2. Data Contributors & Receivers

The Data Server receives data from multiple sources and sends data to several of the DHS subsystems. Figure 1 shows some of the expected contributors of

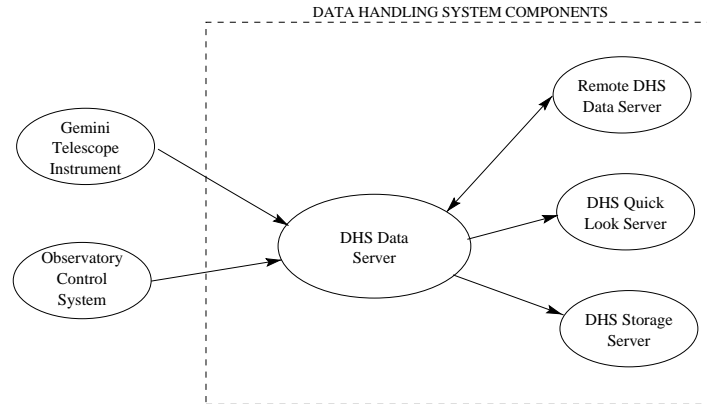


Figure 1. DHS Data Server Context Diagram.

data to the Data Server and the DHS subsystems which will receive data from the Data Server.

2.1. Contributors

The multi-threaded capabilities of the Data Server allow it to handle

1. multiple contributors/sources to a single dataset,
2. different datasets at the same time,
3. and multiple contributors to the same dataset at the same time.

Here, a dataset is considered to be data from a single OBSERVE command.

The flexibility built into the Data Server allows attributes and/or data to be sent to the Data Server all at once or in chunks. When all data have been received by all contributors, the Data Server will assemble the chunks of the jigsaw puzzle to construct a complete dataset.

2.2. Receivers

Each piece of information, whether it is a chunk or complete set, is immediately converted into the Flexible Image Transport System¹ (FITS) format and sent from the Data Server to the Quick Look Server. The Quick Look Server will then send these data to its Quick Look Tool(s), which display the images graphically (Hill et al. 1999b). The DHS Storage Server will receive notification of complete datasets that are destined for permanent storage.

3. Data Formats

Data are sent to the DHS as an SDS structure, which is a hierarchical format. Hierarchical structures are necessary to support the many kinds of data which the Gemini instruments may generate. Data are then stored in a flat FITS format by the Data Server. The SDS format is used because it is part of the message protocol used between DHS and external sources. FITS was chosen

¹<http://heasarc.gsfc.nasa.gov/docs/heasarc/fits.html>

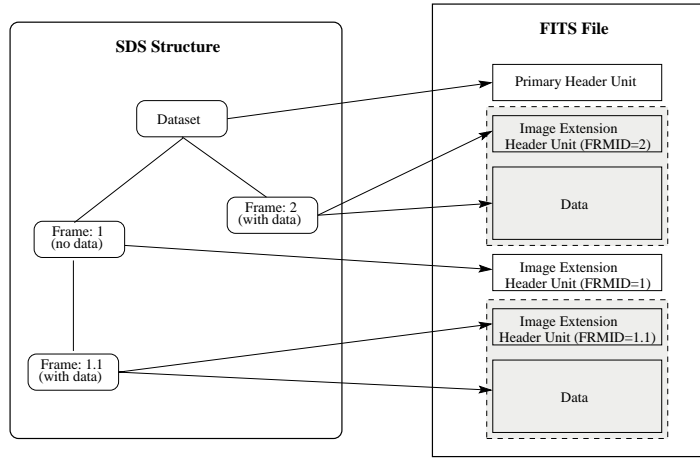


Figure 2. SDS to FITS mapping.

because it is the standard file format used within the astronomical community. All FITS file manipulations use the `cfitsio` library (Pence 1995).

A dataset is the data resulting from an OBSERVE command. A dataset generally consists of header data describing the dataset and one or more frames. A frame is header data and a single array of pixel data. The frame can represent a single exposure or several exposures. It can also be bad pixel masks, quality data, variance data, etc. Frames contain a frame identifier, a frame name, attribute/value data describing the frame, an optional single data array, and an optional set of sub-frames. The hierarchical definition of frames allows related frames to be grouped in a logical way. Header data is a collection of attribute/value pairs which describe a dataset or a frame within a dataset. These data structures are designed to allow for the transmission of arbitrarily complex data.

4. Conversion Process Details

One of the difficulties encountered when designing this system was determining how to preserve the hierarchical information within a flat FITS file format. Figure 2 shows how the SDS hierarchy and data are mapped into FITS files. Note that since data can arrive in any order, Frame 2 actually appears before Frame 1 in the final FITS file.

Each dataset is stored in a single FITS file. A FITS image extension will be created for each frame found in the dataset. The hierarchical information is stored in the FITS keyword `FRMID` which is a value such as 1, 1.1, or 2. Note that if a frame has no associated data then there will be no data following the FITS image extension header.

The Data Dictionary is a configuration file used by the Data Server to map the SDS attributes to FITS keywords. The Data Dictionary is maintained by the instrument support scientists. The properties of the header attributes include the specification of how the data will be stored in the FITS file. For example,

single value attributes are stored as FITS keywords, one-dimensional arrays as indexed keywords and multi-dimensional arrays as FITS binary extensions.

Now how does the Data Server do this conversion? As each chunk arrives, it is converted into FITS for viewing by the Quick Look subsystem. During this first conversion information is gathered about the dataset:

1. which instrument sent the data,
2. how many frames there are,
3. the size of each frame's data array,
4. the number of keywords in each extension and in the primary header unit,
5. and the frame IDs and frame names for each frame.

Once all chunks have been received from all contributors, the chunks are then converted into a single FITS file. At this point the Data Server has all the statistical information about the dataset and can make the conversion for the final FITS file without any reallocation of memory.

5. Conclusion

The Data Server was designed long before any of the Gemini instruments were available for testing. As a result, one of the main design criteria was to make it as flexible as possible in order to meet all potential instrumentation needs. With flexibility came complexity. There are multiple configuration files to be filled out prior to running, with the most important being the Data Dictionary. The use of the Data Dictionary has avoided hard-coding attribute/keyword pairs leaving the Gemini staff and instruments to choose them as they wish.

In November 1998, the Gemini Data Handling System had its "first dark" picture using the DHS with a Gemini High Resolution Wavefront Sensor. It is called "first dark" because the calibration lamp was switched off, and it obtained a dark image contaminated with room light. First light for the Gemini system is scheduled to be January 1999. One advantage of using the DHS is that the DHS Data Server will convert the data and send it on to the Quick Look subsystem and display each new frame automatically.

References

- Hill, N., Gaudet, S., Dunn, J., Jaeger, S., & Cockayne, S. 1999a, this volume, 155
———, 1999b, this volume, 163
Pence W. 1995, *Fitsio User's Guide*, Goddard Space Flight Center