ISDC Data Access Layer

D. Jennings, J. Borkowski, T. Contessi, T. Lock, R. Rohlfs and R. Walter

INTEGRAL Science Data Centre, Chemin d'Ecogia 16, Versoix CH-1290 Switzerland

Abstract. The ISDC Data Access Layer (DAL) is an ANSI C and FORTRAN90 compatible library under development in support of the ESA INTEGRAL mission data analysis software. DALs primary purpose is to isolate the analysis software from the specifics of the data formats while at the same time providing new data abstraction and access capabilities. DAL supports the creation and manipulation of hierarchical data sets which may span multiple files and, in theory, multiple computer systems. A number of Application Programming Interfaces (APIs) are supported by DAL that allow software to view and access data at different levels of complexity. DAL also allows data sets to reside on disk, in conventional memory or in shared memory in a way that is transparent to the user/application.

1. Introduction

INTEGRAL[1] is an ESA Medium Class gamma-ray observatory mission scheduled for launch in 2001. It consists of four co-aligned instruments: two wide field gamma-ray detectors (IBIS, SPI) an x-ray monitor (JEM-X) and an optical monitor (OMC). The IBIS imager, SPI spectrometer and JEM-X x-ray monitor all employ coded mask[2] detection technology which, amongst other complexities, requires the spacecraft to constantly "dither" its pointing position in order to accumulate the required number of coded sky images for analysis.

The dithering strategy creates some unique issues for data organization and analysis. Each celestial *observation* will consist of a collection of *pointings* (5 minute to 2 hour fixed position integrations); conversely, it is possible that a given pointing may belong to several observations at once. A pointing is itself a collection of science, instrument housekeeping and auxiliary data sets, all of which may be grouped into various combinations for reasons of efficiency and conceptual elegance. Thus, the INTEGRAL data analysis system must be capable of supporting distributed data sets composed of many individual files and exhibiting *one to many* and *many to many* associations between the individual data structure elements.

---

[1]http://astro.estec.esa.nl/SA-general/Projects/Integral/integral.html

[2]http://lheawww.gsfc.nasa.gov/docs/cai/coded.html

2. ISDC Data Model

The complicated relationships between observations, pointings, and pointing data set components implies a natural *hierarchy* to the INTEGRAL data. This has led to the current ISDC Data Model which generalizes the concepts of pointings and observations into *Data Objects* and *Data Elements.*

A Data Object is an association, or collection, of one or more Data Elements. A Data Element may itself be Data Object (i.e., a sub-collection of data elements with lower position in the hierarchy) or a terminal *Base Element* containing an atomic data structure. There are three classes of terminal Base Elements that define the atomic data structures: collections of data in tabular (row, column) format known as *TABLE* elements, N dimensional data sets of homogeneous data type known as *ARRAY* elements, and human readable and/or bulk data (e.g., programs, text, GIF images, PostScript output) known as *INFO* elements. In addition to the atomic data structures, there is a fourth class of Base Element used to define compound (i.e., non-atomic) data structures known as *GROUP* elements.

The recursive nature Data Elements allows for the construction of *unbounded hierarchical associations* of atomic data structures. At the opposite extreme, a single atomic data structure may also be considered a Data Object in its own right. Note that it is also possible for a Data Element to belong to many different Data Objects, thus allowing Data Objects to share a given collection of data.

3. The DAL (Data Access Layer)

To implement the ISDC Data Model within the INTEGRAL analysis system, ISDC is currently constructing the DAL, or *Data Access Layer*. The DAL allows applications to create and manipulate data sets at the Data Object and Data Element level of abstraction.

The DAL consists of four logical layers. The physical format I/O modules CFITSIO[3] (for FITS) and SHMEMIO (for shared memory resident data) make up the first layer. These modules handle all the details of the particular storage formats available though DAL. Since the format specific details are isolated in this manner it is possible to add other data format capability with no change to the higher level software layers.

Above the physical format modules are the driver interface modules: FITSdriver, MEMdriver and SHMEMdriver. All these drivers contain an uniform set of interface functions that implement the respective data storage methods (FITS resident data, memory resident data and shared memory resident data). By using the driver level modules it is possible to provide the higher level DAL layers with consistent interface calls regardless of the storage medium in use.

The next DAL layer, and the first used in application programming, is the base element API level. Each of the four base elements supported by DAL (ARRAY, TABLE, INFO and GROUP) has its own Application Programming

---

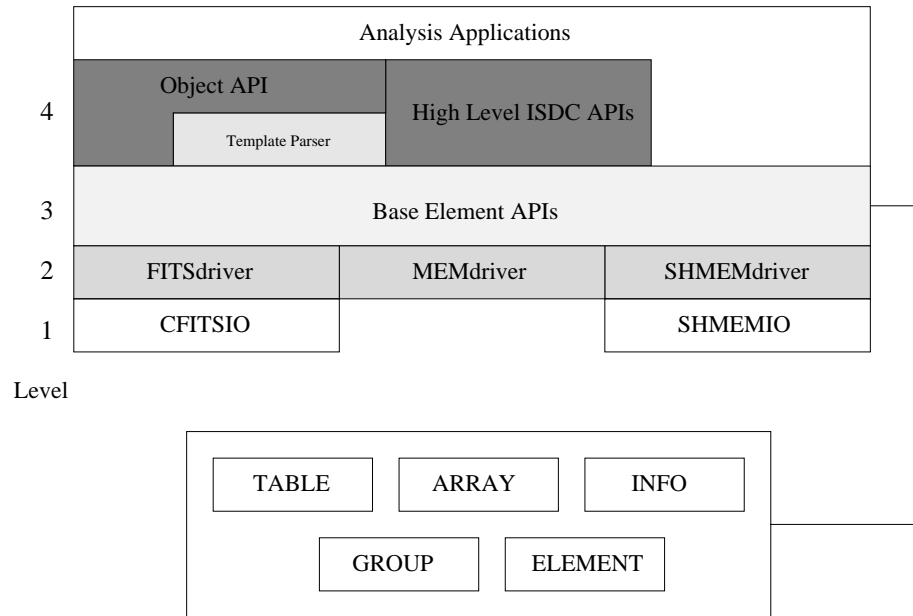[3]http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html

Figure 1.   The Data Access Layer four level API structure, with the analysis applications residing above.

Interface that implements the element type. There is also a fifth API at this level, the ELEMENT API, that allows applications to operate upon individual data elements regardless of the base type.

The top level DAL layer is a collection of APIs that allow analysis applications to make efficient use of the base element APIs. The Object API implements hierarchical associations of data elements (i.e., the data model Data Objects), and the high level ISDC APIs implement scientific data-specific collections of data elements (e.g., spacecraft attitude and orbit, spectra, event lists).

4.   DAL Data Objects: Data Format vs. Data Model

In order for DAL Data Objects, specifically the associations between the data elements, to be *persistent* with time the physical data formats must support the ISDC Data Model. For Data Objects stored on disk in FITS format the FITS Hierarchical Grouping Convention[4] is utilized to achieve Data Format – Data Model conformity.

Each DAL Data Element is stored in a FITS file as a HDU (Header Data Unit). Associations between Data Elements are stored in *grouping tables* as defined by the Grouping Convention. DAL manages all the details involved in creating and updating the grouping tables when a Data Object is written to disk in FITS format, and it attempts to locate and read grouping tables when opening an existing FITS-stored Data Object.

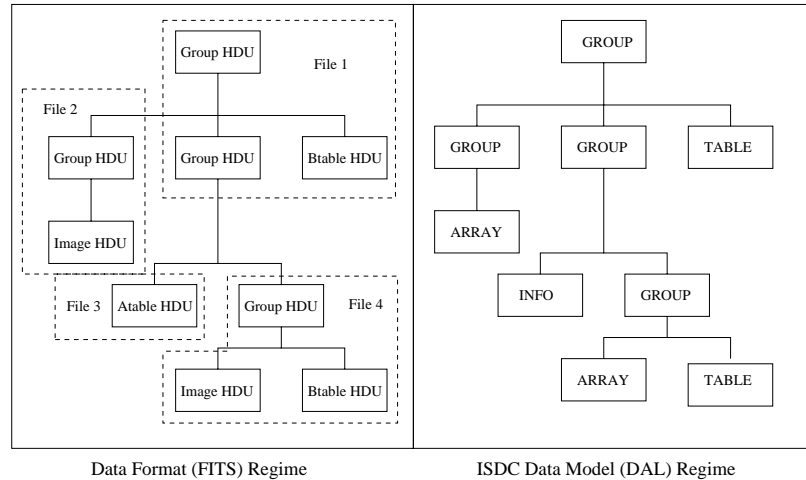[4]http://adfwww.gsfc.nasa.gov/other/convert/group.html

Figure 2.   Data format – data model conformity. The lines connecting the FITS HDUs (left) and DAL data elements (right) define the hierarchical relationships between the data structures.

5.   Conclusion

The DAL concept as presented here has several important implications that are of general utility to astronomical software.

First of all, the data format specifics are hidden from the software application, thus allowing the same API functions to be used for a variety of data formats and access methods. Data residing in memory, for instance, may have a different storage paradigm than data residing in disk files. Also, many different file-based formats (FITS, HDF, CDF) may be used transparently by the analysis applications.

Secondly, it is possible for Data Objects to span multiple data files in a way that is *inherent to the data structure itself*. The relationship between data structures persists even when the supporting archival infrastructure changes (e.g., port to a new OS) or is eliminated entirely (i.e., as can happen at the end of a space mission), thus providing for a *self-contained data organization*. Data Objects may also in theory span multiple computer file systems in a way that is transparent to the analysis applications. The implications for distributed network data archives are significant.

Lastly, different institutes may construct their own DAL-like data interface packages to achieve their own specific goals, but still cooperate to promote commonality across the data formats/structures. The commonality may be added or modified without changing higher level software.

The latest releases of DAL software and documentation may be obtained on line from the INTEGRAL Science Data Centre[5] Web site. Questions or comments regarding DAL should be directed to Don.Jennings@obs.unige.ch.

---

[5]http://obswww.unige.ch/isdc/