Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

The Integral DAL 1-2-3

B. O'Neel, D. Jennings, R. Rohlfs, S. Paltani

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

Abstract. This paper will discuss the three layers of the Integral Science Data Center (ISDC)[1] Data Access Layer (DAL)[2]. The first layer is HEASARC's cfitsio[3] with contributions from ISDC to allow for network and shared memory files as well as FITS grouping tables. The network files allow transparent access to distributed files via HTTP, FTP and the ROOT[4] protocols. The second layer is a mission independent abstraction layer which hides the details of files and FITS HDUs from the program allowing them to work with abstract objects and elements. One of these abstract objects is that of a GROUP which allows you to collect together all the FITS data structures which are related in some way and independent of the exact files they are located in. The third layer is an instrument specific layer tailored to the Integral mission and instruments.

1. Introduction

ISDC has produced a data access layer (DAL) which allows data analysis programs to be written which read and/or write data stored in FITS format but which do not depend on the programs knowing exact locations of either the FITS files they process or the FITS Header Data Units (HDUs) within the FITS files. In addition, for many cases, the scientific end user only has to remember and keep track of one file name. This is all done by using FITS groups and abstracting the data access at a higher level than that of FITS files and FITS HDUs.

The main goal of this was to decouple data access from data analysis. We did this by creating software libraries that both define and isolate the data format, data model, and instrumental specifics, in our case, for Integral. This allows us to keep analysis software development at a high of level as possible. The programmers work with events, images, spectra, lightcurves and other scientific data structures and not with FITS files and HDUs. One consequence of this approach is that it allows changes to the data format, data model or data

[1] `http://isdc.unige.ch`

[2] `http://isdc.unige.ch/bin/sw2html.cgi`

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

[4] `http://root.cern.ch`

implementation by making software changes in one place instead of retrofitting every program. In addition, it allows us to share software development with other data centers where possible.

2. The DAL Layer 1 - cfitsio

DAL layer 1 is cfitsio from the HEASARC[5]. All files in the ISDC DAL are stored in FITS format and FITS is also the processing format. By using cfitsio we use a large body of robust, well tuned and well tested code. The ISDC DAL isolates all file level access in this level.

ISDC contributed FITS grouping (Jennings 1997), template parsing, network file access, regular memory, and shared memory file access to cfitsio. Each one of these additions brings interesting new features to cfitsio. In addition, since these features are added to cfitsio all programs which use cfitsio automatically have access to these features.

2.1. FITS Grouping

FITS HDU grouping lets one group multiple FITS HDUs together in a hierarchy. This allows one to group together logically related FITS HDUs in scientifically interesting ways. One could group together all the files from a particular observation or pointing along with the relevant calibration files. This grouping can occur with no changes to the original FITS files.

2.2. Template Parsing

FITS templates allow you to create FITS files from text templates with one call. This allows one to have a consistent layout to your FITS files and gives one a centralized place to add or remove keywords and/or columns. FITS templates also reduce the amount of code to be written since the one cfitsio call write the basic FITS file.

2.3. Network File Access

Network file access allows one to use a FTP or HTTP file anywhere one could normally use a read/only FITS file. In addition the ROOT protocol is supported for read/write access across the network.

2.4. Shared Memory and Regular Memory File Access

Shared memory file access allows one to create and access FITS files stored in either shared memory or regular memory. This allows one to avoid disk i/o with your FITS files. These are identical except that the shared memory files exist between programs while the regular memory files are not accessible once a program exits.

---

[5] `http://heasarc.gsfc.nasa.gov`

3. DAL Layer 2 - The ISDC DAL

The ISDC DAL provides the Integral data model implementation. It generalizes data access to data objects containing groups of atomic data structures such as tables and arrays. These data structures are arranged into hierarchies that may span multiple data files. In general the scientist or programmer only needs to worry about one file which simplifies both the programming as well as the resulting scientific analysis. All data structure level access is isolated in this level.

3.1. The Advantages of Using the ISDC DAL

The biggest advantage of the ISDC DAL is that programs no longer have to manipulate files, rather, your program manipulates FITS tables or FITS arrays. As a result of no longer manipulating files all of your data can be in one or many files as the user chooses. Files also can be grouped in scientifically useful ways. As an example, the proper set of calibration files can be attached to an observation. Files need not even be on the same system but rather transparently accessible over the network via FTP or HTTP. And finally, while the ISDC DAL was designed to support Integral it is system and mission independent.

4. DAL Layer 3

The ISDC DAL Layer 3 (DAL3)[6] provides an instrument specific data structure implementation for Integral. This allows analysis software to work with scientific data items such as events, images, and spectra for a specific instrument rather than with FITS arrays and tables. Since the exact implementation of the on-disk data structure is hidden from the programs they don't have to know or care exactly how the data structures are stored. While the ISDC DAL3 is system independent it is very Integral mission dependent.

4.1. The Advantages of Using the ISDC DAL3 Concept

The programmers or programmer/scientists can think in terms of event lists, spectra, and light curves rather than FITS data structures. In addition, this allows for possible changes over the life of the mission. For example, it is possible that at the beginning of a the mission one of the columns in one of the tables could be stored. Later in the mission it could turn out that the column would be easier to work with if it was calculated rather than stored. The DAL3 concept isolates in one place the code which will have to change in this case.

5. Main Payoffs of DAL 1-2-3

The DAL 1-2-3 concept has made it possible to introduce changes to the data format, data model and instrument data characteristics in a centralized way that minimizes total project code changes. This has also given ISDC and the

[6] `http://isdc.unige.ch/bin/sw2html.cgi`

different Instrument Teams central control authority over the data model, data formats and data products. Since programmers and scientists can think of their data in scientifically useful products rather than just FITS arrays and tables it has made the writing of analysis software easier.

Finally because analysis software depends on being given a DAL group and this group will contain all the different data structures which are needed it has been easier to integrate software into packages and pipelines. Scientists should also find it easier to use the final software because one should only need to keep track of one file name rather than lists of files. In order to aid this, ISDC is working on an Interactive Browser to browse DAL groups and help you analyze data.

6. Conclusion

The DAL 1-2-3 concept has proven to be very useful for Integral and should be useful for many missions. It gives a common model for data centers to share development work, data models, and data format drivers. Finally, while there might be more work up front, in the long run this should help to reduce one's work to maintain mission specific as well as mission independent tools.

Acknowledgments. We are grateful to Dr. Bill Pence[7] for both cfitsio and his willingness to accept external contributions.

References

Jennings, D. 1997, A User's Guide for the Flexible Image Transport System (FITS), Version 4.0[8], NASA/GSFC Astrophysics Data Facility, Section 5.3

[7] `mailto:William.D.Pence.1@gsfc.nasa.gov`

[8] `http://fits.gsfc.nasa.gov`