Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

ASC Data Structures and Model

M. Conroy, R. Simon, J. McDowell

SAO/ASC, 60 Garden St., Cambridge, Mass. 02138

K. Barry

TRW/ASC, 60 Garden St., Cambridge, Mass. 02138

Abstract. The *AXAF* Science Center (ASC) is using an "open architecture" approach to develop the data analysis system. The Dynamic Data File (DDF) abstraction is being developed to provide sophisticated, format-independent filtering capabilities and support for multiple file formats. The DDF interface library will use the Open-IRAF data structure libraries to provide environment independent access to the DDF data structures from any C or FORTRAN host-level application.

1. Introduction

The characteristics of the X-ray data model and the goals of the ASC data analysis system, combined with the concept of an open architecture, lead to the design of the Dynamic Data Format (DDF) abstraction and interface library. DDF isolates the user from the details of the underlying implementation, while allowing the same analysis programs to accept a variety of data input formats, such as FITS and QPOE.

2. X-Ray Data Model

The definition of an X-ray data model consists of abstract data structures, access functions, and an astronomical interpretation (Farris & Allen 1992). Observational data in X-ray astronomy can be divided into the following data components (each decomposable into data structures): The primary data contain the science information and traditionally appear as either photon event-lists or image arrays. The ancillary data consist of supporting engineering and housekeeping observational information, such as voltages and satellite aspect. The derived data record the results of the analysis of the primary data, such as extracted spectra and detected source lists. The associated data represent information needed for correct interpretation of any of the above quantities, such as units, uncertainties, and errors. Another item, volumes, can be thought of as a quantity to represent the size of the sample space, e.g., exposure time. The meta-data are needed to link all of these data together.

3. Primary Data Access Requirements

The primary data-access functions have received the most attention in current data analysis systems. Most of these access requirements are now well understood, and a base of software tools is accumulating. For example, PROS uses the IRAF QPOE data format to store primary data and provides for most of the access requirements. These requirements include the data access functions for an event-list, which must support a photon record with arbitrary attributes. The photons of the event-list must be selectable both by attributes in the file, such as energy and time, as well as engineering attributes from ancillary files, such as the master-veto rate and the viewing geometry. The selection must be possible without intermediate files and the access must be able to produce both image and event-list output.

The data must also be interpretable in multiple World Coordinate Systems (WCSs), as well as in stored-data coordinates. Items such as exposure time and energy band must be retrievable dynamically, to reflect the selection criteria described above. The event structure must also be extensible, with derived quantities. Finally, derived data must be extractable from the primary events. Data such as light-curves and spectra are derived from the primary data by binning on one or more attributes.

4. Ancillary and Derived Data-Access Requirements

The ancillary and derived data-access functions have received less attention in current data analysis systems. The access functions that do exist are, for the most part, independent of the primary data. The software is less well developed and the access functions to connect these data to the primary data are less mature. For optimal results, the ancillary data lists must be accessible with the same filters as the primary data. Also, the WCS conventions must be extended from the primary data to these data, and include non-celestial coordinate systems such as interpolated functions and sampled points. As well, the observation-specific calibration data must be generated by matching data selections from both primary and ancillary data. The resulting calibration files will then correctly correspond to the filtered observation of the primary data. Finally, the low counting statistics prevalent in X-ray observations dictate the need to calculate uncertainties of all data quantities, and to propagate these uncertainties during the analysis process.

5. Dynamic Data Format (DDF)

One of the key developments of the ASC data system is the DDF data abstraction. The principal features of this abstraction are that the software abstraction allows applications to be completely independent of the underlying physical format. Changes required by a change in physical file format are isolated to the DDF library implementation. In turn, the DDF library implementation can support more than one physical format. The current design plans include support for FITS, QPOE, image, table, array, and ASCII list. The DDF library also makes extensive use of the Open-IRAF data structure libraries. These libraries de-couple the data structures from the IRAF environment, making them exportable and available to stand-alone applications.

Filtering capabilities must be an integral part of the data access I/O routines rather than a feature layered on top of the I/O. Otherwise, applications are required to have knowledge of the file format or contents. The DDF library design accepts a filter specification string with every data retrieval function, thus allowing the user to tailor filter specifications to the content of each file, without application code modifications.

The DDF abstraction, as well as the underlying physical formats listed above, support an extensible, self-defining data structure. This allows application code to be completely independent of the items or data-types contained in the physical file. The DDF design, by pairing a data filter expression with every data retrieval function, allows the application to process a virtual file. That is, the application sees a data file that contains only the records specified by the filter string, thus eliminating the need to create a physical instantiation of the filtered file, unless specifically desired. Many summary quantities describing a data file are functions of the file selections. Access to a virtual filtered file requires the ability to define summary functions and to evaluate them dynamically.

6. Technical Issues to Resolve

Many of the basic functional elements for the DDF implementation already exist in various forms. However, there remain some technical issues to resolve before the design is complete. For example, PROS uses the STSDAS TABLE format to store ancillary and derived data, but this format provides a different set of capabilities from the QPOE structure used for the primary data.

Some utilities that allow the extension of an event structure exist, but more sophisticated tools that allow creation of a new event structure by selecting and merging items from multiple existing lists are needed. Also, the mechanism for linking the primary event-list to ancillary lists and calibration files is missing. The mechanism for associating and propagating uncertainties is a topic of prototyping activities.

Work also continues on FITS/WCS support for image format in the FITS community (Greisen & Calabretta 1995). Work to extend the FITS WCS conventions to TABLE and BINTABLE formats is being pursued at SAO/ASC and HEASARC for the High-Energy Astrophysics domain (Corcoran et al. 1995). Finally, mapping the DDF data abstraction to FITS needs further development. Current work on FITS/HDF interoperability (Jennings, Pence, & Folk 1995), and the work in the AIPS++ group to map their class libraries to FITS are both promising directions.

7. Status

There are several efforts currently underway that represent necessary steps towards a complete DDF design and implementation. The ETOOLS ADP grant is supporting the development of Event Tools as a joint project between CEA/Berkeley and SAO/Cambridge. The products of this work will include:

a package of QPOE support tools, a QPOE browser (with a GUI), and a delivered C-callable library for the QPOE API. Projects are also underway to integrate IRAF TABLES and QPOE data formats. The STSDAS Group at ST ScI is extending the existing tables support to store a TABLES-compatible data structure within an IRAF QPOE file. Similarly, the IRAF Group at NOAO plans to develop a Common Data Format (CDF) that will allow the storage of additional data structures within an IRAF physical data file, such as QPOE.

Work to produce a complete data model of the X-ray data is still preliminary, but data modeling efforts by the AIPS++ project (Farris 1993) offer a promising direction.

Acknowledgments. This work was partially supported by NASA contract NAS8–39073. We also thank A. Farris, B. Glendenning, and G. van Diepen of the AIPS++ project for their generous assistance with astronomical data modeling.

References

Greisen, E., & Calabretta, M. 1995, this volume, p. **??**

Corcoran, M., Angelini, L., George, I., Pence, B., McGlynn, T., Mukai, K., & Rots, A. 1995, this volume, p. **??**

Jennings, D., Pence, W., & Folk, M. 1995, this volume, p. **??**

Farris, A. 1993, in Astronomical Data Analysis Software and Systems II, ASP Conf. Ser., Vol. 52, eds. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco, ASP), p. 145

Farris, A., & Allen, R. J. 1992, in Astronomical Data Analysis Software and Systems I, ASP Conf. Ser., Vol. 25, eds. D. M. Worrall, C. Biemesderfer, & J. Barnes (San Francisco, ASP), p. 157