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.

Convert: Bridging the Scientific Data Format Chasm

D. G. Jennings[1] and W. D. Pence

NASA Goddard Space Flight Center, Greenbelt MD 20771

M. Folk

National Center for Supercomputing Applications, Champaign IL 61820

Abstract. The *Convert* project is a newly funded NASA endeavor whose goal is to provide the scientific community with software tools for data format conversion. Among its goals are the development of utilities that transform data in non-standard formats into selected standard formats and allow for the inter-convertibility of data between those same standard formats. This paper discusses several aspects of the *Convert* project including current software tools, planned software tools, standard data formats under consideration and the efforts to devise transformation mappings between these formats.

1. Introduction

There currently exists within the scientific community a plethora of formats used to store and analyze data. To help unify the data into understandable and transportable formats, we have undertaken an effort known as the *Convert* project. The goals of *Convert* are twofold. First, *Convert* will provide tools to convert data in non-standard data formats into selected standard data formats. These tools will initially perform conversion of data into FITS (Wells et al. 1981), although plans to accommodate HDF (NCSA 1993a) conversion are under consideration. Secondly, *Convert* will allow for the inter-convertibility of data between FITS, HDF, and eventually netCDF (Unidata 1991) formats.

The first goal of *Convert* benefits astrophysical science since it will, for the first time, provide the astronomical community with a set of general purpose tools to transform instrument and mission-specific data products into FITS format. Data from many older astronomy missions (e.g., *EXOSAT*, *SAS-2*, *Vela*) does not reside in FITS and therefore requires post-mission FITS conversion in order to remain useful to future analysis efforts. Even data from some current astrophysics missions (e.g., *GRO*, *ULYSSES*) requires special processing to convert it from instrument specific formats into FITS.

The second goal of *Convert* should allow astronomers to convert their FITS data into formats used by other disciplines. Besides being useful for collabo-

[1]Employed by Hughes STX Corporation

rative efforts between astronomers and scientists from non-astronomical fields, this ability will make it possible for astronomers to utilize the software tools (e.g., data visualization, data management, data analysis) being developed in other fields. As funding sources for science, especially astronomy, continue to shrink, the sharing of resources between disciplines becomes both attractive and necessary.

The following sections elaborate on the work underway by the *Convert* project participants. Section two gives background on the software tools that form the basis of the *Convert* software package, as well as the planned enhancements to these tools. Section three describes the invertible transformation being developed between FITS and HDF. Finally, section four provides a summary of *Convert* project activity and supplies reasons why this activity should be of interest to the astronomical community.

2. Software Tools

Convert shall make use of three pre-existing, public-domain software packages: FITSIO (Pence 1994), ToFU (Jennings 1993) and the HDF application programming interface library (NCSA 1993b). The FITSIO subroutine library, in wide use in NASA-sponsored astrophysics missions, provides an easy-to-use and reliable, low-level I/O programming interface to FITS files. Layered on top of FITSIO is the ToFU (To FITS Utilities) subroutine library. ToFU consists of routines that facilitate the conversion of data into FITS format. Lastly, the HDF application programming interface (API) consists of both high- and low-level routines arranged into six separate modules, with each module corresponding to a supported HDF data object.

Updated and modified versions of ToFU and the HDF API will constitute build 1 of the *Convert* software package. Work is currently underway (and should be complete by the printing of this paper) on porting the ToFU library to C and greatly enhancing its usability. Once complete, ToFU will provide *Convert* with its advertised FITS conversion capabilities. We also intend to integrate the FITSIO subroutine library into the HDF application interface. Thus, the HDF library will be able to use FITSIO to read and write FITS files, just as it currently does with HDF and netCDF files. The integration of FITSIO into the HDF API forms the basis of *Convert's* data format inter-convertibility capabilities, since the HDF API will then be able to read in FITS, HDF, or netCDF files and write out HDF or FITS (or eventually netCDF) files.

An interesting consequence of FITSIO incorporation into the HDF API is that FITS and HDF will then share a common software interface. In this sense, there will be no operational difference between the two formats. Applications may input or output FITS (HDF) formatted files as easily as they input or output HDF (FITS) formatted files. Additionally, the FITSIO/HDF API merger provides FITS with a standard and widely distributed application interface of its own: something that the FITS community has long needed.

3. Inter-Convertibility between FITS and HDF

Before the HDF API can make use of the FITSIO library to read and write FITS formatted files, there must be an invertible transformation defined between FITS and HDF data objects. Without such a transformation, HDF interfaces will not understand how to interpret the contents of FITS primary arrays, images, ASCII or binary tables. Even though FITS and HDF implement their data structures differently, they both make use of the same basic set of abstract data types. (Note, however, that netCDF and CDF use the concept of named N-dimensional variables, which does not conform well to either the HDF or FITS models. This makes the inter-convertibility between FITS and CDF/netCDF a more complex problem than the inter-convertibility between FITS and HDF. Thus, the issue of FITS to CDF/netCDF mappings will be left for future work). The mapping between FITS and HDF is, with a few notable exceptions, a relatively straightforward exercise. The following list demonstrates preliminary mappings between FITS and HDF data objects:

- FITS Primary Array and Image Extensions –> HDF SDS Objects
- FITS ASCII and Binary Table Extensions –> HDF Vdata Objects
- HDF 8-bit and 24-bit RIS Objects –> FITS Image Extensions
- HDF Palettes and SDS Objects –> FITS Image Extensions
- HDF Annotation Objects –> FITS ASCII Table Extensions
- HDF Vdata Objects –> FITS Binary Table Extensions

For a more detailed comparison of the FITS and HDF formats, see Jennings & McGlynn (1993).

The only unresolved issues in the FITS to HDF mapping are the lack of support within HDF for certain FITS data types (i.e., 4-byte complex, 8-byte complex, bit, boolean) and the need for field-associated attributes (i.e., keyword = keyvalue pairs) in HDF Vdata objects. The HDF development group at NCSA intends to solve both problems by modifying the HDF data format to accommodate new data types and Vdata field attributes.

The most significant unresolved issue in the HDF to FITS mapping is the lack of a robust FITS hierarchical grouping structure. HDF often associates its data objects into hierarchical groupings and then operates on the groups as if they were a single entity. Therefore, the FITS to HDF mapping requires that FITS adopt an HDF-like hierarchical grouping structure so that FITS may preserve HDF data object associations during HDF to FITS transformations. To this end, the authors have begun work on a proposal to augment FITS with its own hierarchical grouping convention. Even though the HDF to FITS mapping motivates this effort, the authors will endeavor to develop a grouping structure that is general enough for all FITS applications to use.

4. Summary

It could be argued that the field of astronomy focuses most of its efforts on internal concerns, neglecting to stay abreast of work being done in other disciplines. This attitude is ultimately self-limiting, because in terms of money and resources astronomy is just one small part of the world scientific community. The earth science, space science, and atmospheric science disciplines invest significant portions of their (often greater) resources into developing software to visualize, manage, and analyze data. If astronomers wish to leverage their own resources and take advantage of the infrastructure built by other disciplines, then astronomical data must be convertible to the formats used by those infrastructures.

Even within astronomy, research teams do not always concern themselves with producing data products in standard formats. This creates situations in which data becomes increasingly less understandable over time and more difficult to analyze with standard analysis packages. To ensure that the astronomical community has a cost effective means to transform data into standard formats and between standard formats, the *Convert* project has begun work on two software tool sets. The first, ToFU, converts generic data streams into FITS format. Secondly, the HDF application programming interface will allow the inter-convertibility of FITS and HDF formatted data sets. Both software packages will make use of the FITSIO subroutine library to perform low level FITS file manipulations.

Before true inter-convertibility between FITS and HDF can be achieved, an invertible mapping between the two formats is needed. This mapping is reasonably straightforward because of the high level of abstract conformability between FITS and HDF. Only minor changes need occur to the FITS and HDF formats to make the mapping possible. Amongst these changes are additions of new data types to HDF, support for field attributes in HDF Vdata objects, and the creation of a hierarchical grouping convention for FITS. The work necessary to modify both data formats is underway.

Acknowledgments. We gratefully acknowledge the support of the NASA Applied Information Systems Research Program (AISRP), under which this effort is funded.

References

Jennings, D. J., et al. 1993, BAAS, 25, 962

Jennings, D. J., & McGlynn, T. A. 1993, in Astronomical Data Analysis Software and Systems III, ASP Conf. Ser., Vol. 61, eds. D. R. Crabtree, R. J. Hanisch, & J. Barnes (San Francisco, ASP), p. 526jenningsd

NCSA 1993a, NCSA HDF Specifications and Developers Guide, Version 3.2 (Champaign, NCSA)

NCSA 1993b, Getting Started with HDF, Version 3.2 (Champaign, NCSA)

Pence, W. D. 1994, this volume, p. **??**

Unidata 1991, NetCDF User's Guide: An Interface for Data Access (Boulder, Unidata Program Center)

Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363