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.

A Proposed Convention for Writing FITS Data Tapes: DRAFT 0

ROSAT/ASCA/XTE Development Team

Astrophysics Data Facility, NASA Goddard Space Flight Center, Code 631, Greenbelt, Md. 20771

Abstract. Even with today's advances in networking, file system capacities and CD technology it is often necessary to transport and store scientific data sets on magnetic tape. The FITS data format standard contains guidelines on how to write FITS files to magnetic tape but does not address the problem of indexing or organizing tape files. Currently available magnetic tape media can store multiple gigabytes of information on a single tape, which translates into thousands of FITS files per tape. Thus, the lack of a standard tape indexing and organizing scheme can, in many instances, become a serious problem.

Faced with the above dilemma, the Astrophysics Data Facility at Goddard Space Flight Center has developed a simple in-house convention for indexing the contents of FITS data tapes that allows software to quickly and easily inventory tape contents. This paper describes the convention used by our organization. We propose that this convention be adopted into the FITS standard as the way to index and organize the contents of magnetic tape media.

1. Introduction

Originally defined as an exchange format for data on nine-track tape (NOST, 1993), the Flexible Image Transport System (Wells et al. 1981) has in recent years expanded into a very general and useful logical data format. The current FITS format accommodates data exchange and archival storage on a wide range of media, and is also used as a working native data format (Pence et al. 1992) for many new software applications.

While the main emphasis of FITS has moved away from magnetic tape-based data storage and transport there are, perhaps surprisingly, new and growing needs for its use as a data tape format. Data volumes from many current and planned astronomy missions preclude the use of electronic data distribution. Examples of such missions are the *ASCA* X-ray Observatory Satellite which produces $\sim$ 500 MB (200–1000 files) of FITS formatted data per observation, and the soon to be launched *XTE* (X-ray Timing Explorer) satellite with an estimated data size of $\sim$3 GB (3000–20000 files) files per observation. Storing and distributing data on CD-ROM media can be a good alternative to electronic distribution in some cases, however, this technology costs an order of magnitude more to use than magnetic tape media and the storage capacity is many times less (600 MB per CD-ROM vs. 8–16 GB per 4mm tape).

Storing and transporting data sets on magnetic tape solves the problem of large data volume, but it does not provide FITS readers with the information needed to construct a catalog of the tape contents (short of reading every file from the tape). Keeping tape content information (e.g., file names, file sizes, file order) tends to be less of a problem when the number of FITS files per tape remains small, since this information can be easily stored external to the tape. However, when the numbers of files on a single tape grow into the hundreds or thousands, it becomes desirable for the tape to be *self-describing* just as a single FITS file is considered self-describing.

The FITS standard provides simple guidelines on how to write FITS formatted data to magnetic tape (Grosbøl & Wells 1994), but it does not address the issue of tape content indexing and cataloging. This paper presents a convention for writing self-describing magnetic tapes that contain FITS formatted data files. By using this convention, FITS readers may quickly access a catalog of the tape contents and determine the names, sizes, positions and meanings of every data file contained on a tape.

2. Tape Structure

In addition to the recommended guidelines for tape block sizes (up to 28800 bytes per block in increments of 2880), file separators (single tape marks) and logical tape labels (ANSI standard labels or no labels) this convention requires that the first file written to every FITS formatted data tape be a catalog of the tape contents. The catalog file is itself a FITS formatted file with a null primary array and an ASCII table (Harten et al. 1988) as its first extension. The second and subsequent tape files may be written in any order as long as this corresponds to the recorded order in the tape catalog file.

3. Tape Catalog File

All information pertaining to tape content resides in an ASCII table, which must be the first extension of the tape catalog file. Other FITS extensions may follow this ASCII table but the primary array shall be empty (null). This ensures that a dump of the tape catalog file produces readable output, at least up to the end of the first extension.

The ASCII table containing tape content information is composed of four table columns and one row for each FITS file on the tape. The four column entries provide the following information about each FITS file on tape: (1) original file name, (2) file size, (3) a brief description of file contents, and (4) the file's tape position number (first file on tape = tape catalog file = 1).

The order of the columns within the extension is not important and any additional columns describing tape contents are allowed. However, the four required columns must have associated TTYPE keywords with the following values:

- TTYPEnnn= 'filename', TTYPE keyword value for column holding original names of files on tape.
- TTYPEnnn='filesize', TTYPE keyword value for column containing sizes of files on tape. Recommended file size unit is bytes, kilobytes or megabytes.
- TTYPEnnn='descrip', TTYPE keyword value for column describing contents of files on tape.
- TTYPEnnn='filenum', TTYPE keyword value for column holding positional values of files on tape.

4. Example Tape Catalog File

The following is an example of a tape catalog file currently being used for guest observer distribution tapes for the ASCA X-ray Observatory.

```
SIMPLE  =           T / file does conform to FITS standard
BITPIX  =          16 / number of bits per data pixel
NAXIS   =           0 / number of data axes
EXTEND  =           T / FITS dataset may contain extensions
FNAME   = 'ad13000000_050_tape.cat' / Original file name
SEQNUM  =    13000000 / Sequential number from ODB
PROCVER = 'P4.0.0  '  / Processing Configuration number
SEQPNUM =         050 / Number of times sequence processed
USPINUM =         5000
END

XTENSION= 'TABLE   '  / ASCII table extension
BITPIX  =           8 / 8-bit ASCII characters
NAXIS   =           2 / 2-dimensional ASCII table
NAXIS1  =         135 / width of table in characters
NAXIS2  =         201 / number of rows in table
PCOUNT  =           0 / no group parameters (required)
GCOUNT  =           1 / one data group (required)
TFIELDS =          4  / number of fields in each row
TTYPE1  = 'filenum '  / label for field   1
TBCOL1  =           1 / beginning column of field   1
TFORM1  = 'I4      '  / Fortran-77 format of field
TTYPE2  = 'filename'  / label for field   2
TBCOL2  =           6 / beginning column of field   2
TFORM2  = 'A57     '  / Fortran-77 format of field
TTYPE3  = 'filesize'  / label for field   3
TBCOL3  =          64 / beginning column of field   3
TFORM3  = 'I7      '  / Fortran-77 format of field
TUNIT3  = 'kilobytes' / physical unit of field
TTYPE4  = 'descrip '  / label for field   4
TBCOL4  =          72 / beginning column of field   4
TFORM4  = 'A64     '  / Fortran-77 format of field
```

```
HISTORY   This FITS file was created by the FCREATE task.
SEQNUM  =     13000000 / Sequential number from ODB
END
```

5. Summary

The above simple scheme for producing self-describing FITS data tapes allows both humans and software to understand the contents of a tape without unloading and examining every file. The true utility of this feature can be realized when one considers data tapes containing gigabytes of data and hundreds (or thousands) of individual FITS files.

If adopted as a standard FITS convention, this method of writing data tapes will allow users to pseudo-randomly access files or groups of files from a tape, knowing in advance the disk space necessary to hold them. It will also ensure that the contents of a tape is documented, since the catalog becomes part of the data set and FITS readers will always know where to find the information.

References

NASA Office of Standards and Technology 1993, Definition of the Flexible Image Transport System (FITS) (Greenbelt, NASA/OSSA)

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

Pence W., Blackburn J. K., & Greene E. 1992, 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. 541

Grosbøl, P., & Wells D. 1994, Blocking of Fixed-block Sequential Media (Greenbelt, NOST Office GSFC), available via anonymous ftp at `nssdca.gsfc.nasa.gov`

Harten, R. H., Grosbøl, P., Greisen, E. W., & Wells, D. C. 1988, A&AS, 73, 365