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

The Chandra Data Archive

A. Rots
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

Abstract. We describe the hierarchy and organization of the data products in the archive of the Chandra X-Ray Observatory. The CDA has adopted the set of FITS conventions developed by the HEASARC, extending them where necessary. This has resulted in a more comprehensive set of conventions on which future missions can build. It is implemented in the Chandra Data Archive environment through a set of standard FITS header modules prescribed for all data products. All CXCDS software is attuned to these standard components but will attempt to perform reasonably for data products from other missions. A brief mention is made of data distribution and time convention issues.

1. The Functional Design of the Chandra Data Archive

The Chandra Data Archive (CDA) is depository as well as source of information for all Chandra-related activities (Proposal submission and management, Mission Planning, Automated Processing, Data Retrieval). The observation catalog (ObsCat), embodied in the database *axafocat* plays a central role in this.

axafocat, with the help of related databases:

- Receives information from approved proposals,
- Provides that information to Mission Planning,
- Receives scheduling information from Mission Planning,
- Provides information on scheduling, target data, and instrument configurations to Automated Processing,
- Receives updates to that information from Automated Processing,
- Informs the user what observations are being planned and what data are available,
- And aids in retrieving those data products.

For more information on the interfaces, see Estes et al. (2000).

2. The Data Products in the CDA

The CXC Automated Processing system generates data products in three main levels (Levels 0, 1, 2) and two sublevels (Levels 0.5, 1.5) of processing. All products are archived and divided into three categories: primary, secondary, and supporting products. The primary products contain all (science) Level-2 products and some crucial science Level-1 products. The secondary products contain the remaining science Level-1 products, Level-1 aspect products, and Level-1 orbit ephemeris. The remaining products are classified as supporting. Primary and secondary products are shipped to our users automatically. Individual supporting products are available upon request.

The basic data product (file) retrieval keys for observational data are:

- Time (or observation),
- Data product type (CONTENT).

(As far as retrieval by the user is concerned, the immediately visible keys are the usual observation-related ones, such as source, type of object, time, observing mode, RA, Dec, etc. (which largely translate into time), and standard data product packages.)

The vast majority of the data products that the user will encounter are FITS files; for these the metadata are extracted from the headers. In addition, the archive holds raw binary files, with accompanying log files for the metadata, and is equipped to deal with other standard formats, such as PostScript, PDF, GIF, JPEG, etc.

A FITS data product's type is identified by its CONTENT keyword. The list of recognized CONTENT values has, in principle, a flat structure, but for most querying activities a hierarchy is imposed by grouping them into instrument/subsystem and level-of-processing categories.

3. FITS Conventions

Starting with the HEASARC FITS Working Group guidelines and conventions, we have developed an extended set of conventions to which all FITS files produced by the CXC are supposed to adhere.

FITS files consist of one or more Header-Data Units (HDU), each containing a header and a data part. For all practical purposes, an HDU may be one of the following:

- Primary array,
- Image extension,
- Binary table extension.

The first HDU in a FITS file must be a primary array; all subsequent ones (if any) are extensions. The length of the header as well as the data part of each HDU must be an integer (positive for headers, non-negative for data parts) multiple of 2880 bytes.

All FITS files are to be checksummed: the 32-bit one's complement checksum over each HDU (and therefore over the entire file) must be -0 (all ones). Hence, first order data product integrity may be ascertained by verifying the checksum.

The FITS headers are required to have a number of standard components:

- **M:** Mandatory FITS keywords for HDU type,
- **CC:** Configuration control component,
- **T:** Timing component,
- **O:** Observation info component,
- **IC:** Image coordinate system keywords (CTYPEn, etc.),
- **TC:** Table column specifications (TTYPE, etc.), coordinate system, and ranges (TLMIN/TLMAX).

There are different specifications for these components, depending on HDU type:

Table 1.

Type	FITS Primary or extension	ASC Principal or auxiliary	Header components
Image	Primary	Principal	M, CC, T, O, IC
Null	Primary	Auxiliary	M, Null CC, Short T, Short O
Table	Extension	Principal	M, CC, T, O, TC
Table	Extension	Auxiliary	M, Short CC, Short T, Short O, TC
Image	Extension	Auxiliary	M, Short CC, Short T, Short O, IC

4. Data Distribution

Although we will continue the practice of sending proprietary data out to Guest Observers on physical media (4-mm DAT tape, 8mm Exabyte tape, CD-ROM), we expect that the bulk of our data transfers will happen through our web interface, for public as well as proprietary data.

The CDA Search and Retrieval interface is discussed by Estes et al. (2000). It will enable the community to download any set of CDA products, public or proprietary (provided one has the proper authorization), or stage the data for anonymous ftp transfer.

In the meantime, we have implemented a provisional web interface that allows retrieval of primary and secondary products for public observations that have been approved by the CXC Verification & Validation process.

5. Time and Ephemerides

Time is kept in FITS files as MET: seconds since 1998-01-01T00:00:00 TT. It is tied to absolute time by the keyword MJDREF (50814.0). At that epoch, the difference with UTC was: TT-UTC = 63.184 s; by launch: TT-UTC = 64.184 s.

We have adopted the FITS formats developed in the context of the RXTE mission for orbit ephemeris files and the JPL DE-200 and DE-405 solar and planetary ephemerides. The latter come with an ANSI C interface, using HEASARC's cfitsio, that is functionally equivalent to the JPL Fortran software. Consequently, the RXTE multi-mission software package for time conversions, barycenter corrections, and pulsar phase binning (including absolute timing and phase-resolved spectroscopy) can be used for Chandra data after only slight modifications.

References

Estes, A. N., DuPrie, K., Harbo, P., Pak, K., & Zografou, P. 2000, this volume, 457