The "Data Model" and the Data Manipulation Tools

More info and examples can be found in:

***Data Manipulation Users Guide** (not completely updated)*:
http://asc.harvard.edu/ciao/download/doc/manual_dm.ps

***Introduction To the Data Model** thread:*
http://asc.harvard.edu/ciao/threads/dm_intro.thread.html

Also by doing "ahelp" on the following subjects:

dm, dmintro, dmsyntax, filtering, binning, dmimages, region, dmcols, dmimfiltering

- The CXC analysis and processing software is built on a common versatile interface library called the CXC Data Model (or just DM).

- The DM provides users with a powerful built-in data filtering and binning capability.

- The name “Data Model” reflects the fact that the interface can be used on data files of different format (all described by a single abstract description - the same “model”) in a transparent way.

- As of CIAO2.1, the format supported by the DM library include: FITS, IRAF QPOE and IRAF IMH.

- An important characteristic of the DM is that ANY program that asks for a data file name as input, will accept a virtual file string which will cause the program to see a filtered version of the file in question.

- The "virtual file" syntax is also commonly used to create on disk a filtered version of the input file.

- Another important characteristic of the DM is that all columns of event lists are treated "equally": for example binning is allowed not only in spatial coordinates but also in e.g. time, or energy coordinate, giving the ability of creating multidimensional images in space-energy, or space-time, etc.

SUMMARY OF MAIN DATA MODEL PROPERTIES:

FORMAT INDEPENDENT

POWERFUL FILTERING AND BINNING

USAGE OF “VIRTUAL FILES”

Data Manipulation Tools

The four DM "core" tools are:

dmlist: list contents or structure of a file

dmcopy: filter and bin tables and images

dmextract: make a histogram table file (e.g. PHA file) from a table column. Generate count histogram on supplied regions for a spatial table or image file.

dmgti: create custom Good Time Intervals (GTIs) from a constraint expression

The full list of DM tools in CIAO 2.1 includes also:

dmappend - Append multiple blocks/extensions to an existing output file
dmarfadd - Add multiple ARF files together , weighting by exposure time
dmcontour - Make contour regions from a 2-D image
dmcoords - Convert between Chandra instrumental coordinate systems
dmgroup - Group a specified column in a table
dmhedit - Edit data model file headers
dmimg2jpg - Make a color JPEG image from three image files
dmimgcalc - Perform arithmetic on images
dmimghist - Make histogram of values in a 2-D image
dmkeypar - Retrieve information about a keyword from an input file
dmmakepar - Write header keywords to a parameter file
dmmakereg - Create a FITS region file from an ASCII region description
dmmerge - Merge two or more compatible tables into one
dmpaste - Add new columns to a table
dmreadpar - Add parameters from a .par file to a file header
dmrega2fts - Convert a grating ASCII region into a FITS format
dmregrid - Rebin a stack of 2 dimensional images
dmsort - Sort a table block on a given column
dmstat - Compute standard statistics for the column in a table or image
dmtcalc - Define new table columns as functions of old ones
dmtype2split - Create a type 1 file for specified rows of a type 2 file
dmwritefef - Create a FITS Embedded Function (FEF) file from ASCII files

DATA MODEL SYNTAX

- All CIAO tools use the DM library and therefore accept as input "virtual files" described using the DM syntax.

- In the DM context a "virtual file" is represented by a filename followed by a series of optional qualifiers in square brackets [ ]:

filename[block][filter][columns/binning][options]

where:

block - is the "section" of the file to use

filter - is the filter to be applied

columns/binning - specifies either the columns from a table to be included in an output table or the binning. When binning the data to generate an n-dimensional image, the range and binsize (min:max:bin) must be specified.

options - a sequence describing special options for the DM library

Note that:

- the order of the qualifiers matters, however ...
- not all qualifiers need to be present always

Examples of "virtual files":

- Select the first three columns of the EVENTS block by number:

acisf01843N001_evt2.fits[EVENTS][time=84245787:84247000][cols #1,#2,#3]

or by name:

acisf01843N001_evt2.fits[EVENTS][grade=0,2,3][cols time,ccd_id,node_id]

after filtering in time or grade

- Bin an events file to create a PI spectrum for a specified region (input of dmextract):

acisf01843N001_evt2.fits[EVENTS][sky=region(mysrc.reg)][bin pi=1:1024:1]

or an image (input of dmcopy):

acisf01843N001_evt2.fits[EVENTS][pha<100][bin x=320:480:4,y=320:480:4]

In the examples above:

block: [EVENTS]

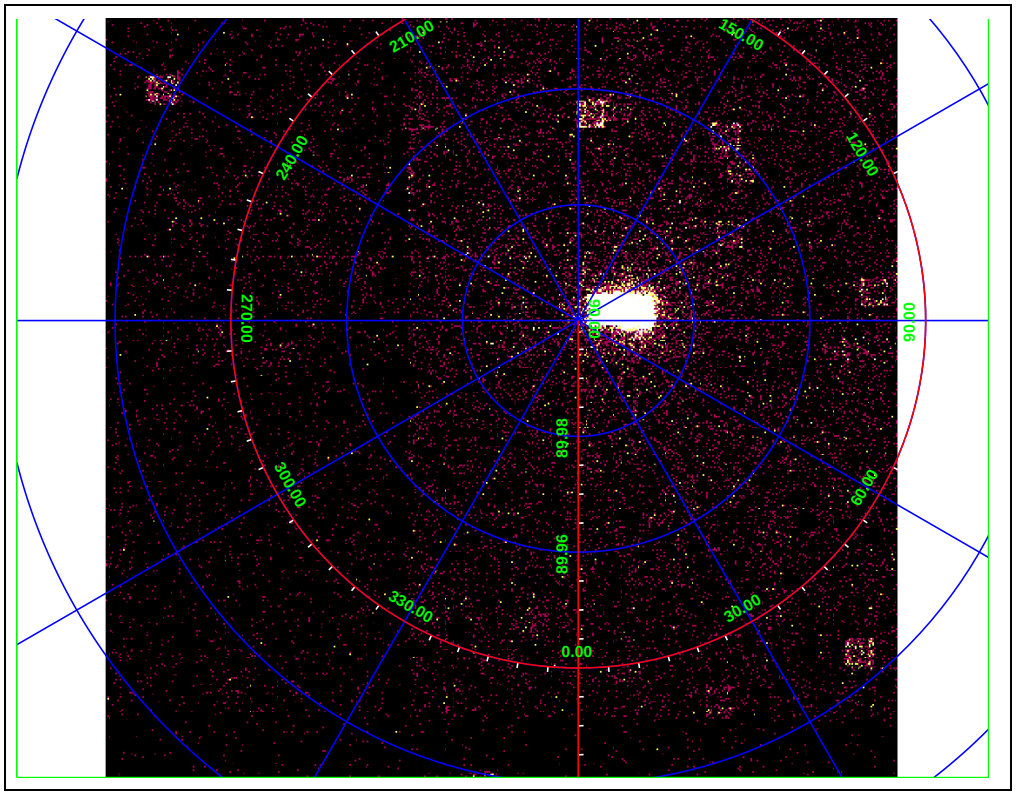
filter: [time=84245787:84247000]
[grade=0,2,3]
[sky=region(mysource.reg)]
[pha<1000]

columns/binning:
[cols time,ccd_id,node_id]
[cols #1,#2,#3]
[bin pi=1:1024:1]
[bin x=320:480:4,y=320:480:4]

DM Examples 1: Detector Image

Imaging on multiple coordinate systems: first, let's look at a region in detector coordinates, filtered on energy and time.

```
dmcopy "merge3e.fits[energy=500:2000,
    time=:63940080,63940180:][bin detx=3500:4500:2,
                              dety=3500:4500:2]" det.img
```



DM Examples 2: Sky Image

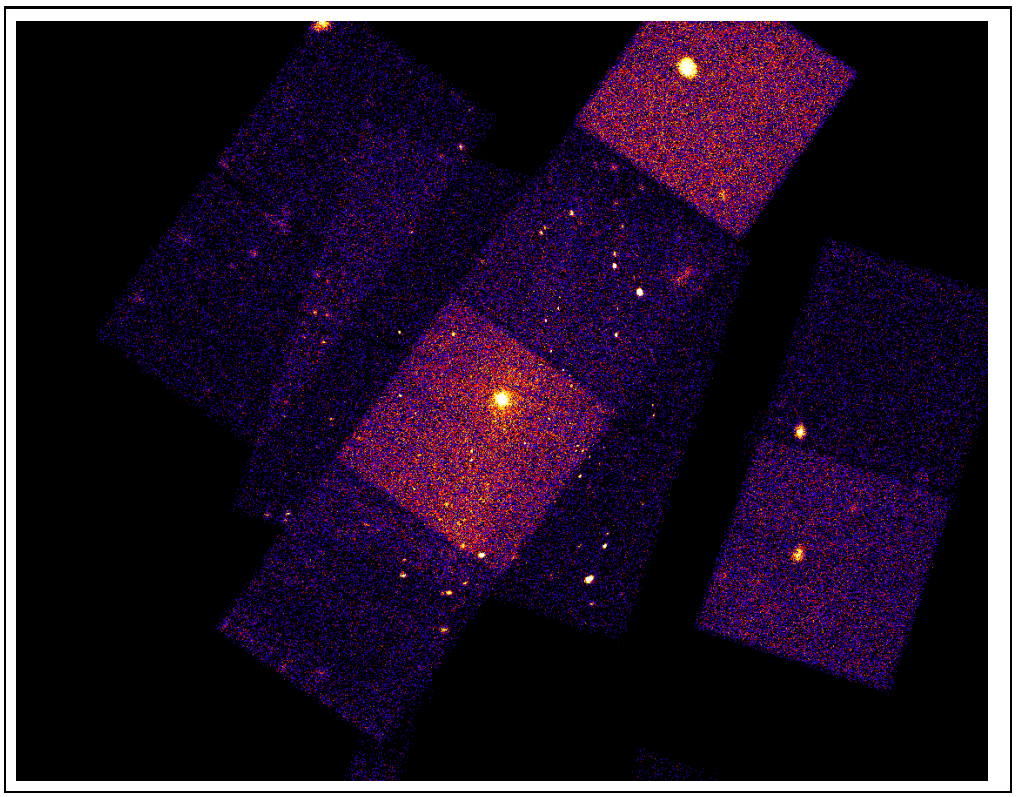
Now look at the same photons but in sky coordinates

```
dmcopy "merge3e.fits[energy=500:2000,time=:63940080,63940180:,
        detx=3500:4500,dety=3500:4500]
       [bin x=3200:4800:2,y=3200:4800:2]" sky.img
```

DM Examples 3: Merged sky image

The whole field was created by merging three separate observations.

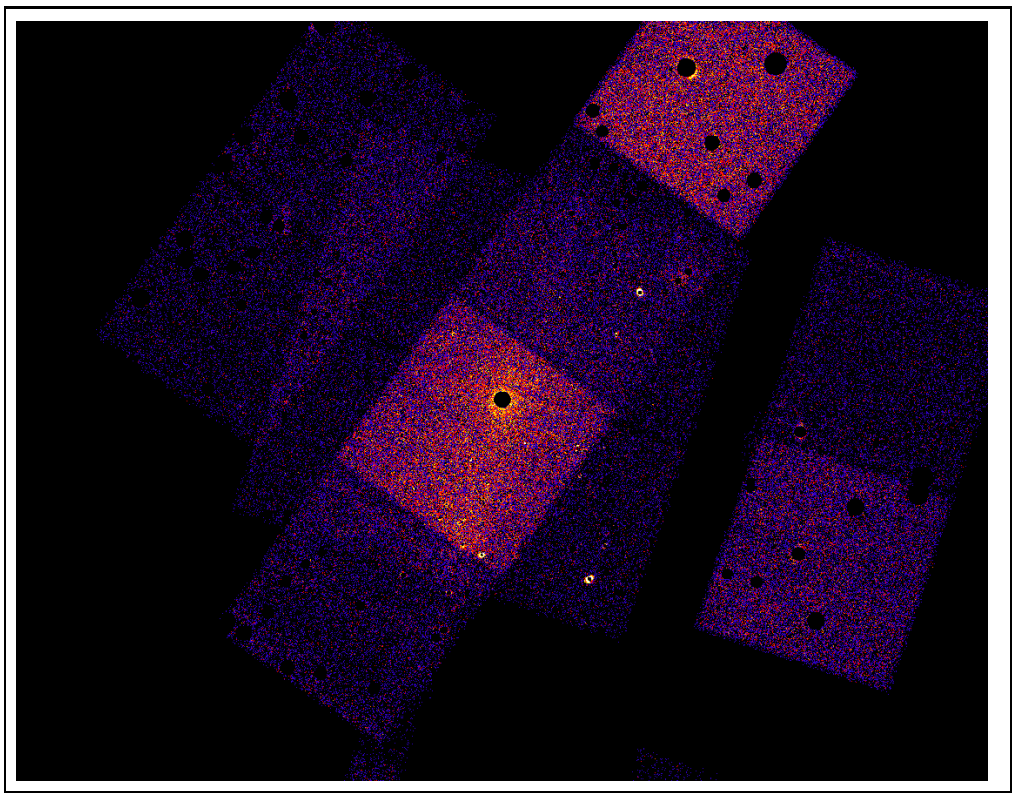
```
dmmerge "786.fits,787.fits,1730.fits" outfile=merge3e.fits
```



DM Examples 4: Removing sources

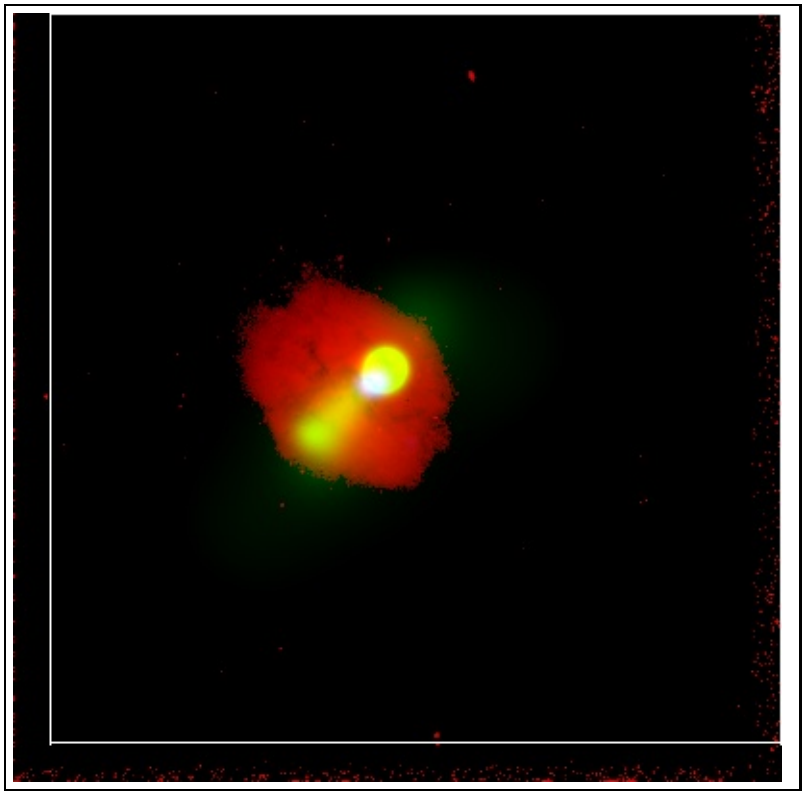
We can generate a background image by removing sources found by the automatic source detection program.

```
dmcopy "merge3e.fits[exclude sky=region(gg.reg)]" exclude.fits
```



- The dmimg2jpg tool is useful for comparing images in different energy bands

```
dmimg2jpg dss.fits "merge3e.fits[energy=400:1500]"
  "merge3e.fits[energy=1500:6500]" ha2.jpg clob+
    maxred=3 maxblue=15
```



CIAO 2.1 Science Threads

Introduction To the Data Model (4 April 2001)

Return to Threads Page.

For illustration, this thread utilizes the ObsID 1843 (ACIS-I, G21.5-0.9) data that was downloaded in the How To Download Chandra Data From The Archive thread.

Contents:

The CIAO "data model" is a versatile interface that allows any standard format datafile (FITS, QPOE, IMH) to be used in a transparent way. This interface also allows powerful filtering of datafiles. The filtered output can either be directly inputed to an analysis task (as a "virtual file") or be written to disk.

A brief introduction to the data model interface and examples on how to use the four core tools are provided here. Full details on all the dmtools can be found in the *Data Manipulation Users Guide*.

Data Model Tools

The four core tools, those that are used most often, are:

- **dmlist** - List contents or structure of a file
- **dmcopy** - Filter and bin tables and images
- **dmextract** - Make a histogram table file (e.g. PHA or PI file) for spectral fitting

- **dmgti** - Define custom Good Time Intervals (GTIs)

In addition, there are a number of more specialized tools:

- **dmappend** - Append multiple blocks/extensions to an existing output file
- **dmarfadd** - Add multiple ARF files together, weighting by exposure time
- **dmcontour** - Make contour regions from a 2-D image
- **dmcoords** - Convert between Chandra instrumental coordinate systems
- **dmgroup** - Group a specified column in a table
- **dmhedit** - Edit data model file headers
- **dmimg2jpg** - Make a color JPEG image from three image files
- **dmimgcalc** - Perform arithmetic on images
- **dmimghist** - Make histogram of values in a 2-D image
- **dmkeypar** - Retrieve information about a keyword from an input file
- **dmmakepar** - Write header keywords to a parameter file
- **dmmakereg** - Create a FITS region file from an ASCII region description
- **dmmerge** - Merge two or more compatible tables into one
- **dmpaste** - Add new columns to a table
- **dmreadpar** - Add parameters from a `.par` file to a file header
- **dmrega2fits** - Convert a grating ASCII region into a FITS format
- **dmregrid** - Rebin a stack of 2 dimensional images
- **dmsort** - Sort a table block on a given column
- **dmstat** - Compute standard statistics for the column in a table or image
- **dmtcalc** - Define new table columns as functions of old ones
- **dmtype2split** - Create a type 1 file for specified rows of a type 2 file
- **dmwritefef** - Create a FITS Embedded Function (FEF) file from ASCII files

Running Data Model Tools

- Running from the Command Line

  The dmtools are usually executed from the command-line prompt:

  ```
  unix% dmcopy "acisf01843N001_evt2.fits[EVENTS]" acis_events.fits
  ```

- Running from a GUI

  All of the capabilities of the data model tools are available from the "Analysis" menu of both *Prism* and *FilterWindow*. See the introductory threads for those tools for more information.

Syntax for Defining a Subspace

When examining large, multi-dimensional files, it is useful to list a subset of interest; this is known as defining a subspace. The basic syntax of a virtual file has four arguments:

```
filename[block][filter][columns/binning][newblock]
```

- `block` - the section of the file to use

- `filter` - filter to be applied
- `columns/binning` - specifies either the columns from a table to be included in an output table or the binning. When binning the data to generate an n-dimensional image, the range and binsize (min:max:bin) must be specified.
- `newblock` (optional) - the name for the new block in the output file, default is the block used from the input file

Examples of a virtual file:

- Select the first three columns of the EVENTS block by number:

  ```
  acisf01843N001_evt2.fits[EVENTS][cols #1,#2,#3]
  ```

  or by name:

  ```
  acisf01843N001_evt2.fits[EVENTS][cols time,ccd_id,node_id]
  ```

- Bin an events file to create a PI spectrum (using `dmextract`):

  ```
  acisf01843N001_evt2.fits[EVENTS][bin pi=1:1024:1]
  ```

  or an image (using `dmcopy`):

  ```
  acisf01843N001_evt2.fits[EVENTS][bin x=3200:4800:4,y=3200:4800:4]
  ```

Using dmlist

FITS files generally contain several different blocks, each containing multi-dimensional data. The `dmlist` tool allows the user to inspect all or part of a data file by column, row, etc.

- To list the blocks in a file:

  ```
  unix%  dmlist acisf01843N001_evt2.fits blocks

  ------------------------------------------------------------------------------
  Dataset: acisf01843N001_evt2.fits
  ------------------------------------------------------------------------------

       Block Name                          Type         Dimensions
  ------------------------------------------------------------------------------
  Block    1: PRIMARY                        Null
  Block    2: EVENTS                         Table        14 cols x 475869   rows
  Block    3: GTI7                           Table         2 cols x 1        rows
  Block    4: GTI0                           Table         2 cols x 1        rows
  Block    5: GTI1                           Table         2 cols x 1        rows
  Block    6: GTI2                           Table         2 cols x 1        rows
  Block    7: GTI3                           Table         2 cols x 2        rows
  Block    8: GTI6                           Table         2 cols x 1        rows
  ```

- To list the columns of the events block in a file (note the required double quotes):

  ```
  unix%  dmlist "acisf01843N001_evt2.fits[events]" cols

  ------------------------------------------------------------------------------
  ```

```
Columns for Table Block EVENTS
-------------------------------------------------------------------------------

ColNo  Name                 Unit        Type             Range
   1   time                 s            Real8            84270902.16820:   842
   2   ccd_id                            Int2           0:9                  CCD
   3   node_id                           Int2           0:3                  CCD
   4   expno                             Int4           0:2147483647         Exp
   5   chip(chipx,chipy)    pixel        Int2           1:1024               Chi
   6   tdet(tdetx,tdety)    pixel        Int2           1:8192               ACI
   7   det(detx,dety)       pixel        Real4              0.50:     8192.50 AC
   8   sky(x,y)             pixel        Real4              0.50:     8192.50 sk
   9   pha                  adu          Int4           0:36855              tot
  10   energy               eV           Real4          0:    1000000.0      nor
  11   pi                   chan         Int4           1:1024               pul
  12   fltgrade                          Int2           0:255                eve
  13   grade                             Int2           0:7                  bin
  14   status[4]                         Bit(4)                              eve

-------------------------------------------------------------------------------
4:
-------------------------------------------------------------------------------

ColNo    Name
5:    CPC(CPCX) = (+0) +(+0.0240)* (chip(chipx)-(+0.50))
         (CPCY)   (+0)  (+0.0240)  (    (chipy) (+0.50))
7:    MSC(THETA) = (+0) +TAN-P[(+0.000136667)* (det(detx)-(+4096.50))]
         (PHI  )   (+0)        (+0.000136667)  (   (dety) (+4096.50))
8:    EQPOS(RA ) = (+278.0474) +TAN[(-0.000136667)* (sky(x)-(+4096.50))]
           (DEC)   (-10.5711 )      (+0.000136667)  (   (y) (+4096.50))
```

- To list the first 5 events in a file:

```
unix%  dmlist "acisf01843N001_evt2.fits[events]" data row=1:5

-------------------------------------------------------------------------------
Data for Table Block EVENTS
-------------------------------------------------------------------------------

ROW    time                 ccd_id node_id expno      chip(chipx,chipy) tdet(tde

     1    84272488.55042922    6    3          3    (861,148)  (3736,1850) (
     2    84272488.55042922    6    3          3    (962,609)  (3837,2311) (
     3    84272488.59146923    7    2          3     (524,10)  (4441,1712) (
     4    84272488.59146923    7    3          3     (807,52)  (4724,1754) (
     5    84272488.59146923    7    1          3    (448,135)  (4365,1837) (
```

- To list the file header:

```
unix% dmlist acisf01843N001_evt2.fits header

-------------------------------------------------------------------------------
Header keys for block EVENTS
-------------------------------------------------------------------------------

 --  COMMENT              This FITS file may contain long string keyword value
 --  COMMENT              continued over multiple keywords.  The HEASARC conve
 --  COMMENT              character at the end of each substring which is then
 --  COMMENT              on the next keyword which has the name CONTINUE.
0001 HDUCLASS            OGIP                          String
```

```
0002 HDUCLAS1            EVENTS                          String
0003 HDUCLAS2            ALL                             String
0004 ORIGIN              ASC                             String       Source of
0005 CREATOR             cxc - Version CIAO 2.0b         String       tool that
0006 REVISION            1                               Int4
0007 ASCDSVER            R4CU5UPD8.2                     String       ASCDS vers
0008 CHECKSUM            kiJFmhIFkhIFkhIF                String       HDU checks
0009 DATASUM             2822987099                      String       data unit
0010 DATE                2000-09-04T05:09:46             String       Date and t
0011 DATE-OBS            2000-09-02T08:35:02             String       Date and t
0012 DATE-END            2000-09-02T11:17:25             String       Date and t
0013 TIMESYS             TT                              String       Time syste
0014 MJDREF                  50814.0                     Real8        MJD zero p
0015 TIMEZERO            0                               Real8        Clock corr
0016 TIMEUNIT            s                               String       Time unit
0017 TIMEREF             LOCAL                           String       Time refer
0018 TASSIGN             SATELLITE                       String       Time assig
0019 CLOCKAPP            TRUE                            Logical      default
0020 TIERRELA            1E-09                           Real8        default
0021 TIERABSO                0.0010                      Real8        default
0022 TIMVERSN            ASC-FITS-2                      String       Timing sys
0023 TSTART                 84270902.16820               Real8        Observatio
0024 TSTOP                  84280645.81856599            Real8        Observatio
0025 TIMEPIXR                0.50                        Real8        default
.
. (and so on)
.
```

- All of the above examples may be combined into one command that yields the greatest possible amount of information:

  ```
  dmlist "acisf01843N001_evt2.fits[events]" full
  ```

  which is shorthand for:

  ```
  dmlist "acisf01843N001_evt2.fits[events]" blocks,header,cols,subspace,data
  ```

- To save the output to an ASCII file:

  ```
  unix% dmlist acisf01843N001_evt2.fits blocks outfile=blocks.ascii
  unix% more blocks.ascii

  ---------------------------------------------------------------------------
  Dataset: acisf01843N001_evt2.fits
  ---------------------------------------------------------------------------

       Block Name                          Type         Dimensions
  ---------------------------------------------------------------------------
  Block    1: PRIMARY                       Null
  Block    2: EVENTS                        Table        14 cols x 475869   rows
  Block    3: GTI7                          Table         2 cols x 1        rows
  Block    4: GTI0                          Table         2 cols x 1        rows
  Block    5: GTI1                          Table         2 cols x 1        rows
  Block    6: GTI2                          Table         2 cols x 1        rows
  Block    7: GTI3                          Table         2 cols x 2        rows
  Block    8: GTI6                          Table         2 cols x 1        rows
  ```

Using dmcopy

The `dmcopy` tool is a versatile program which you can use to manipulate data. Unlike `dmlist`, which produces text output, this tool produces a new data file in one of the supported formats and writes it to disk.

- Copy the events from the central region of a file into a new FITS file:

  ```
  unix% dmcopy "acisf01843N001_evt2.fits[events][X=3600:4000,Y=3800:4200]" acis_ce
  ```

- Generate a blocked image to allow inspection of the full field of view.

  The pixel range of many X-ray event files is much larger than the default size of an image display (typically 512 square pixels). A convenient way to display the full field of view is to bin the image. In this example, the full range of x and y sky coordinates are blocked by a factor of 4:

  ```
  unix% dmcopy "acisf01843N001_evt2.fits[events][bin x=::4,y=::4][IMAGE]" acis_im
  ```

  Since the image file is a binned version of the events file, only the selected variable (i.e. sky coordinates) is retained. All other information (photon arrival times, energy, etc.) is lost.

- Filter using a region defined in sky coordinates:

  ```
  unix% dmcopy "acisf01843N001_evt2.fits[events][sky=ellipse(1628,4116,92,160,0)]'
  ```

  Note that the name of the column in the filter must match the name within the file itself; in Chandra data, "sky" is shorthand for "(x, y)."

- Filter by the 32-bit event ’status’:

  ```
  unix % dmcopy "acisf01843_000N001_evt1.fits[events][status=11xx0]" acis_filt.fi
  ```

  This filter matches 11110, 11000, etc; the ’x’ is a wildcard, accepting either 0 or 1. Leading zeros may be omitted.

Using dmextract

The `dmextract` tool is intended to be similar to `dmcopy`. It is used to bin tables into images, but writes the binned data to a table instead of creating an image format file.

- Make a histogram from a table column:

  ```
  unix% dmextract "acisf01843N001_evt2.fits[bin pha=1:2048:2]" acis_histogram.pha
  ```

- Extract the PI spectrum of a source in sky coordinates:

  ```
  unix% dmextract "acisf01843N001_evt2.fits[events][sky=ellipse(1628,4116,92,160,(
  ```

Using dmgti

Pipeline processing of Chandra data uses `dmgti` to calculate good time intervals (GTIs) based on input MTL files. Although this is what the tool was designed for, as long as the first column of the input file is

time, `dmgti` will work equally well on any GTI filter.

A look at a lightcurve file (created with the `lightcurve` tool) shows background flares where the count rate reaches values much higher than the mean (several rows were omitted for the sake of space):

```
unix% dmlist "bkg_lc.fits[cols time,exposure,rate,error]" data

--------------------------------------------------------------------------------
Data for Table Block LIGHTCURVE
--------------------------------------------------------------------------------

ROW    TIME                  EXPOSURE             RATE          ERROR

     1    84272506.06420149        39.7920000404         6.6596302986     0.43480700
     2    84272545.85620153        39.7920000404         7.0617208481     0.44695943
     3    84272585.64820157        39.7920000404         7.6397266388     0.46383923
     4    84272625.44020160        39.7920000404         6.6344995499     0.43403545
     5    84272665.23220164        39.7920000404         6.0313630104     0.41506135
     6    84272705.02420168        39.7920000404         6.7852835655     0.43864300
     7    84272744.81620172        39.7920000404         6.5591073036     0.43171215
     8    84272784.60820174        39.7920000404         6.3580617905     0.42545068
     9    84272824.40020178        39.7920000404         6.6847605705     0.43557709
    10    84272864.19220182        39.7920000404         6.5088458061     0.43015587
    11    84272903.98420186        39.7920000404         6.9109368324     0.44244375
    12    84272943.77620190        39.7920000404         6.5842380524     0.43248805
    13    84272983.56820193        39.7920000404         7.614595890     0.463118851
    14    84273023.36020197        39.7920000404         6.9611983299     0.44395437
    15    84273063.15220201        39.7920000404        23.5474472046     0.79470002
    16    84273102.94420205        39.7920000404        20.4563732147     0.74245637
    17    84273142.73620208        39.7920000404        18.0186977386     0.69840312
    18    84273182.52820212        39.7920000404        26.1359062195     0.83586293
    19    84273222.32020216        39.7920000404        41.5410118103         1.0471
    20    84273262.11220220        39.7920000404        58.7052688599         1.2399
    21    84273301.90420224        39.7920000404        89.6662673950         1.5264
    22    84273341.69620226        39.7920000404        55.9408912659         1.2110
    23    84273381.48820230        39.7920000404        62.8266983032         1.2818
    24    84273421.28020234        39.7920000404        44.1797332764         1.0790
    25    84273461.07220238        39.7920000404        61.9471244812         1.2730
     .
     .
     .
    45    84274256.91220312        39.7920000404       141.7872924805         1.9129
    46    84274296.70420316        39.7920000404       106.6546020508         1.6624
     .
     .
     .
   190    84280026.75220849        39.7920000404        73.3815841675         1.3832
   191    84280066.54420853        39.7920000404        74.9899444580         1.3980
   192    84280106.33620857        39.7920000404        72.4014892578         1.3741
   193    84280146.12820861        39.7920000404        75.2915191650         1.4008
   194    84280185.92020865        39.7920000404        80.7197418213         1.4495
   195    84280225.71220867        39.7920000404        80.7700042725         1.4500
   196    84280265.50420871        39.7920000404        72.8538436890         1.3784
   197    84280305.29620875        39.7920000404        72.1501846313         1.3718
   198    84280345.08820879        39.7920000404        75.7187347412         1.4047
   199    84280384.88020882        39.7920000404        72.6025314331         1.3760
   200    84280424.67220886        39.7920000404        75.3166427612         1.4010
```

To exclude the times when the count rate was less than 20.0 and greater than 100.0:

```
unix% dmgti infile=bkg_lc.fits outfile=acis_gti.fits userlimit="(rate>20.0)&&(rate<1
unix% dmlist "acis_gti.fits[gti]" data

--------------------------------------------------------------------------------
Data for Table Block GTI
--------------------------------------------------------------------------------

ROW    START                STOP

     1    84273063.15220201    84273142.73620208
     2    84273182.52820212    84274256.91220312
     3    84274336.49620320    84279788.00020827
     4    84279827.79220831    84280464.46420890
```

The events file may now be filtered on the newly calculated GTI:

```
unix% dmcopy "acisf01843N001_evt2.fits[@acis_gti.fits]" acis_filtered_evt2.fits
```

Getting More Information

Full details on all the dmtools can be found in the *Data Manipulation Users Guide*. Typing `ahelp dm` on the CIAO command line will also bring up more information on the data model.

Updates
05 March 2001 -- original version; identical to CIAO 2.0 version.
22 March 2001 -- modified date format
4 April 2001 -- added link

Return to Threads Page.

Last modified: 4 April 2001