Data conventions

August 15, 2005

Abstract

Several conventions are described for passing around data in the SAS, such as parameter names, file structure, attribute names, column names etc.

1 Introduction

With this note we would like to set the wheels turning on the definition of a number of conventions to be used across the XMM SAS. Eventually this document would be part of the SAS user documentation. We are aware that the SSC have started some work in this area, and we would like to reach a common view of the matter.

2 Dataset structure

Input and output data is processed in the form of data sets, usually saved on disk as a FITS file. A data set can contain several blocks, such as tables and arrays (FITS's HDU).

Traditionally blocks in a data set have been addressed by number (their position in the data set). However the DAL allows the task programmer to address blocks by name, so that it would seem appropriate to ask that all blocks (extensions) have a name (that means the `EXTNAME` FITS key word is set), and that in a data set block names should be unique.

The OGIP conventions for data products all have the block name set, so it is a matter of extending this concept also to utility tasks that normally wouldn't bother with the extension name.

The OGIP `HDUCLASS` set of key words will be used. Possibly it will be necessary to extend the OGIP conventions to describe some other intermediate data set that may be produced during the interactive analysis (TBD).

2.1 Event lists

Events are stored in a table in a data set. There is only one event list in a data set. The name of the table is 'events'.

Each row in a column can contain associated information. This has the form of a column in the table.

Work on the XMM event list definition is done by the SSC. The various columns in the event list would be described here, or a suitable reference given.

2.1.1 Event status

Data type	integer (16 bits?)
Values	0 means the event is good, > 0 the event is flagged as being bad[1]
Location	an extra column in the events table
Naming	column name: `STATUS`

2.1.2 Errors

Data type	real
Units	same as column it refers to
Location	a column in the same table as the column it refers to
Naming	column name: name of the column plus 'Error'

2.1.2.1 Note The error column could also contain the fractional error ('fError'?). Also, should there be a flag indicating what type os statistics applies to the error?

2.2 Images

Images are stored in an array (≥ 2 dimensions) in a data set. There can be multiple images in a data set. The name of the image should indicate its contents.

An image can have several associated arrays. The size of these arrays has to be identical to the image it refers to. The name of the associated array is the name of the image followed by a label indicating the contents of the arrays.

2.2.0.1 Note Should the associated arrays immediately follow the data array they refer to?

2.2.1 ExposureMap

Data type	real
Units	seconds
Location	a block in the same file as the image
Naming	array name: name of the image plus 'Exposure'

2.2.1.1 Note Exposure maps are also products [2].

2.2.2 BadPixelMap

Data type	boolean
Values	true: good pixel, false: bad pixel
Location	a block in the same file as the image
Naming	array name: name of the image plus 'BadPixel'

2.2.2.1 Note Note that the task `BADPIX` is writing a bad pixel extension into the event list.

2.2.3 Errors

Data type	real
Units	same as image
Location	a block in the same file as the image
Naming	array name: name of the image plus 'Error'

3 Parameter names

Parameters are used to configure the behaviour of a task. Several parameters are shared amongst different tasks. The following list contains parameter names and what they are used for.

Name	Data type	Value/units	Description
infile	string		Input data set in case there is only one input
infileN	string		Input data set in case there are multiple inputs
outfile	string		Output data set in case there is only one output
outfileN	string		Output data set in case there are multiple outputs

As a general note: no underscores, parameter names are only lower case, names are as long as necessary to be understabdable.

3.0.0.1 Note Has Dean Hinshaw circulated a note on this?

4 Attributes

Attributes are keyword-value pairs that are used to describe component of a data file. For the FITS input files and FITS output files there are several standards that are obeyed by the SAS:

- OGIP (ref?)
- ODF ICD [1]
- SAS products ICD [2]

4.1 OGIP compliance

4.1.1 Relevant OGIP standards

To be written.

4.1.2 XMM-specific extensions of the OGIP standard

To be written.

4.2 Automatically generated attributes

4.2.1 CREATOR

The attribute CREATOR is automatically added to each data set created by a SAS task (via the DAL). The value is the name and version of the task that created the data set.

4.2.2 DATE

The date of creation is automatically added to the data set as the attribute DATE (via the DAL).

4.2.3 HISTORY

The SAS (via the DAL) will automatically log the command line format of the task, containing the values of all arguments, to the history attribute of a data set.

5 Column names

Column names are upper case.

5.1 Events lists

Name	Values/Units	Description

6 Units

IN general the OGIP convention will be used.

Quantity	Units	Comments
time	s	in UTC TBC
spacecraft position	km	
position in spacecraft reference frame	mm	
energy	eV	

7 Coordinate systems

TBD

7.0.0.1 Note Here we should list column names versus reference frame. For instance, `RAWX` and `RAWY` are CCD node-oriented pixel coordinates etc.

References

[1] ESA. XMM Interface Control Document: Observation and Slew Data Files (XSCS to SSC) (SciSIM to SOCSIM). Technical Report XMM-SOC-ICD-0004-SSD Issue 2.5, ESA/SSD, June 2000. Found at the URL: `ftp://astro.estec.esa.nl/pub/XMM/documents/odf_icd.ps.gz`.

[2] SSC. XMM Survey Science Centre to Science Operations ICD for SSC Products. Technical Report XMM-SOC-ICD-0006-SSC Issue 2.1, SSC, Mar 2000.