CORONAL DIAGNOSTIC SPECTROMETER

SOHO

CDS SOFTWARE NOTE No. 3

Version 4 6 March 1998

CONVERTING CDS TELEMETRY TO FITS FILES

W. Thompson
Applied Research Corporation
NASA Goddard Space Flight Center
Laboratory for Astronomy and Solar Physics
Code 682.1
Greenbelt, MD 20771, USA

William.T.Thompson.1@gsfc.nasa.gov
pal::thompson

1 Overview

This document considers the problem of extracting data from the CDS telemetry stream, and from the planning database to format the data into properly annotated FITS files and to populate the CDS catalog. The purpose is to outline exactly what data will be stored in the FITS files and catalog, and how that data will be obtained.

All keywords are designed to be compatible with those outlined in the document *"Proposed Keywords for SOHO"* which is available from

http://sohowww.nascom.nasa.gov/data/catalogues/

All angles are given in degrees, and are measured in a counter-clockwise sense. All pointing coordinates are in arcseconds from the center of the sun.

2 FITS Binary Tables

In a previous document [3] a number of possible ways of storing CDS data in FITS files [2] were examined, and a recommendation was made for using the binary tables extension together with the optional multidimensional array facility [1]. Here I expand on and refine that concept.

The format of a CDS FITS binary table file can be described as having three parts:

1. The main FITS header. This will describe the observation as a whole. There will be no data array associated with the main header.
2. The FITS header for the binary table. Except for those records peculiar to the FITS structure, this will replicate the information in the main header. It will also describe the columns of the binary table.
3. The body of the binary table. This will contain one complete raster. Every data window will be stored in a separate column. Additional columns will be used to store data about the individual exposures (e.g. time, pointing, etc.).

In the following, each of these components will be discussed in turn.

3 The main and binary table headers

Tables 1–6 list keywords to be used in both the main and binary table FITS headers to describe the observation, organized by how the data is gathered.

3.1 Keywords extracted from the science plan

For some of the keywords it will be necessary to relate the telemetry stream with the planned observations. The keywords are listed in Table 4. The anticipated mechanism would involve

Table 1: Keywords to be used in the FITS primary and binary table headers for SOHO CDS level-0 data. Part 1—trivial keywords.

FILENAME	Name of data file (no path or extension)
SFDUADID	SFDU Authority and ADID label (if applicable)
DATE	Date file was written
ORIGIN	The character string "SOHO-EOF" or "RAL"
TELESCOP	The character string "SOHO"
INSTRUME	The character string "CDS"

determining the time of the observation and looking up what was planned for that time. Knowledge about the previous observation(s) may also be required. This approach must be flexible enough to take into account changing the plan in real time, and for changing observation modes based on flag operations.

In order to link the telemetry stream with the science plan, the software will need to keep a running track of which study is currently being processed. As the study counter changes in the telemetry stream, the software will then compare the information in the telemetry against what it expects from the next item in the science plan. However, in order to find the correct initial study, the software will need to go through the following steps.

1. If not processing real-time telemetry, then first check the CDS catalog. The information may already be recorded for this observation.
2. If not already in the CDS catalog, then check to see if one is still processing the same study as previously. If so, then the parameters should be consistent with where one should be in that study.
3. If not the same study as before, perhaps it's the next study.
4. Otherwise, taking the date of the beginning of the observation, find the study in the science plan which best corresponds to that time.
5. Check the information in the telemetry against that in the science plan. Make sure that this study isn't already in the CDS catalog. (This test is also applied in all the subsequent steps in this list.)
6. If the information doesn't match, then there may be a small mismatch in the actual start time and the planned start time. If the raster counter is one, then find the study with the closest start time.
7. If the information still doesn't match, then the previous study may not have completed yet. Check the telemetry against the previous study.
8. Or, the previous study may have completed earlier than expected. Check the telemetry against the next study.
9. If a match still has not been found, then the spacecraft may have reacted to a flag. Reapply the above steps to the alternate flag plan.

Table 2: Keywords to be used in the FITS primary and binary table headers for SOHO CDS level-0 data. Part 2—keywords derived from the telemetry.

STUDY_ID	The study ID number
OBS_PROG	Name of the study—implied by STUDY_ID
CATEGORY	The study category, either "T" for Test, "S" for Science or "C" for Calibration—implied by STUDY_ID
PROG_NUM	Study counter number
SEQ_IND	The raster index counter within the study
RAS_ID	The raster ID number—determined from STUDY_ID and SEQ_IND
DETECTOR	Either "NIS" or "GIS"
SLIT_NUM	The number of the slit used
EXPTIME	Exposure time
EXPCOUNT	Number of exposures per raster
OBS_MODE	Observing mode—a character string combining the values of DETECTOR, SLIT_NUM, RAS_ID, and EXPTIME, e.g. "NIS-S1-R3-E100"
DW_ID	Data extraction window ID
NWINDOWS	Number of data extraction windows
LL_ID	The linelist ID number—implied by DW_ID
LL_TITLE	The name of the linelist used—implied by DW_ID
XSTEP	Step size in X direction—deduced based on STUDY_ID and SEQ_IND
YSTEP	...
NX	Number of X steps
NY	Number of Y steps
COMP_ID	Compression ID
COMP_OPT	Compression option
DATE_OBS	Starting date/time of raster (CCSDS format)
DATE_END	End date/time of raster
OBT_TIME	On-board value starting time (double-precision integer)
OBT_END	On-board end time
IXWIDTH	Size of instrument field of view—calculated from the pointing information, data window table, and binning factors
IYWIDTH	...
WAVEMIN	Minimum wavelength observed
WAVEMAX	Maximum wavelength observed
INS_X0	Instrument origin pointing (legs) X-axis
INS_Y0	Instrument origin pointing (legs) Y-axis
INS_ROLL	Instrument roll angle (from calibration)
TRACKING	True if solar feature tracking is used—otherwise false

Table 3: Keywords to be used in the FITS primary and binary table headers for SOHO CDS level-0 data. Part 2 continued—keywords derived from the telemetry.

SER_ID	On-board sequence load ID—taken from the "Sequence ID" in the telemetry stream (**TBC**)
OPSLBITS	OPS-L status bits at start of raster, as a character string of "0"s and "1"s.
OPSRBITS	...
OPS_L	OPS-L position at start of raster
OPS_R	...
SLIT_POS	Slit position at start of raster
MIR_POS	Mirror position at start of raster
EV_ENAB	True if event recognition is enabled
COMP_ERR	True if a data compression scheme error was encountered
	The following keywords are only present for VDS data
VDS_ORI	True if VDS telemetry data oriented by columns
VDS_ACC	True if VDS is operated in accumulate mode

Table 4: Keywords to be used in the FITS primary and binary table headers for SOHO CDS level-0 data. Part 3—keywords derived from the science plan.

STUDYVAR	The study variation index
RAS_VAR	The raster variation index
PROG_ID	Observing program ID, linking studies together
PROGNAME	Observing program name, if applicable
SCI_OBJ	General science objective
SCI_SPEC	More specific science objective
CMP_NAME	Campaign name (if applicable)
CMP_NO	Campaign ID number (if applicable)
OBJECT	Object planned to be observed
OBJ_ID	Object identification (if applicable)

Table 5: Keywords to be used in the FITS primary and binary table headers for SOHO CDS level-0 data. Part 4—keywords derived from spacecraft data.

XCEN	Center of instrument field of view
YCEN	...
ANGLE	Orientation of instrument to solar north
SC_X0	Spacecraft X pointing
SC_Y0	Spacecraft Y pointing
SC_ROLL	Spacecraft roll

Table 6: Keywords to be used in the FITS primary and binary table header for SOHO CDS level-0 data. Part 5—miscellaneous keywords.

VDS_MODE: VDS readout mode—taken from the state database based on the time of observation.

SEQ_NUM: Unique number for each file. This number is intended to serve as a unique identifier to a specific instance of an raster. Each data file generated by a CDS raster would have a unique value of SEQ_NUM. Also, for each raster file, the combination of PROG_NUM and SEQ_IND would be unique. However, there is no straightforward way to generate SEQ_NUM from PROG_NUM and SEQ_IND. Instead, the correspondences will be maintained in the CDS catalog. See also section 6.3 for a discussion of how SEQ_NUM is generated.

SEQVALID: Either true or false, normally true. The intent of this keyword is to be either true or false depending on whether or not the raster contains useful data. One possible use of this keyword is to signal whether or not the pointing was changed during the execution of the raster. If so, then the software will need to be able to distinguish between solar feature tracking and true pointing changes.

10. If no match can be found, then set all the database information to nulls, and flag the error to the operator.

These steps would also be taken whenever the telemetry diverged from that expected in the science plan.

3.2 Keywords derived from spacecraft data

Some of the keywords will require knowledge not only about the instrument, but also about the spacecraft. These keywords are listed in Table 5. This will require that some process keep track of the spacecraft information (telemetry and orbital data) from which the relevant data can be extracted. Calibrations for both the instrument and the spacecraft will need to be applied to these data.

Table 7: Keywords to be used for the columns in the FITS binary table header for SOHO CDS level-0 data.

TFORM*n*	Data format
TTYPE*n*	Label
TDIM*n*	Dimensions
TDESC*n*	Dimension labels
TCUNI*n*	Units along each of the dimensions
TRPIX*n*	Reference pixel position
TRVAL*n*	Axes values at reference pixel
TDELT*n*	Pixel spacings along axes
TUNIT*n*	Units
TNULL*n*	Data missing flag value
TDMIN*n*	Data minimum value
TDMAX*n*	Data maximum value
TDETX*n*	Start column on detector
TDETY*n*	Start row on detector (NIS)
TWAVE*n*	Principle wavelength
TWMIN*n*	Minimum wavelength
TWMAX*n*	Maximum wavelength
TWBND*n*	Wavelength band
TGORD*n*	Grating order

4 The binary table column headers.

There are a number of keywords that are used to define the columns in the binary table. These are listed in Table 7. Each of these keywords is described in detail below.

The following keywords would be used for all columns in the binary table, for both columns containing data windows and columns containing auxiliary information.

TFORM*n*: The format for column n, describing the data type (e.g. integer*2) and the number of data points. This is generated automatically by the FITS binary table writer.

TTYPE*n*: This is the label. For columns containing auxiliary data, this will be the associated keyword, e. g. "DEL_TIME". For columns containing particular data windows, then this will be the name of the data window, extracted from the linelist database. It would be nice if this could be somehow descriptive, e.g.

```
TTYPE3  = 'HE304   '
```

This label would have to be stored as part of the linelist definition. An alternate but less desirable scheme would simply combine the name linelist with the index of the data window within that linelist, e.g.

```
TTYPE3  = 'L10-4   '
```

referring to window 4 within linelist "L10".

TDIMn: The dimensions of the data. Not needed if only one-dimensional. The IDL FITS binary table writer will generate this keyword automatically.

TDESCn: The labels associated with the dimensions defined by TDIMn. These labels will be the appropriate keywords associated with the physical dimensions of the data, e.g.

```
TDESC3  = '(WAVELNTH,SOLAR_X,SOLAR_Y)'
```

for a three dimensional array, or simply

```
TDESC5  = 'SOLAR_Y '
```

for a one dimensional array (no TDIMn keyword). These names would be based on the properties of the raster used, and would be taken from the following set:

DEL_TIME
SOLAR_X
SOLAR_Y
WAVELNTH

TCUNIn: The units associated with the dimensions defined by TDIMn, e.g.

```
TCUNI3  = '(ANGSTROM,ARCSEC,ARCSEC)'
```

There's a one-to-one correspondence between the value of TDESCn and the value of TCUNIn, namely

TDESCn	TCUNIn
DEL_TIME	SECOND
SOLAR_X	ARCSEC
SOLAR_Y	ARCSEC
WAVELNTH	ANGSTROM

TRPIXn: Character string giving the index or indices of a reference pixel. This will have the same form as TDESCn, e.g.

```
TRPIX3  = '(1,1,1) '
TRPIX5  = '1       '
```

Here, the pixel index refers to its position within the data window, not the detector. The first pixel has the index 1. Integral pixel indices refer to the center of the relevant pixel. Fractional pixel indices are allowed.

TRVALn: Character string giving the value along each axis corresponding to the reference pixel position given by TRPIXn. The units of spectral dimensions will be Ångstroms, the units of spatial dimensions will be arcseconds, and the units of time dimensions will be seconds. The spatial position will refer to an absolute position on the sun, in the coordinate system defined in the SOHO SOP. TRVALn will have the same structure as TRPIXn. If any dimensions are based on mechanism movements, then the TRVALn values for these dimensions will be approximate only, and the precise values for the position of center of each pixel will be stored in separate binary table columns.

TDELTn: Character string giving the pixel spacing along each axis. The units will be the same as for TRVALn (see above). TDELTn will have the same structure as TRPIXn and TRVALn. Any pixel averaging (e.g. binning within the VDS detector) will be reflected here. If any dimensions are based on mechanism movements, then the TDELTn values for these dimensions will be approximate only, and the precise values for the position of center of each pixel will be stored in separate binary table columns.

TROTAn: Character string giving the rotation angles between the coordinate system that the data were taken in, and the coordinate system that is used for TRVALn. Only spatial data will have nonzero values, which will be the same as the ANGLE keyword, e.g.

```
ANGLE   =                          3.52 /Alignment to solar north
TDIM3   = '(10,100,100)'                /Dimensions of column 3
TDESC3  = '(WAVELNTH,SOLAR_X,SOLAR_Y)'  /Dimension names
TROTA3  = '(0,3.52,3.52)'               /Rotation angles of dimensions
```

Rotation will be about the pixel position given by TRPIXn, i.e. the coordinates of the reference pixel position will be set solely by TRVALn, and the coordinates of all other pixels will be determined by calculating from the values of TRVALn, TDELTn, and ANGLE.

TUNITn: Character string describing the units of the data in column n. For data windows this will be "ADC" (Analog-to-Digital Conversion) for NIS data, or "Counts" for GIS data.

TNULLn: Value used to flag missing data. For data windows this will be -1. Note that TNULLn is not used for floating point data.

The following keywords would be used only for those columns storing data windows:

TDMINn, TDMAXn: The minimum and maximum values for the data in column n.

TDETXn: The value of the starting column on the detector for the data window in column n, in pixels. This is taken directly from the telemetry.

TDETYn: The value of the starting row on the detector for the data window in column n in pixels (NIS only).

TWAVEn: The value of the central wavelength of observation for the data window in column n. This will require calibration.

TWMINn, TWMAXn: The minimum and maximum wavelengths of observation for the data window in column n. The extent of the data pixels in the spectral dimension is taken into account. This will require calibration.

TWBNDn: The wavelength band. For NIS this can be either 1 or 2, denoting the short or long wavelength band accordingly. For GIS this can be from 1 to 4.

TGORDn: The grating order for the principal line of interest. No matter what the grating order is, the other wavelength keywords—TWAVEn, TWMINn, TWMAXn—will be based on a strictly first-order calculation. For example, for the Helium II line at 303.8Å the grating order will be 2, but the value of TWAVEn will be something close to 607.6Å.

5 Data to be stored in binary table columns

There are two classes of data which will be stored in the body of the binary table: data windows, and auxiliary information describing the individual exposures making up the raster.

5.1 Data windows

Each data window will be stored in a separate column in the binary table.

In [3] I had proposed that each row in the binary table represent a separate exposure. After much reflection I have come to the conclusion that this plan, although it appears to work fine for VDS, does not fit the requirements of the GIS instrument. A series of GIS exposures could represent a time series, or a one-dimensional spatial scan across the sun, which would be a good match with one exposure per row, but it could also represent a two-dimensional raster scan to form a solar image, which would not.

Rather than using different approaches for GIS and NIS, I now recommend that each binary table contain one and only one record. The dimensions of each column will be increased by either one or two to reflect the multiple exposures, depending on the raster.

There is an additional benefit of writing the data out at the end of a raster, rather than writing it out exposure by exposure. The dimensions of the data can be reordered to facilitate data analysis, rather than being stored in the order the data was taken in. After examining all the different ways that the CDS data can come down, I recommend that the data be organized into one of the following dimensionalities:

(SOLAR_X,SOLAR_Y)
(SOLAR_X,SOLAR_Y,DEL_TIME)
(WAVELNTH,SOLAR_X,SOLAR_Y)
(WAVELNTH,SOLAR_X,SOLAR_Y,DEL_TIME)

In addition, the data should be rearranged so that each dimension increments in the positive dimension. The WAVELNTH and DEL_TIME dimensions would only be present if non-trivial. However, the SOLAR_X and SOLAR_Y dimensions would always be present, even if the number of points along either or both of these dimensions would be 1. The advantage of this is that one can then easily associate pointing information with the data windows through the keywords in Table 7.

5.2 Auxiliary information

As well as the data windows, there will be other columns storing auxiliary data pertaining to the exposures. These are listed in Table 8. They are described in detail below:

DEL_TIME: Floating point array of the start time of each exposure as an offset in seconds from the absolute time given by the keyword DATE_OBS in the header.

INS_X, INS_Y: Floating point arrays of the pointing of the instrument in arcseconds relative to the pointing given by INS_X0, INS_Y0. For CDS, INS_X and INS_Y are controlled by the scanning mirror and slit mechanism. For VDS the value of INS_X and INS_Y represent only

Table 8: Auxiliary information to be stored in columns within the binary table.

DEL_TIME	Time offsets
INS_X	Internal instrument pointing X-axis
INS_Y	...
OPS_L	OPS-L coordinate
OPS_R	...
MIR_POS	Mirror position
SLIT_POS	Slit position
OPSLBITS	OPS-L status bits, written out as a byte array
OPSRBITS	...
SLITBITS	Slit mechanism status bits
PITCH	Sun sensor pitch
YAW	Sun sensor yaw
PTCHBITS	Sun sensor pitch status bits
YAW_BITS	Sun sensor yaw status bits
SOLAR_X	Solar cartesian west (X) pointing
SOLAR_Y	Solar cartesian north (Y) pointing
EV_LEVEL	Event level
EV_RECOG	Event recognition enabled
EV_DETEC	Event detected
EV_VALID	Event valid

the pointing of the central pixel of the detector. The spatial information for each window is stored within the TRPIXn, TRVALn, and TDELTn keywords. INS_X0 and INS_Y0 are controlled by the Offset Pointing System (OPS).

OPS_L, OPS_R: The raw value(s) of the Offset Pointing System coordinates from the telemetry as a function of exposure.

MIR_POS: The raw values of the mirror position as a function of exposure.

SLIT_POS: The raw values of the slit position as a function of exposure.

The dimensionality of the auxilliary data will reflect the dimensionality associated with the exposures, and would have the following possible dimensions:

SOLAR_X
DEL_TIME
(SOLAR_X,SOLAR_Y)
(SOLAR_X,DEL_TIME)
(SOLAR_Y,DEL_TIME)

Each auxilliary column would have the same dimensionality associated with it. It wouldn't matter if the quantity changed during the raster or not. This is a change from the more complicated scheme discussed in an earlier version of this document.

6 Requirements

One of the purposes of this exercise is to determine what requirements are put by the FITS formatting software on the other components of the CDS ground support software system. These are outlined below. (It should be noted that since this section was first written, many of these components have been realized.)

6.1 General databases

The following requirements are placed on various general databases within the CDS EGSE. The term database here is not meant to imply a relational database system—much of this data could simply be stored in flat ASCII files, or some other format(s) as yet **TBD**. Many are stored in a database management system written in IDL, and are also used by the planning software.

Study and raster databases: Contains each valid value of the "Study ID" combined with its name. This database also lists the series of rasters that make up the individual studies. It also contains information about any adjustable parameters within the definitions of the studies.

Linelist and data-window databases: Contains each valid value of the "Line List ID" combined with the name of that linelist and the window definitions (NIS and GIS). This database also contains names for each window in the linelists.

There are also some databases that come out of the calibration process. In each of these databases, successive calibrations will be maintained in the same database, so that a calibration as a function of time can be maintained.

Wavelength calibration database: Contains the calibration from detector pixel to wavelength.

Internal pointing calibration database: Contains the parameters needed to convert slit and scan mirror mechanism values into arcseconds relative to the OPS.

External pointing calibration database: Contains the parameters needed to convert the OPS L and R values into arcseconds relative to the spacecraft pointing.

Spacecraft pointing database: Contains the information about the spacecraft attitude needed to convert the CDS pointing information into an absolute coordinate system. Predicted data will be stored in the database until the final data are available.

6.2 Planning database

In order to properly link together the CDS data with the scientific objectives associated with that data, it will be necessary for the FITS formatter software to have access to the science plan. At a minimum, the science plan must contain the information given in Table 9. The database also contains the parameters that are passed to the study (e.g. number of repeated rasters, exposure times, etc.).

Table 9: Data to be stored within the CDS science plan.

PROG_ID	Observing Program ID
STUDY_ID	Number defining the study
SCI_OBJ	General science objective
SCI_SPEC	More specific science objective
CMP_NAME	Campaign name
CMP_NO	Campaign number
OBJECT	Object planned to be observed
OBJ_ID	Object identification (if applicable)
DATE_OBS	Date and time of beginning of study
DATE_END	Date and time of end of study

The science plan database is organized by time, and is kept up to date as the plan changes. Thus, when changes are made, either before the execution of the plan, or during real time operations, then the plan will be updated. The plan will be kept as a function of time, but old data will be archived off-line. An alternate parallel plan will also be maintained to reflect flag operations, i.e. what CDS will do if it responds to a flag.

6.3 Interaction with the catalog

In order to assign the proper SEQ_NUM identifier to the data, it will be necessary for the FITS formatting software to interact with the catalog software. The following types of interaction will need to be supported:

- The software must be capable of determining whether or not a given study or raster is already listed in the catalog, and return the proper identifier if it is. This will never happen when processing the RT telemetry as it comes down, but will be the case for all but a small fraction of the PB and FD data processed. It will be necessary to insure that small changes in the calibration between when the RT, PB and RT data are processed does not obscure the fact that these data are already in the catalog. To handle those cases which were lost in the RT telemetry but recovered in the PB or FD data, it is proposed that SEQ_NUM be required only to be *unique*, and not necessarily incrementing in strict chronological order.
- If the given study or raster is not already in the catalog (always true for RT telemetry, unless reprocessing from archived files) then a unique number must be assigned. The most likely procedure would involve performing this within the relational database software, and then returning the result to the FITS formatting software.
- Upon completion of the FITS formatting of a study and/or raster, the catalog information must be updated. This will entail either adding new records to the catalog, or replacing the records that are already there.

6.4 Time calibration

It will be necessary to apply a calibration procedure to convert the times encoded within the telemetry into UTC, and to correct for the difference in light arrival times between the spacecraft and the earth. To perform this calibration, the following pieces of information will have to be combined:

- The time anomalies catalog supplied by the ECS.
- The spacecraft orbital data, both predicted and final.
- A database listing the dates of leap seconds.

References

[1] W. D. Cotton and D. B. Tody. Binary table extension to FITS: A proposal. preprint, 1991.

[2] NOST. Implementation of the Flexible Image Transport System (FITS). Technical Report NOST 100-0.3b, NASA/OSSA Office of Standards and Technology, Code 933, NASA Goddard Space Flight Center, Greenbelt MD 20771, USA, November 6 1991.

[3] W. Thompson. Proposed file format for CDS. CDS internal memo, 1992.

APPENDICES

A Alternative keyword configuration

Recently the High Energy Astrophysics Science Archive Research Center (HEASARC) at Goddard adopted a different set of keywords to accomplish the same task as the above proposed "TDIM-like" keywords. Basically, they are using TDIMn to define the dimensionality of the data, but instead of using TDESC, etc., they are using the following equivalents:

TDIM-like keyword	HEASARC equivalent
TDESCn	mCTYPn
TCUNIn	mCUNIn
TRPIXn	mCRPXn
TRVALn	mCRVLn
TDELTn	mCDLTn
TROTAn	mCRTAn

where m is the axis index. For example, instead of having

```
TDESC3  = '(WAVELNTH,SOLAR_X,SOLAR_Y)'
```

one would instead have

```
1CTYP3  = 'WAVELNTH'                      /Wavelength dimension
2CTYP3  = 'SOLAR_X'                       /Solar X (Cartesian West) dimension
3CTYP3  = 'SOLAR_Y'                       /Solar Y (Cartesian North) dimension
```

The HEASARC community was aware of my proposed "TDIM-like" keywords, but decided not to use it because the string values could, in principal, become too long for FITS headers. Also, they felt it would be easier to have one value per FITS keyword, rather than an array of values. The advantage of the HEASARC approach is that one can attach comments to the values for each dimension—the disadvantage is that the binary table header is larger.

We in the SOHO community should consider very seriously whether or not we should switch over to the HEASARC convention to be more compatible with others in the FITS community. I believe that this group represents an important an influential part of the FITS world.