DATA FORMATS FOR IRAM TELESCOPES

January 1983

In devising a data format for the IRAM telescopes, the following general goals have been kept in mind:

1) As far as possible, the data format should be the same for all telescopes, and all forms of data. This means that raw data, spectral line data, continuum data, etc. will have essentially the same format at all sites.

2) The format should be easily useable by an outside observer. Visiting observers should easily be able to decode the header information (of raw data as well as reduced and partially reduced data) and be able to install their own reduction programs on the IRAM computer with the minimum of effort if they want to.

3) The data should be easily transportable to another computer facility. A visitor should be able convert his data to a standard interchange format like FITS with a minimum of effort.

4) The data format and storage of data on the disk should use the DEC/VAX system facilities, DCL language, etc. as much as possible. This is more efficient than developing our own special purpose programs, cuts down on unnecessary software development, and is easily understood by most visiting astronomers since the use of DEC systems is widespread in the astronomical community.

5) The data format can be easily modified in an upward compatible way to allow for new header variable.

To satisify these goals, we propose the following concept for a data format, which we will call IDFITS (IRAM Disk-oriented FITS):

I. The data format will be a slightly modified FITS format (see Wells, et al. 1981, Astron. Astrophy. Suppl., 44, 363). We are essentially proposing a disk version of the FITS format (which was intended as a tape format). The FITS standard will be followed with the following modifications to make it more efficient on the VAX.

A. HEADER

1) Rather than using a fixed length card image of 80 characters for each keyword and associated parameters, the card images (which will also be records - see below) will be of variable length up to 80 characters in length.

2) The parameter may be left justified to column 11 instead of right justified to column 30, as it is in FITS.

3) Blanks are to be ignored between column 10 and the start of the parameter.

4) More tnan 2 parameters are allowed per keyword. Thus, arrays are allowed for keyword parameters. The parameters are separated by blanks or commas.

5) Array keywords are only allowed for variables that are not already defined in the standard FITS definition.

6) If it is necessary to go to the next card image (record) to list all of the paramters (such as for an array), then they will be continued with the keyword CONTINUE.

7) The sizes of the parameter fields within a card image will be the same as for the FITS standard.

B. DATA

8) For the data itself, floating point format will be allowed (either REAL*4 or REAL*8). This will have to be identified with a keyword in the header. If integer data is used, it will be stored in DEC integer format, which is byte swapped in comparison to the FITS standard. In conformity with the FITS standard, integer data can be stored in 1, 2, 4, or 8 byte words (8, 16, 32, or 64 bits).

II. Each raw data scan, reduced spectral line scan, map, etc. will be one IDFITS data set with header and associated data. The record size for the header will be variable length, but never greater than 80 characters. Thus, the header information will be stored in several small records of up to 80 characters in length (roughly one record per keyword). The data will be in in fixed length records of 512 bytes (or 128 32-bit words), stored as binary data.

III. Each data scan, map, etc. will be stored as two files. One file will be the header records and the second the data records. For example, the raw telescope data will be 1 pair of files per scan, all put into a directory assigned to the observer. The name of each file could reflect the scan number and the qualifier would indicate that it is header or data (for example 10776.HEA and 10776.DAT would be the header and data files, respectively, for scan 10776).

We feel that the above scheme offers several advantages to the observer. The data is in an international, standard format; thus, it is easy for the observer to decode the format for his own special use if he so desires. Since the header information is in ASCII and each scan is stored on separate files, the observer can look at the header information, and sort the data using only the DCL/VAX operating system. Thus some simple data "reduction" can be done without any reduction program! And finally, the data is easily converted to a standard FITS format for export to another institution.

The major disadvantage that the above scheme has is that it is not as efficient as could be for header storage. But in comparing it to some other standard headers we estimate that it is only about a factor of 2 longer in header length (the data itself is the same length). This would mean about a 20% increase in size for a 1024 channel spectral line scan. For the data file sizes we are concerned with, this increased header size is not a major concern. In some instances, this format may be more efficient than a "conventional" format, since only the necessary keywords need to be specified with FITS and no extra space is needed for unused parameters and future expansion.

On the following pages are a list of proposed keywords for the header, and some examples of how the proposed data format would be implemented. Additional ones can of course be used, but there should be a consistent set that is relevant to the IRAM telescopes. This list is intended as a starting point for a discussion of which keywords to settle on eventually. It should be kept in mind that ONLY THE NECESSARY KEYWORDS are used with any given data set. Thus, what at first appears as redundancy is in fact different method of handling different types of data. There should be a recommended subset of these keywords for continuum, spectral line, etc.

RECOMMENDED KEYWORDS

I. Required by FITS: Must be first keywords and in order given.

SIMPLE	(logical)	File conforms to basic format?
BITPIX	(integer)	# bits used for each pixel value.
NAXIS	(integer)	# axes in array.
NAXIS1	(integer)	# pixels on fastest varying axis.
NAXIS2	(integer)	# pixels on second fastest varying axis.
NAXISnnn	(integer)	# pixels on nnnth fastest varying axis.

II. Required by FITS: Must be last keyword, and last record of header.

END	(none)	no parameter. Keyword signifies end of header.

III. Optional keywords in FITS standard:

BSCALE	(floating)	scale factor used to convert pixel values to true values (true=[value * BSCALE]+BZERO).
BZERO	(floating)	Offset applied to true pixel values (see above).
BUNIT	(floating)	Brightness units.
BLANK	(integer)	Pixel value assigned to undefined-value pixels.
OBJECT	(character)	Image name.
DATE	(character)	Date file written ('dd/mm/yy')
DATE-OBS	(character)	Date of data acquisition.
ORIGIN	(character)	file writing institution.
INSTRUME	(character)	data acquisition instrument.
TELESCOP	(character)	data acquisition telescope.
OBSERVER	(character)	Observer name/ identification.
(blank)	(none)	cols. 9-80 are a comment.
COMMENT	(none)	cols. 9-80 are a comment.
HISTORY	(none)	cols. 9-80 are a comment.
CRVALn	(floating)	value of physical coordinate on axis n at the reference pixel.

CRPIXn	(floating)	Array location of reference pixel along axis n.
CDELTn	(floating)	Increment in physical coordinate along axis n as FORTRAN counter increases by 1.
CTYPEn	(character)	Type of physical coordinate on axis n.
CROTAn	(floating)	Rotation angle of actual axis n from stated coordinate type.
DATAMAX	(floating)	Maximum data value in the file.
DATAMIN	(floating)	Minimum data value in the file.
EPOCH	(floating)	EPOCH of coordinate system.

Units for BUNIT: 'K', 'JY/BEAM', 'JY/PIX', 'MAG/PIX', 'M/SEC', 'DEGREES', ' '

Units for CTYPE: 'RA', 'DEC', 'LL', 'MM', 'GLON', 'GLAT', 'ELON', 'ELAT', 'TIME', 'FREQ', 'LAMBDA', 'VELO', 'VELO-LSR', 'VELO-HEL', 'VELO-OBS', 'PIXEL', ' ', 'STOKES', 'COMPLEX', 'DISTANCE', 'ANGLE'

Units are consistent with the SI system; m, sec, Kg, and K. Addition of degrees for angles, and Jy for flux.

IV. Suggested Additional Keywords for IRAM Telescopes:

CONTINUE	(any)	continuation of keyword values from last card image.
PIXTYPE	(character)	Specifies whether the data is 'INTEGER' or 'FLOATING'. Default is integer.
SCANNUMB	(integer)	scan number.
VERSNUMB	(floating)	version number of acquision program.
DATATYPE	(character)	type of data. 'RAW', 'CONT', 'LINE',
REFSCAN	(integer)	reference scan number to which the data is paired.
OPERATOR	(character)	telescope operator name/identification.
TITLE	(character)	Program title.
UT-START	(floating)	UT start time (sec).
LST-STRT	(floating)	LST start time (sec).
INTEGTIM	(floating)	integration time of each pixel value.
SITELONG	(floating)	site longitude.
SITELAT	(floating)	site latitude.
SITEELEV	(floating)	site elevation.
NDUMPS	(integer)	number of on/off pairs, or dumps, or ?
NUMSCANS	(integer)	number of scans if an accumulation.
RESTFREQ	(floating)	Rest frequency of center of band for observation.
SYNFREQ	(floating)	synthesizer frequency
LOADD	(floating)	LO addition factor
LOMULT	(floating)	LO multiplication factor. FSKY=SYNFREQ*LOMULT+LOADD.
FSSHIFT1	(floating)	Frequency offset for frequency switching.
FSSHIFT2	(floating)	Frequency offset for frequency switching if a second offset is used.
IFFREQ1	(floating)	frequency of 1st IF.
IFFREQn	(floating)	frequency of nth IF.
LOFREQ1	(floating)	frequency of 1st LO.
LOFREQn	(floating)	frequency of nth LO.
SBCODE	(character)	side band code. 'DSB', 'USB', 'LSB
SENSITIV	(floating)	instrument sensitivity
SENSUNIT	(floating)	sensitivity units for SENSITIV. 'K', 'JY', 'NEP'
ANTEFFIC	(floating)	antenna efficiency
CALTEMP1	(floating)	calibration temperature.
CALTEMPn	(floating)	nth calibration temperature.
CALMETHO	(character)	calibration method. 'CHOP', 'TIP', 'NT', ...

```
TSCALING (floating)      scaling system temperature.
ATMOSTAU (floating)      atmospheric opacity at the zenith.
ATMOSTEM (floating)      atmospheric temperature.
ATMOSPP  (floating)      atmospheric partial pressure of H2O.
ATMOSPRE (floating)      atmospheric pressure.
ATMOSWIN (floating)      wind speed.
ATMOSSKY (character)     sky type.
POINT1   (floating)      pointing constand number 1.
POINTn   (floating)      pointing constant number n.
FOCUS1   (floating)      focus constant number 1.
FOCUSn   (floating)      focus constant number n.
BASELINE (floating)      array of 3 numbers for the A, B, and C factors in
                         a baseline correction made to a spectral line scan.
                         Tcorr = C*Traw**2 + B*Traw + A
RA1950   (floating)      RA in epoch 1950.0
DEC1950  (floating)      DEC in epoch 1950.0
RACURR   (floating)      RA at the current epoch
DECCURR  (floating)      DEC at the current epoch
RA2000   (floating)      RA in epoch 2000.0
DEC2000  (floating)      DEC in epoch 2000.0
LII      (floating)      galactic longitude
BII      (floating)      galactic latitude
AZIM1    (floating)      azimuth at start of record
ELEV1    (floating)      elevation at start of data record
POLARTYP (character)     polarization type. 'R', 'L', 'V', 'H', 'LIN',
POLARANG (floating)      polarization angle. (degrees east of north)
PVLn-m   (floating)      if one of the dimensions of the data array is
                         listed as 'PHASE', then the values that change
                         during this phase can be given with PVLn-m, and
                         PTYn-m.  This is the mth parameter that is changed
                         for the nth phase.  All parameters are the values
                         listed with normal parameters unless given otherwise
                         for that phase, and the value remains changed only
                         for that phase.
PTYn-m   (character)     type of parameter for value given by PVLn-m.
                         'RA1950', 'RESTFREQ', 'DECCURR', etc.
```

V. Additional values for CTYPEn

'PHASE', 'SUBSCAN'

Example

Spectral line scan containing signal and reference and zero offset (considered as 3 phases of data).

File name: 1098.HEA

Record #	Contents
1	SIMPLE = F
2	BITPIX = 32
3	NAXIS = 2
4	NAXIS1 = 512 / NR. OF PTS. PER SPECTRUM

```
NAXIS2   = 3 / NR. OF PHASES
BUNIT    = 'K'
OBJECT   = 'ORION'
DATA     = '01/04/84'
DATA-OBS= '01/04/84'
INSTRUME= '3MM RX'
TELESCOP= '30M'
OBSERVER= 'ABC'
OBERATOR= 'XYZ'
HISTORY   RAW DATA SCAN.
COMMENT   THIS IS AN EXAMPLE OF A DATA FORMAT.
CRVAL1   = 8000.0
CRPIX1   = 255.5
CDELT1   = 2600
CTYPE1   = 'VELO-LSR'
CRVAL2   = 1
CRPIX2   = 1
CDELT2   = 1
CTYPE2   = 'PHASE'
PIXTYPE  = 'FLOATING'
SCANNUMB= 1098
DATATYPE= 'PS-LINE'
RA1950   = 60.1234
DEC1950  = -5.4455
PVL2-1   = 60.5566
PVL2-2   = -5.0000
END
```

QUIDELINES FOR USING THE DATA FORMAT:

1) Each data set should be self contained for simplicity. This means that for spectral line scans, the signal and reference positions will be stored in the same data file (and header file) for position switching. If a scan creates two separate spectra (such as with a split receiver) then the data should be stored as two separate data sets.

2) If all subscans, or dumps, of a scan are kept instead of averaging together, then they can be treated as another dimension in the data.

3) Since any data format that is specified is likely to be extended, there should be a master list of "approved" keywords for both 30m and P.d.B. telescopes. Fortunately, the sytem proposed here is upward compatible, since any new keyword would be treated as a comment by an old program.