Any comments? Rachael

E-S

Minimum single-dish spectral-line data format [MT.DOC]DATFMT02.RNO

1 RATIONALE

It seems to me to be beyond question that if we want observatories in general (many of which do not support computer professionals on their staffs) to adopt our proposed standard data format, we have to make it accessible. This means it must be possible to write a recognizable file without (1) requiring any specialist knowledge of data-base systems, and (2) without having to provide lots of data items which may be either unknown or unnecessary in most contexts. Furthermore, if we all decide to adopt this format as the basis of our data reduction system, then it must be possible to include data from other wavebands with a minimum of difficulty. In practice this means that we should not require radio-specific data (aperture efficiency etc).

The data items I have selected refer specifically to the spectrum, and not to the observation. I don't doubt that we will include lots of observation specific data in the rest of the header, for our own purposes, but so much of it is specific to our individual packages that it seems foolish to require any of these items.

There is a second reason for wanting to define a minimum format. This is one of completeness. It is not always intuitively obvious just which parameters are required to specify an observation fully. As an example, it is worth noting that none of the UKIRT programs produce complete data headers, in the sense that it is not possible to reconstruct the frequency axis of the plot accurately, although to date the errors involved have been minimal.

As noted in a number of places below, there are several possible complete subsets in most cases. I would like us to choose one of these, rather than having several possible alternatives, which would mean a data item to say 'I have used minimum format number n'. This will assist in documentation ("If you do not supply AT LEAST the following data the reduction programs cannot work properly"), and should also reduce the amount of programming required for reduction packages. The minimum set I have supplied strikes me as being in some way the most fundamental, in that it makes the minimum number of assumptions about conventions etc.

I am not sure yet how these data items should be included in the general single-dish data format (GSDD format). I would guess that they should be in a class by themselves at the front of the header. Right now, I would like some feedback from others as to whether this sounds like a reasonable idea. (1) is this in fact a complete format?, (2)What do you think of the data-types and units suggested?, (3) is this even necessary? Answers on a postcard please.

2 MINIMUM REQUIRED DATA

In order to identify a spectrum absolutely for later use the following items are required:

1. The time at which the spectrum was measured (by convention the start time)

2. The position on the sky at which the centre of the beam was pointing.

3. The X co-ordinate system of spectrum (e.g. centre frequency and channel spacing).

4. The noise level on the spectrum.

Any reduction system must be capable of operating on data which have only this information, even if not all possible operations in the reduction system can be carried out without further details. With the information given above it is always possible to:

1. Display the data with properly labelled axes.

2. Perform any 1-d data-type-independent operations (baseline removal, addition, subtraction, normalization by a control spectrum, etc).

3. Merge the data (via averaging or otherwise) with data from other instruments (subject to an intelligent understanding on the part of the observer).

3 POSSIBLE COMPLETE DESCRIPTIONS

3.1 Time:

The possibilities include UT vs LST or local time, given as arrays (IHRS, IMIN, ISEC...), characters ('hh:mm:ss.ss'), or reals (in radians, hours, days).
IF the time is local time or LST, THEN we also need to include the difference between local time and GMT, or the telescope longitude, for a complete description which would allow comparison of data from different observatories.
Only if we specify the time as a real number is it immediately useful without conversion routines. On the other hand, for most purposes the time is not actually used. I suggest we adopt a character format as above, but clearly it doesn't matter.
If we adopt a real format, then most people, me among them, understand hours better than radians or fractions of days. Given that we are discussing a VAX specific format, and that the VAX now has routines to generate the trig. functions with arguments in degrees, I see no reason to prefer radians.

My preference: UT hours as character*16 variable 'hh:mm:ss[.sss]'

3.2 Date:

Here the serious contenders include Julian day number and Gregorian date. Most low-resolution spectra do not require accurate LSR correction, and so the Julian date is not necessarily readily available in the data. If we were to insist on Julian date to describe the date of observation we might be making a significant amount or work for people who wish to use our packages, but do not have Julian date immediately available. Once again, although the date specifies the spectrum, it is not used in most applications.

My preference: UT date as character*16 variable '[yy]yy-mon-dd'

3.3 Position:

There are clearly any number of options - Galactic, ecliptic or equatorial co-ordinates, in any epoch, and equinoxes B1950 and J2000, at the least. Plus an infinite number of others, which we can't possibly code for in advance. I believe we should adopt one of these, and I really don't care which, although for most people equatorial co-ordinates convey information most readily (when does the source rise, how far over is it, how long can I see it for???). This is not to say that we can't give the others in the expanded (non-minimum) header, but if we had to settle on one, then I would guess that there are very few telescopes where the RA and Dec. are not at least worked out along the way.

Probably, in order to allow for observations of comets, the moon, various planets and their satellites, we need to allow a 3-dimensional co-ordinate system.

```
My preference: RA as REAL*8 Hours
               Dec as REAL*8 Degrees
               Horizontal Parallax as REAL*8 arcseconds.
               Epoch as REAL*8 years
               Equinox as INTEGER*4 variable - 0 = B1950.0
                                               1 = J2000.0
               Observatory Latitude   REAL*8 Latitude degrees
               Observatory Longitude  REAL*8 Longitude degrees
               Height above reference geoid  REAL*4 metres

Second preference:  Include a further variable

               Co-ordinate system as INTEGER*4 - 0 = Equatorial
                                                 1 = Galactic lII,bII
                                                 2 = Ecliptic
                                                 3 = Alt-azimuth.
                                                 etc.
```

3.4 X-axis Scaling:

For many purposes the X-axis of the spectrum is well-described by a centre frequency, either observed or LSR corrected, and a frequency increment. However, when observing spectra in other wavebands, or with an AOS, this may not be so. It is therefore essential that we include the option of giving both the datum and the corresponding X-value.

Assuming a linear scale, this will almost certainly be linear in the observers frame, and not in the rest frame of the emitter. Hence for accuracy we should specify the observed centre frequency and increment, rather than the sky frequency at the centre of the band.

In order to display this we therefore need the assumed velocity offset. Some of the many (complete) specifications include:

1. Assumed lsr velocity of source plus (known) velocity of observatory w.r.t. the lsr at the time of observation.

2. Just the total velocity of the source w.r.t the earth (less convenient for display)

3. Assumed lsr velocity of source, the time of observation, position of source and position of the observatory.

4. Rest-frame centre frequency, time of observation, position of source and position of observatory.

If we had to cater for just one of these (in our minimum) I would choose the latter. The rest-frame centre frequency is useful in itself (as an indicator of the line being observed) and is independent of the lsr convention being used.

My preference:	Observed centre frequency	REAL*8 Hz
	Rest frame centre freq.	REAL*8 Hz
	Number of data channels	INTEGER*4 channels
	Frequency increment	REAL*8 Hz
		(if = zero, then observed frequencies are included with data as (x,y) pairs).

3.5 Noise Levels:

These are usually specified by a system noise temperature and an integration time. This has the disadvantage that it is not clear ever if integration time refers to time on source, clock duration of observation, observing time only but including the offs, etc.. Further, we then assume that spectra are calibrated in antenna temperatures, which may not always be true. For purposes of optimal averaging, or more importantly, for discerning the significance of a feature, all we require is the actual noise level.

My preference: Noise level, REAL*4 (arbitrary units, to suit data)

Totals: 2 Character*16, 8 Real*8, 2 Real*4, and 2(3) Integer*4 variables.

r padman 15 July 1985