Metadata for VO-compatible Spectral Files in FITS Format

Inga Kamp (Space Telescope Division of ESA/STScI), Randy Thompson (CSC/STScI), Myron Smith (CSC/STScI), Alberto Micol (ESA/ST-ECF), Daniel Durand (CADC), Karen Levay (CSC/STScI), Rick White (STScI)

Abstract

The Multimission Archive at the Space Telescope Science Institute (MAST) in collaboration with the Space Telescope European Coordination Facility (ST-ECF) and the Canadian Astronomy Data Centre (CADC) is working on making its spectral data (UV and optical) available to the VO through the Simple Spectral Access Protocol. The bottleneck so far has been the collection of metadata, especially for heritage instruments. We follow closely the VO Data Characteristics Model for determining the different levels of data characterization. We also address the need for additional metadata such as a 98 percentile flux to display spectra and characterize continuum levels. Creating the proper keywords for FITS serialization has also required much effort. FITS conventions such as TDMINn, TDMAXn, etc. will be used where possible. There is a need to standardize algorithms for derived metadata such as SNR in order to assure meaningful comparison between VO retrieved spectral data.

Collecting the Metadata

• Metadata from original FITS files: mine the FITS keywords of the individual missions for metadata such as plate scale, errors, begin and end time of observation.

• Metadata from instrument manuals: look through the various instrument/mission handbooks and search for "typical/representative" metadata such as aperture, or spectral resolution; take metadata from the instrument manuals when it cannot be computed from the actual data.

• C… …er metada… …ion on how me…

New metadata:

No. of spectral samples per res. element:

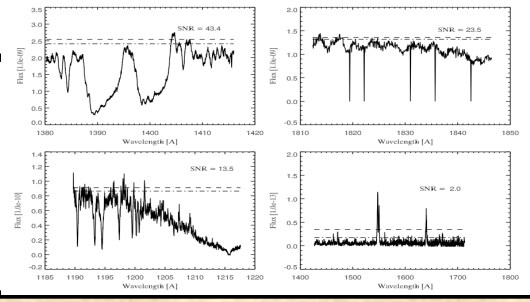
SPECSAMP = SPECDISP/(SPECRES*SPCOMID)
with SPECRES in [ ] & SPECDISP [Å/pixel]

98 percentile flux to represent the continuum:

fluxsort = flux(sort(flux)) contin_2 =
fluxsort(fix(0.98*nflux))

Our algorithm for calculating SNR (in IDL) for the SSAP

adiff = abs(flux-shift(flux,2))
noise = 1.0483581 * median(adiff)
snr = mean(flux)/noise



98 and 95 percentiles (dashed line and dash-dotted line) calculated for a range of GHRS spectra. In most cases, even emission line spectra, the 98 or 95 percentile is a good representation of the continuum flux level. In all cases, it is a useful quantity for tools that display spectra.

Also given on the graph is the signal-to-noise ratio derived from the data using the above algorithm.

Creating the Spectral Container

The existing FITS spectral files archived within MAST were converted using programs written in IDL. New keywords were populated using the sources of metadata described above. For consistency, calculated values are preferred over other sources. Database tables for handling VO service requests are generated based on the values extracted from the above shown VO-FITS keywords. The units in the FITS header follow the OGIP/93-001 recommendation, while the database units follow the SSAP paper recommendation (SI units).

Storing the metadata in the FITS headers

Following the Data Characterization Model paper (v.0.93), the metadata on the four axes (spatial, temporal, spectral, and observable) can be divided into three categories: Coverage (**location**, **bound**, **support**), **Resolution**, and **Sampling Precision**. The example header below illustrates the FITS serialization of this metadata and its sources. We use the FITS convention to store metadata for table columns in TFORMn, TTYPEn, TUNITn etc. and to store units in [ ] as described in Greisen et al., WCS paper I.

FITS keyword	value	description	source
General:			
RADECSYS=	'FK5'	/	hardcoded
TIMESYS =	'TT'	/	"
TIMESDIN=		T /	"
EQUINOX =	2000.00	/	"
MJDREF =	0.00000	/ [d] MJD zero point for times	"
Spatial:			
RA_NOM =	194.2597814434	/ [deg] right ascension of aperture	original header keyword
DEC_NOM =	22.0314330156	/ [deg] declination of aperture	"
APERTURE=	'52X2'	/ [arcsec] aperture length & width	translate header keyword
PLATESC =	0.051	/ [arcsec/pixel] plate scale	original header keyword
Temporal:			
TMID =	50575.45075772	/ [d] MJD mid time	(TSTART+TSTOP)/2
TSTART =	50575.44703086	/ [d] MJD exposure start time	original header keyword
TSTOP =	50575.45448457	/ [d] MJD exposure stop time	"
Spectral:			
TFORM1 =	'1024D '	/ Data Format for Field 1	hardcoded
TTYPE1 =	'SPCO '	/ Data Type for Field 1	"
TUNIT1 =	'angstrom'	/ Units for Field 1	"
TUCD1 =	'em.wl'	/ UCD for field 1	"
SPECSDIM=	'10-10 L'	/	"
SPCOMID =	2369.952	/ [angstrom] Mean Wavelength	(TDMIN1+TDMAX1)/2
TDMIN1 =	1667.648	/ [angstrom] Min Value for Field 1	original header keyword
TDMAX1 =	3072.257	/ [angstrom] Max Value for Field 1	"
SPCO_RES=	910.000	/ [ ] Approx. spectral resolution	"
SPCODISP=	5.0	/ [angstrom/bin] dispersion	(TDMAX1-TDMIN1)/npoints
SPCOSAMP=	2.5	/ [bin**(-1)] sampling	SPCODISP/SPCO_RES
SPCOWID =	1404.609	/ [angstrom] Bandpass Width Wmax - Wmin	TDMAX1-TDMIN1
Observable/Flux:			
TFORM2 =	'1024E '	/ Data Format for Field 2	hardcoded
TUNIT2 =	'erg/cm**2/s/angstrom'/	Units for Field 2	"
TUCD2 =	'phot.fluDens' /	UCD for field 2	"
FLUXSDIM=	'10+7 ML-1T-3'	/	"
FLUXSAT =		/ [] saturation	difficult for many missions!
FLUXLIM =		/ [erg/cm**2/s/angstrom] Min Value for Field 2	(TDMAX2-TDMIN2)/npoints
TDMIN2 =	2.25E-14	/ [erg/cm**2/s/angstrom] Min Value for Field 2	original header keyword
TDMAX2 =	1.50028E-13	/ [erg/cm**2/s/angstrom] Max Value for Field 2	"
SNR_MEAN=	9.688	/ [ ] Mean signal-to-noise ratio	calculated from data
CONTIN_2=	1.02E-13	/ [erg/cm**2/s/angstrom] 98th percentile flux called flux	

Suggestions for Simple Spectral Access Protocol and Data Models

• new FITS keywords to describe the metadata such as TUTYPn, TUCDn

• new metadata relevant to astrophysical applications such as number of spectral samples per resolution element, and "continuum flux"

• standardization of definitions for derived metadata such as SNR and resolution.

• VO documents should define the metadata better including its units to avoid ambiguities. Examples for confusion are resolution, aperture, errors, noise and not yet standardized quantities such as the "continuum flux".