Simple Spectral Data Model
Version 0.01

IVOA Working Draft
2003-12-05

This version:
http://www.ivoa.net/Documents/WD/DAL/DMSSP-20031210.html
Latest version:
http://www.ivoa.net/Documents/WD/DAL/DMSSP.html
Previous versions:
None.
Note: ivoa.net links are not yet active; the above are placeholders.
Editors:
Jonathan McDowell, Doug Tody, Tamas Budavari, Arnold Rots, Roy Williams
Authors:
IVOA Data Access Layer Working Group

Abstract

We propose a convention for representing one-dimensional photon spectra in a FITS Binary Table. This convention does not require extensions to the FITS standard. It is back compatible with the PHA file format used in the high energy astronomy community.

Status of this document

This is a Working Draft. The first release of this document was 2003 XXX XX.

This is an IVOA Working Draft for review by IVOA members and other interested parties. It is a draft document and may be updated, replaced or obsoleted by other documents at any time. It is inappropriate to use IVOA

Working Drafts as reference materials or to cite them as other than "work in progress". A list of current IVOA recommendataions and other technical documents can be found at http://www.ivoa.net/Documents/.

Contents

1 Introduction and Motivation

Spectra are stored in many different ways within the astronomical community. In this document we present a proposed abstraction for spectra and serializations in VOTABLE and FITS, for use as a standard method of spectral data interchange.

We distinguish in several places between the implementation proposed in this document, referred to as Version 1, and capabilities proposed for possible later implementation.

2 Requirements

We need to represent a single 1-dimensional spectrum in sufficient detail to understand the differences between two spectra of the same object and between two spectra of different objects.

We need to represent time series photometry, with many photometry points of the same object at different times.

We need to represent spectral energy distributions (SED) which consist of multiple spectra and photometry points, usually for a single object.

3 Spectral data model summary

Our model for an SED is a set of spectra, some of which may have ony one or few data points (photometry) and each of which may have different contextual metadata (time, aperture, position, etc.). Specifically, a spectrum will have arrays of the following values:

- Spectral coordinate (e.g. wavelength), central and bin min and max
- Flux value, with upper and lower statistical (uncorrelated) errors
- Quality mask
- Spectral resolution

and will have the following associated metadata:

- Time of observation, given as MJD UTC
- Instrument and filter ID
- Exposure time in seconds
- Position of aperture center, given as J2000 ICRS degrees.
- Aperture in arcseconds
- Systematic (correlated) error
- Bibcode

In the following sections we elaborate these concepts in detail, including some complications that we explicitly do not attempt to handle in this version.

4 Spectral data model concepts

4.1 Spectral coordinate

Astronomers use a number of different spectral coordinates to label the electromagnetic spectrum. The cases enumerated by Greisen et al. (2003) are listed below with their UCDs. At least one of these fields should be present. We distinguish between the VO data model field name, the FITS WCS name, and the UCD1 and UCD2 names. For UCD2, I propose UCDs for those not currently covered by the document.

Note 1: For version 1, we adopt only the first three entries, Wavelength, Frequency, Energy. The others are considered reserved names to be considered for future implementations.

Note 2: For the velocity cases, I propose an *em.velocity* tree, rather than a *phys.velocity.radial* tree, because the velocity here is really a labelling of a spectral coordinate, and the link to the physical radial velocity of the different emission sources contributing to the spectrum is rather indirect.

Note 3: I propose to use the same data model field name, Velocity, for all the different kinds of velocity and allow software to distinguish between them using the UCD. Thus, in this proposal there is not a 1-1 relationship between data model fields and UCDs - UCDs are parameters to the data model object and can be used to express different frames. We should discuss whether radio velocity and optical velocity are different parameterized instances of the SpectralVelocity class or whether they are different subclasses.

Field	FITS WCS	UCD1	UCD2	Meaning
Wavelength	WAVE	SPECT_WAVELENGTH	em.*wavelength*	Wavelength
Frequency	FREQ	OBS_FREQUENCY	em.*frequency*	Frequency of photo
Energy	ENER	SPECT_LINE_ENERGY	em.*energy*	Photon energy
Channel	-	ID_SET	em.*channel*	Instrumental spect
LogWavelength	WAVE-LOG		em.*wavelength.log*	Log Wavelength
LogFrequency	FREQ-LOG		em.*frequency.log*	Log Frequency of p
LogEnergy	ENER-LOG		em.*energy.log*	Log Photon energy
Wavenumber	WAVN		em.*wavenumber*	Wavenumber
Wavelength	AWAV		em.*wavelength.air*	Wavelength
SpectralVelocity	VELO	SPECT_VELOC	em.*velocity*	Apparent radial ve
SpectralVelocity	VRAD		em.*velocity.radio*	Radio velocity
SpectralVelocity	VOPT		em.*velocity.optical*	Optical velocity
SpectralVelocity	BETA		em.*velocity.beta*	Velocity (c=1)

4.2 Spectral Intensity Fields

At least one of the following fields should be present. If more than one is present it is up to the application to decide which to use.

Field	FITS WCS	UCD1	UCD2	Meaning
Flux	-	PHOT_FLUX	phot.flux,em.*wavelength*	Flux density p
Flux	-	PHOT_FLUX	phot.flux,em.*frequency*	Flux density p
Flux	-	PHOT_FLUX	phot.flux,em.*energy*	Flux density p
Flux	-	PHOT_FLUX	phot.flux,em.*wavelength.log*	Flux density p
Flux	-	PHOT_SB	phot.sb, em.*wavelength*	Surface bright
Counts	-	PHOT_COUNTS	phot.count	Counts in spec
Rate	-	PHOT_COUNT-RATE	phot.count?	Count rate in s
FluxRatio	-	SPECT_FLUX_RATIO		Flux ratio of tw
Luminosity	-	PHYS_LUMINOSITY		Luminosity per
Luminosity	-	PHYS_LUMINOSITY		Luminosity per
Luminosity	-	PHYS_LUMINOSITY		Luminosity per
Luminosity	-	PHYS_LUMINOSITY		Luminosity per
EnergyDensity	-	PHYS_DENSITY_ENERGY		Radiation ener
Intensity	-			Flux per unit s
TempAntenna	-	INST_ANTENNA-TEMP		Antenna tempe
TempBrightness	-	INST_SKY_TEMP		Brightness tem
Magnitude		PHOT_MAG		Magnitude in c
MagnitudeAB		PHOT_MAG		AB (spectroph
BeamFlux	-	-		Flux per resolu
Brightness	-	PHOT_SB	phot.sb, em.*wavelength*	Surface brightn
BrightnessMag	-			Surface brightn

5 Accuracy Fields

We express the bandpass for each spectral bin as a low and high value for the spectral coordinate. The same is done for photometry points, which amounts to approximating a filter by a rectangular bandpass.

We also use a very simple error model: we include plus and minus flux errors, and a quality flag. The errors are understood as 1 sigma gaussian errors which are uncorrelated for different points in the spectrum. If the data provider has only upper limit information, it should be represented by setting the flux value and the lower error value equal to the limit. In general applications may choose to render measurements as upper limits if the flux value is less than some multiple (e.g. 3) of the lower error. We also allow a systematic error value, assumed constant across a given spectrum and fully correlated (so that, e.g. it does not enter into estimating spectral slopes).

We also include a trivial spectral resolution model: a single number nominally representing a FWHM spectral resolution expressed in the same units as the spectral coordinate.

Field	FITS WCS	UCD1	UCD2	Meaning
Wavelength.BinLow	-	SPECT_WAVELENGTH	stat.min,em.*wavelength*	Wavelength
Wavelength.BinHigh	-	SPECT_WAVELENGTH	stat.max,em.*wavelength*	Wavelength
Frequency.BinLow	-	OBS_FREQUENCY	stat.min,em.*frequency*	Frequency of
Frequency.BinHigh	-	OBS_FREQUENCY	stat.max,em.*frequency*	Frequency of
Energy.BinLow	-	SPECT_LINE_ENERGY	stat.min,em.*energy*	Photon energ
Energy.BinHigh	-	SPECT_LINE_ENERGY	stat.max,em.*energy*	Photon energ
Flux.ErrorLow	-	ERROR	meas.error,phot.flux,em....	Negative err
Flux.ErrorHigh	-	ERROR	meas.error,phot.flux,em....	Positive erro
Flux.SysError	-	ERROR	meas.error.*sys*	Systematic e
Flux.Quality	-	CODE_QUALITY	meta.quality,phot.flux,em....	Quality mas
Flux.Resolution	-	SPECT_RESOLUTION	meas.resolution, phot.flux	Spectral reso

6 Coverage Fields

The coverage fields will have a constant value for a given spectrum.

We allow strings to identify the instrument and filter used, and a value giving the effective exposure time (useful for selecting among multiple spectra from the same instrument). The aperture is important to determine what part of an extended object is contributing to the spectrum; in version 1 we allow a simple aperture description consisting of a single number representing the aperture diameter in arcseconds; it is anticipated that a full region description will be supported in later versions.

Field	FITS WCS	UCD1	UCD2	Meaning
Instrument	INSTRUME	INST_ID	*inst.id*	Instrument ID
Filter	-	INST_FILTER_CODE	*inst.filter.id*	Filter or dispers
Coverage.Position		POS_EQ_MAIN	pos.eq	RA and Dec
Coverage.Aperture		INST_APERT		Aperture region
Coverage.Time		TIME	time.*obs*	Midpoint of exp
Coverage.Exposure	-	TIME_INTERVAL	time.interval, phot.spectrum	Exposure time

6.1 Frame fields

For all the numeric fields, units must be supplied. For the spectral coordinate, position and time we need to add further metadata to define the frames used. Here we try and match existing FITS conventions (TIMESYS is a convention used in the X-ray community).

Other frame information needed for velocity spectral coordinates include the observation-fixed spectral frame, the observatory location, the source redshift, and the velocity zero point (in Greisen et al, SSYSOBS, OBSGEO, VELOSYS, RESTFRQ/RESTWAV).

Field	FITS WCS	UCD1	UCD2	Meaning
Position.Frame.Type	RA,DEC	-	*frame.pos.type*	e.g. Equatorial
Position.Frame.Equinox	EQUINOX	TIME_EQUINOX	time.equinox, pos.eq	e.g. 2000.0
Position.Frame.System	RADECSYS	ID_FRAME	*frame.pos.system*	e.g ICRS or FK5
Time.Frame.Type	TIMESYS		*frame.time.scale*	Timescale, UTC TT et
Time.Frame.Zero	MJDREF		*frame.time.zero*	Zero point of timescale
Wavelength.System	SPECSYS		*frame.em.system*	e.g. BARYCENT; use

6.2 Curation model

The Curation field is an object derived from the VO Resource Metadata document.

Field	UCD1	UCD2	Meaning
Publisher	ID	human.publisher, meta.curation	Publisher
PubID	ID	meta.curation.pubid	URI for VO Publisher
Creator	ID	meta.curation.creator	VO Creator ID
Logo		meta.curation.logo	URL for creator logo
Contributor	ID	human.contributor, meta.curation	Contributor (may be many)
Date	TIME	time, soft.dataset, meta.curation	Data processing/creation date
Version	ID_VERSION	soft.dataset.version, meta.curation	Version of dataset
Contact		human.contact, meta.curation	Contact name
Contact.Email		meta.curation.email	Contact email

7 Relationship to general VO data models

The Spectrum model involves objects addressed by the proposed VO Observation and Quantity data models. Although these models have not yet been fully worked out, we may note that a single Spectrum maps to the Observation model, which will include the Curation and Coverage objects. The Flux and the spectral coordinate entries together with their associated errors and quality will be special cases of the Quantity model, as will the simpler individual parameters. The field structure presented here is consistent with current drafts of the models.

8 FITS serialization

We define a reference serialization of this data model as a FITS binary table. The format is similar in spirit to the X-ray PHA type II dataset. It represents each spectrum or photometry point as a single row of the table. Variable-length arrays are used to contain the array quantities.

Here we give the mapping of data model fields to FITS columns and keywords. For each column, the standard keywords TTYPEn, TUNITn, TFORMn should be provided. In addition, we define a new keyword TUCDn which should contain the UCD2 string for each column. Order of keywords and columns is not significant, except that it is strongly recommended that RA and Dec be in adjacent columns or keywords.

We adopt the convention that columns which are constant (same value for all rows) may if desired be omitted and the value given as a keyword instead. (e.g. the column TTYPEn='INSTRUME' is replaced by a keyword INSTRUME = 'value'). This is a trivial overhead in the FITS reading interface.

We add a new keyword VOCLASS to describe the VO object represented by the FITS table.

FITS keyword	Data model field	Value
VOCLASS	Spectrum	Spectrum (required)
EQUINOX	Position.Frame.Equinox	2000.0 (required)
RADECSYS	Position.Frame.System	either ICRS or FK5
TIMESYS	Time.Frame.Type	UTC (required)
MJDREF	Time.Frame.Zero	0.0 (required)
SPECSYS	Wavelength.Frame.System	(see Greisen et al)
VOPUB	Curation.Publisher	
VOPUBID	Curation.PubID	
VOCREATE	Curation.Creator	
VOLOGO	Curation.Logo	
CONTRIBn	Curation.Contributor	
DATE	Curation.Date	
VERSION	Curation.Version	
CONTACT	Curation.Contact	
EMAIL	Curation.Contact.Email	

The following fields are scalar columns. *They may be used as keywords if they are constant for the whole table.* The Position.Type field is not explicitly serialized: the names of the two columns used for Position are used to infer the type.

FITS TTYPEn	Data model field
INSTRUME	Instrument
FILTER	Filter
RA	Coverage.Position
DEC	Coverage.Position
APERTURE	Coverage.Aperture
TIME	Coverage.Time
EXPOSURE	Coverage.Exposure
SYS_ERR	Flux.SysError

The following fields are variable-length array columns.

FITS TTYPEn	Data model field
WAVE	Wavelength
WAVE_LO	Wavelength.BinLow
WAVE_HI	Wavelength.BinHigh
ENERGY	Energy
ENERG_LO	Energy.BinLow
ENERG_HI	Energy.BinHigh
FREQ	Frequency
FREQ_LO	Frequency.BinLow
FREQ_HI	Frequency.BinHigh
FLUX	Flux
ERR_LO	Flux.ErrorLow
ERR_HI	Flux.ErrorHigh
QUALITY	Flux.Quality
SPEC_RES	Flux.Resolution

Note: The ENERG_LO, ENERG_HI columns are already used in the X-ray community.

We summarize this with a full sample FITS extension header.

```
XTENSION= 'BINTABLE'           / binary table extension
BITPIX  =                    8 / 8-bit bytes
NAXIS   =                    2 / 2-dimensional binary table
NAXIS1  =                   80 / width of table in bytes
NAXIS2  =                 2048 / number of rows in table
PCOUNT  =                  208 / size of special data area
GCOUNT  =                    1 / one data group (required keyword)
TFIELDS =                   14 / number of fields in each row
EXTNAME = 'SPECTRUM '           / name of this binary table extension
DATE    = '2004-08-30T14:18:17' / Date and time of file creation
VOCLASS = 'Spectrum'            / VO Data Model
TIMESYS = 'UTC      '           / Time system
SPECSYS = 'TOPOCENT'            / Wavelengths are as observed
MJDREF  =  0.0                 / MJD zero point for times
TELESCOP= 'MMT '          / Telescope
INSTRUME= 'BCS    '          / Instrument
FILTER  = 'G220    '           / Grating
OBJECT  = 'ARP 220 '           / Source name
OBSERVER= 'Johann Kepler '     / Principal investigator
VOPUB   = 'CfA Archive'        / VO Publisher authority
VOPUBID = 'ivoa://cfa.harvard.edu' / VO Publisher ID URI
VOCREATE= 'MMT Archive'       / VO Creator
VOLOGO  = 'http://cfa.harvard.edu/vo/cfalogo.jpg' / VO Creator logo
VERSION =  2                     / Reprocessed 2004 Aug
EQUINOX =  2.0000000000000E+03 / default
```

```
RADECSYS= 'ICRS    '           / default
CONTACT = 'Jonathan McDowell'  / Curation contact
EMAIL   = 'jcm@cfa.harvard.edu' / Curation contact email
SPEC_RES=              5.0  / [Angstrom] Spectral resolution
TTYPE1  = 'INSTRUME  '           / Instrument ID
TFORM1  = '8A      '           / format of field
TTYPE2  = 'FILTER  '           / Filter ID
TFORM2  = '8A      '           / Format
TTYPE2  = 'RA '                / Position RA of aperture center
TFORM2  = '1D      '           /
TUNIT2  = 'deg     '           /
TTYPE3  = 'DEC     '                / Position Dec of aperture center
TFORM3  = '1D      '           /
TUNIT3  = 'deg     '           /
TTYPE4  = 'APERTURE'           / Aperture diameter (physical or extraction)
TFORM4  = '1E       ' /
TUNIT4 =  'arcsec '   /
TTYPE5 = 'TIME'
TFORM5 = '1D '
TUNIT5 =  'd'  /  MJD days
TTYPE6 = 'EXPOSURE' /  Effective exposure time
TFORM6 = '1E'
TUNIT6 = 's'
TTYPE7 = 'SYS_ERR' / Fractional systematic error
TFORM7 = '1E'
TUNIT7 = 's'
TTYPE8 = 'WAVE'   / Wavelength
TFORM8 = '1PE'
TUNIT8 = 'Angstrom'
TTYPE9 = 'WAVE_LO' /
TFORM9 = '1PE'
TUNIT9 = 'Angstrom'
TTYPE10= 'WAVE_HI' /
TFORM10= '1PE'
TUNIT10= 'Angstrom'
TTYPE11 = 'FLUX' /
TFORM11= '1PE'
TUNIT11= 'erg cm^-2 s^-1 Angstrom^-1'
TTYPE12= 'ERR_LO' /
TFORM12 = '1PE'
TUNIT12 ='erg cm^-2 s^-1 Angstrom^-1'
TTYPE13= 'ERR_HI' /
TFORM13 = '1PE'
TUNIT13 ='erg cm^-2 s^-1 Angstrom^-1'
TTYPE14= 'QUALITY' /
```

```
TFORM14 = '1PI'
```

The data would look like

```
MMT/BCS  G300  233.73791  23.50333 2.0 52984.301203 1500.0 0.15  3200.0 3195.0 3205.0 1.48E-12
                                                                 3210.0 3205.0 3215.0 1.52E-12
                                                                 3220.0 3215.0 3225.0 0.38E-12
                                                                 3230.0 3225.0 3235.0 1.62E-12
MMT/BCS  G300  233.73792  23.50334 2.0 52102.103211 1480.0 0.15  3200.0 3195.0 3205.0 3.48E-12
                                                                 3210.0 3205.0 3215.0 2.52E-12
                                                                 3220.0 3215.0 3225.0 1.38E-12
                                                                 3230.0 3225.0 3235.0 1.62E-12
FLWO/4S  B     233.73791  23.50333 4.5 48776.001234  300.0 0.05  4400.0 4200.0 4600.0 1.82E-12
FLWO/4S  V     233.73791  23.50333 4.5 48776.012012  300.0 0.05  5400.0 5200.0 5600.0 3.82E-12
FLWO/4S  R     233.73791  23.50333 4.5 48776.019013  240.0 0.05  7000.0 6200.0 7500.0 5.82E-12
FLWO/4S  I     233.73791  23.50333 4.5 48776.024988  240.0 0.08  9000.0 8200.0 9900.0 8.12E-12
```

8.1 Alternate FITS serialization

Some implementors may find variable length arrays inconvenient. An alternate 'relational normalization' serialization, considered but currently rejected by the VO group, is recorded here for information. It would consist of two binary or ASCII tables, each with only scalar columns. The Spectrum table would contain the same scalar columns as above, but the variable array columns would be moved to a second table, called SpecData, and used as scalar columns arranged vertically. A new integer column "OBS_ID" would be added to each table; the values in the Spectrum table would be unique and the corresponding values in the SpecData table would indicate which data points belonged to which rows in the Spectrum table.

Another approach would be to have one FITS HDU per spectrum or photometry point. However this was rejected as unworkable, as the overhead of 5760 bytes (2 FITS blocks) per photometry point would inflate the data for the photometry-only SED case by factors of around 50-100.

9 VOTable Serialization

The VOTable version of Spectrum will represent an SED by a series of tables, one for each individual spectrum. The data model fields described above as arrays map to VOTable FIELDs, while the remaining fields map to PARAM.

We use the GROUP construct to delimit data model objects within the main object. Names of fields and parameters are left to the data provider. The utype and ucd attributes are used to denote data model and UCD tags.

```
<VOTABLE>
<RESOURCE>
<TABLE utype="Spectrum">   # Defines local namespace for utype
```

```
<PARAM utype="Instrument" ucd="inst.id">BCS</PARAM>
<PARAM utype="Filter" ucd="inst.filter.id">G300</PARAM>
<GROUP utype="Wavelength">
<FIELD utype="Value" ucd="em.wavelength" unit="Angstrom"/>
<FIELD utype="BinLow" ucd="stat.min,em.wavelength" unit="Angstrom"/>
<FIELD utype="BinHigh" ucd="stat.max,em.wavelength" unit="Angstrom"/>
<PARAM utype="Frame.System"  ucd="frame.em.system">Barycentric</PARAM>
</GROUP>
<GROUP utype="Coverage">
<GROUP utype="Position">
<PARAM utype="Value" ucd="pos.eq" unit="deg">132.4210, 12.1232</PARAM>
<PARAM utype="Frame.Equinox" ucd="time.equinox,pos.eq">2000.0</PARAM>
<PARAM utype="Frame.System" ucd="frame.pos.system">ICRS</PARAM>
</GROUP>
<PARAM utype="Aperture" ucd="pos.region.diameter" unit="arcsec">20</PARAM>
<GROUP utype="Time">
<PARAM utype="Value" ucd="time.obs">52148.3252</PARAM>
<PARAM utype="Frame.Type" ucd="frame.time.scale">UTC</PARAM>
<PARAM utype="Frame.Zero" ucd="frame.time.zero">0.0</PARAM>
</GROUP>
<PARAM utype="Exposure" ucd="time.interval,phot.spectrum" unit="s">1500.0</PARAM>
</GROUP>
<GROUP utype="Curation">
<PARAM utype="Publisher" ucd="human.publisher,meta.curation">SAO</PARAM>
<PARAM utype="PubID" ucd="meta.curation.pubid">ivoa://cfa.harvard.edu</PARAM>
<PARAM utype="Creator" ucd="meta.curation.creator">SAO/FLWO</PARAM>
<PARAM utype="Logo" ucd="meta.curation.logo">http://cfa-www.harvard.edu/nvo/cfalogo.jpg<
<PARAM utype="Date" ucd="time,soft.dataset,meta.curation">2003-12-31T14:00:02</PARAM>
<PARAM utype="Version" ucd="soft.dataset.version,meta.curation">1</PARAM>
<PARAM utype="Contact" ucd="human.contact,meta.curation">Jonathan McDowell</PARAM>
<PARAM utype="Contact.Email" ucd="meta.curation.email">jcm@cfa.harvard.edu</PARAM>
</GROUP>
<GROUP utype="Flux">
<FIELD utype="Value" ucd="phot.flux,em.wavelength" unit="erg cm^-2 s^-1 Angstrom^-1">
<FIELD utype="ErrorLow"  unit="erg cm^-2 s^-1 Angstrom^-1">
<FIELD utype="ErrorHigh" unit="erg cm^-2 s^-1 Angstrom^-1">
<FIELD utype="Quality">
# In this case Resolution is demoted from Field to Param since it is constant
<PARAM utype="Resolution" unit="Angstrom">14.2</PARAM>
<PARAM utype="SysError" unit="">0.05</PARAM>
</GROUP>
<TABDATA>
# Note slightly nonlinear wavelength solution
# Second row is upper limit
# Third row has quality mask set
```

```
<TR><TD>3200.0<TD>3195.0<TD>3205.0<TD>1.38E-12<TD>5.2E-14<TD>6.2E-14 <TD>0
<TR><TD>3210.5<TD>3205.0<TD>3216.0<TD>1.12E-12<TD>1.12E-12<TD>0<TD>0
<TR><TD>3222.0<TD>3216.0<TD>3228.0<TD>1.42E-12<TD>1.3E-14<TD>0.2E-14<TD>3
</TABDATA>
</TABLE>
</RESOURCE>
</VOTABLE>
```

References

Greisen, EW, Valdes F G, Calabretta M R and Allen S L 2003, in prep., www.aoc.nrao.edu/~egreisen/scs_112103.ps

Hanisch, R., (ed)., Resource Metadata for the VO 2003 Oct 16 Draft, Do Not Reference (hah!), www.ivoa.net/Documents/WD/ResMetadata/WD-RM-20031016.pdf