# The ESO science data product standard for 1D spectra

A.Micol, N. Delmotte, M. Arnaboldi, J. Retzlaff et al., ESO Data Management & Operations [DMO]

amicol@eso.org

**Goal**: To mange the submission, archiving, and publishing of high science level data products generated by various data providers, including External Data Products [**EDP**] by PIs of various ESO observing programs (public surveys, large programmes, and from willing PIs of GTO/GO programmes), from various instruments (VIRCAM, OMEGACAM, HAWK-I, EFOSC2, UVES, etc.), of different calibration levels (from science grade to advanced data products [ref. 1]), and Internal Data Products [**IDP**] from quality-controlled pipeline-processed data automatically-generated using the existing ESO pipelines, and doing all that while *minimising* the (archive) operational costs and *maximising* science exploitation.

At the root of the **solution**, in what is called ESO **Phase 3** (*the process of preparing, validating and ingesting science data products (SDPs) for storage in the ESO science archive facility, and subsequent data publication to the scientific community*) [ref. 2], is the definition of a common data format that all parties (PIs, ESO data flow, and (planned) the ESO pipelines) have to adhere to: the **ESO Science Data Product standard** [aka, **SDP**].

This poster describes the Spectroscopic part of the standard (1D spectra), its role within the ESO science archive facility [**SAF**], and its utilisation.

---

## Definition of the ESO science data product standard for 1D spectra

### Study phase

**Matrix comparison** of various ground- and space-based instrument spectral data formats with respect to different criteria (ability to support all ESO instruments, flexibility and future proofness, impact on ESO metadata harvester, on Phase 3 validation, on archiving, compatibility with existing data analysis and visualisation tools, VO compliance, etc).

Input received by all involved ESO operational groups (archive science group, quality control, science-grade data products, archive content handling)

### Main requirements

The contingent of existing and future ESO spectroscopic instrumentation offers quite a vast panorama of instrument modes (from optical to mid-IR and sub-millimeter, from single to multi-arm spectrographs, from single slit to MOS, etc). The 1D spectra generated by all those instruments must be supported by the selected data format.

One important aspect is the ability to convey in a single file all the necessary information that an astronomer needs when analysing the data. The spectral flux array, its associated error, sky background, data quality, and weight map arrays must be found all in the same file.

Some instrument modes cover large spectral ranges; varying spectral bin and resolution along a spectrum is quite a typical scenario. Some instruments (e.g., the red arm of UVES) split different wavelength regions into differently optimised detectors with different bin sizes. The ability to store such non-equally sampled spectra in a single array is of fundamental importance.

Willingness to follow the Virtual Observatory [VO] standard played a role.

### Selected format: FITS Binary Table

The FITS binary table format was found to best support all requirements for the 1D spectrum case.

The adopted format inherits most of the features of the FITS serialisation of the IVOA SpectrumDM 1.0 [ref. 3], adding to it the required Phase 3 keywords.

All important scientific parameters (e.g.: space/time/spectral coverage, spectral sampling/resolution, signal-to-noise ratio, etc) and all relevant Phase 3 and bookkeeping information (e.g.: provenance, product category, product level, etc) are stored in the headers of the FITS file, using standard keyword names, and standard formats and units for their values (see example).

The actual 1D spectral data are stored in the first FITS extension as a single row of a binary table. Each table cell contains one of the arrays: the first one always contains the wavelength array; any other cell in the table row hosts one of the data arrays [flux, error, sky, etc]. All arrays are of the same length. Variable length arrays are not permitted.

Given that the wavelength array is stored with the data, no WCS information is to be provided. The data arrays do not need to be equally-sampled.

ESO instrument calibration pipelines will soon support the SDP format.

---

### Incomplete example of a PRIMARY HEADER UNIT

```
        Product Category & Unique identifier (within release)
PRODCATG= 'SCIENCE.SPECTRUM'   / Data product category
OBJECT  = 'SN2012hr'           / Original target and unique identifier
                Spatial characterisation
RA      =          95.41025 / [deg] Spectroscopic target position (J2000.0)
DEC     =         -59.71406 / [deg] Spectroscopic target position (J2000.0)
                Temporal characterisation
EXPTIME =             720.0 / Total integration time per pixel (s)
MJD-OBS = 58242.3139210
        Wavelength solution and spectral characterisation
LAMNLIN =              12.0 / number of arc lines used for fit
LAMRMS  =            0.0133 / residual RMS [nm]
WAVELMAX=    1645.202167344094 / [nm] Maximum wavelength
WAVELMIN=    934.7408880472184 / [nm] Minimum wavelength
SPECSYS = 'TOPOCENT'         / Reference frame for spectral coordinate
SPEC_BIN= 0.6938098430633551 / Wavelength bin size [nm/pix]
SPEC_ERR= 0.003839379290111012 / statistical uncertainty
SPEC_RES=      544.8170230880778 / Spectral resolving power
SPEC_SYE=               0.0 / systematic error
                Flux/Spectrum characterisation
FLUXCAL = 'ABSOLUTE'           / type of flux calibration
FLUXERR =              34.7 / Fractional uncertainty of the flux [%]
TOT_FLUX=                 F / TRUE if phot cond and all src flux is captured
SNR     =    22.25848342228699 / Average signal to noise ratio per pixel
CONTNORM=                 F / spectrum normalized to the continuum
EXT_OBJ =                 F / TRUE if spectrum of extended object
                Associated ancillary files
ASSOC1  = 'ANCILLARY.2DSPECTRUM' / Category of associated file
ASSON1  = 'SN2012hr_20130113_GB_merge_56477_1_si.fits' / Name of associated file
                Provenance information (contributing files)
PROV1   = 'SOFI.2013-01-14T05:32:17.420.fits' / Originating file
[...]
PROV5   = 'SOFI.2013-01-14T05:46:30.248.fits' / Originating file
        Extra keywords not covered by the standard are admitted
HIERARCH ESO OBS NAME     = 'SN2012hr_BG' / OB name
```

### Incomplete example of a FITS EXTENSION HEADER

```
XTENSION= 'BINTABLE'           / binary table extension
...
NAXIS   =                    2 / number of array dimensions
NAXIS1  =                24576 / Length of data axis 1
NAXIS2  =                    1 / length of dimension 2
TFIELDS = 4                    / Number of cells
NELEM   =                 1024 / Length of the data arrays
TITLE   = '56306.249 SN2012hr GB GBF long_slit_1' / Dataset title
                Arrays Names and formats
TTYPE1  = 'WAVE    '           / Label for field 1
TTYPE2  = 'FLUX    '           / Label for field 2
TTYPE3  = 'ERR     '           / Label for field 3
TTYPE4  = 'SKYBACK '           / Label for field 4
TFORM1  = '1024E   '           / Data format of field 1
[...]
TFORM4  = '1024E   '           / Data format of field 4
                Units and VO descriptors of arrays
TUNIT1  = 'nanometers'         / Physical unit of field 1
TUNIT2  = 'erg cm**(-2) s**(-1) angstrom**(-1)' / Physical unit of field 2
[...]
TUCD1   = 'em.wl'              / UCD of field 1
TUCD2   = 'phot.flux.density;em.wl;src.net;meta.main' / UCD of field2
TUCD3   = 'stat.error;phot.flux.density,meta.main' / UCD of field3
TUCD4   = 'phot.flux.density;em.wl' / UCD of field4
TUTYP1  = 'Spectrum.Data.SpectralAxis.Value'
TUTYP2  = 'Spectrum.Data.FluxAxis.Value'
TUTYP3  = 'Spectrum.Data.FluxAxis.Accuracy.StatError'
TUTYP4  = 'Spectrum.Data.BackgroundModel.Value'
VOCLASS = 'SPECTRUM V1.0'      / VO Data Model
VOPUB   = 'ESO/SAF'            / VO Publishing Authority
CHECKSUM= 'CD6cDC4cCC4cCC4c'   / HDU checksum updated 2013-08-01T15:08:37
DATASUM = '4065290576'         / data unit checksum updated 2013-08-01T15:08:37
```

On the right an example of an SDP spectrum visualised through a VO tool.

On the left its (incomplete) header.
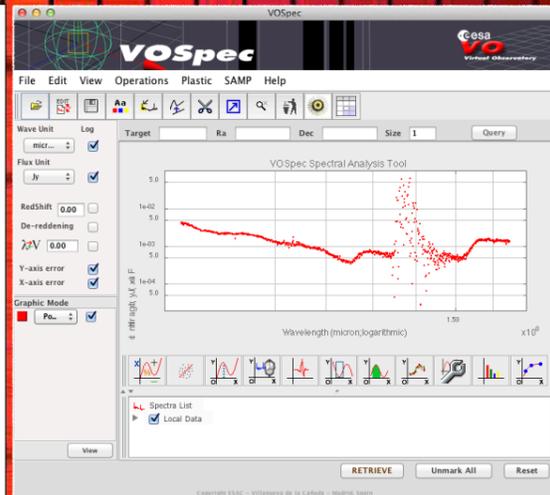
## References       p-poster o-oral

[1] R. Hanuschik: *"Phoenix: automatic science processing of ESO-VLT data"* - P

[2] N. Delmotte: *"ESO Phase 3 user support and operations"* - P

[3] J. McDowell: *"IVOA recommendation: spectrum data model 1.1"*

[4] M. Arnaboldi: *"The ESO public surveys: science goals, current status and policies"* - O

[5] J. Retzlaff: *"Data products in the ESO science archive"* - O

See also in this conference:

J. Retzlaff: *"Releasing ESO public survey data through the Phase 3 catalogue facility"* - P

---

### The Phase 3 process

The Phase 3 process handles the submission, validation, and archiving of data products (whether IDPs or EDPs). The main steps involved are:

• Uploading data products and release description to the Phase 3 staging area

• Formal validation of the uploaded data (automatic checks)

• Content validation of the uploaded data (by a Phase 3 scientist)

Once all validations are successful, and therefore the submitted data are proved fully compliant with the SDP standard, the archiving of the data begins:

• Data files are ingested in the archive storage system (on online hard-drives (NGAS))

• FITS headers are extracted and placed in a SYBASE IQ keyword repository

• Access control metadata are assigned to each data product, according to the ESO data policy (defining proprietary period, data accessibility, data visibility)

• Relevant metadata are published from the keyword repository into the SYBASE ASE database tables used by the archive query forms.

### Role of the SDP standard within Phase 3

It acts as a contract between the submitter and the SAF.

It guarantees that all metadata are transmitted to the archive with proper formats and units; among them for each product it conveys:

• product category

• space/time/spectral coverage with errors

• resolving power, bin size, aperture, exposure time, SNR, calibration level

• provenance: which observations or other products contributed to it

• associations: which other ancillary files are linked to it

• scientific associations: OBJECT acts as unique identifier within a survey allowing to find all scientific products related to it

It ensures homogeneity of all incoming data and metadata, no matter from which observer, instrument, calibration process they were obtained.

The archive query forms can rely on a homogenous set of metadata.

The archive users are exposed to a single data format when downloading 1D spectra.

### The 1D spectrum format in action

Phase 3 is currently receiving its first spectra from 2 ESO spectroscopic public surveys [ref. 4]:

GAIA-ESO (G.Gilmore, S.Randich) targeting > $10^5$ stars with FLAMES for an homogeneous overview of the distributions of kinematics and elemental abundances of the Mily Way. This first submission will provide around 6000 spectra.

PESSTO (S.Smartt) aiming to classify around 2000 supernovae, and to provide 150 supernovae with full spectroscopic time series coverage (~10 epochs). First submission: ~1000 spectra and ~1800 images in various bands.

Also, starting September 2013, UVES/ECHELLE spectra will be re-calibrated, quality-controlled, converted to the SDP format, and then archived via Phase 3 (expected ~100.000 products) [ref. 1].