Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

Survey among Spectral Data Providers and Consumers

M. Dolensky

European Southern Observatory, Karl-Schwarzschild-Straße 2, D-85748 Garching, Germany

D. Tody

National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801-0387

Abstract.
The IVOA Data Access Layer (DAL) working group is defining a Simple Spectral Access (SSA) protocol for accessing 1D spectra and SEDs. To help plan this effort a survey was conducted to provide use-cases to guide the design. This poster summarizes the results from the point of view of both the data provider and the data consumer.

The survey shows that a wide variety of spectroscopic data is readily available on-line. Data structures and qualitative aspects, however, vary substantially. A preliminary analysis shows that the definition of a common format for spectra is a big challenge. Therefore, the first step is to concentrate on the access protocol and then to work on the formal representation in a standard format based on a spectral data model.

1. Introduction

The survey consists of two parts. The first concerns spectral data providers, and the second software tools which might use the SSA protocol to remotely access spectra. Such software tools are referred to as data consumers in this paper. Below is a summary of the information gathered from ESA IDC & XMM archives, ESO/ST-ECF, HEASARC, Hyperleda, INTA/LAEFF, NAOJ/SMOKA, NCSA, NOAO and CFLIB project, OHP and SDSS.

This survey did not aim to be complete, rather the goal was to obtain a large enough sample to capture the major aspects to help design the SSA protocol and spectral data model.

2. Data Provider

2.1. Spectral Data Collections

The following spectral data collections were examined:

- Elodie: 25000 extracted Echelle spectra, resolution 42000, 4000 - 6800 Å on the fly processing options available: wavelength resampling, flux calibration etc., total volume 50 - 100 GB
- ESO/ST-ECF, 250000+ observations taken with 14 spectrographs in UV, optical, near IR; archive is rapidly growing
- HEASARC: Variety of spectral data sources from X-ray and gamma-ray missions, including both relatively high and low resolution spectral data
- Hyperleda: 10000 reference spectra of stars and galaxies, resolution 300 - 1000000, 1300 Å - 2.5 microns total volume 5 GB, collected from literature and reformatted
- Indo-U.S. Library of Coude Feed Stellar Spectra (NOAO & CFLIB) 1300 spectra, 3465-9469 Å @ 0.44Å/pix
- ISO 1D Spectral Data: 10000 files, 2.4-197 micron, resolution 40-30000
- IUE Newly-Extracted Spectra (INES): 110000 spectra, 9500 objects, 1150-1980 Å & 1850-3350 Å resolution 1.676 & 2.669 Å/pix
- NAOJ/SMOKA: 40000+ observations with 7 spectrographs in optical and IR including Echelle and multi-slit detectors, 150+ GB volume
- NCSA Astronomy Digital Image Library - NCSA BIMA Data Archive
- NOAO: various high resolution atlases and libraries of spectra mainly in the optical and IR
- Sloan Digital Sky Survey (SDSS): Data Release 1, 186.000 objects, 3800-9200 Å , 1300 square degrees, resolution 1800
- XMM-Newton 1D Spectral Data: XMM/RGS, 20000 spectra (expected), 0.35-2.5 keV, resolution 200-800; XMM/EPIC, 1.5 mio. spectra (expected), 0.35-15 keV, resolution 20-50

2.2. Characteristics of Data

The characteristics of spectral data present in the above collections may be summarized as follows:

- number: 100 K - 1 mio. spectra/data provider; often only several 100s in very specific catalogs
- size: few KBs - few MBs/spectrum
- wavelength/energy ranges: gamma, X-ray, UV, optical, IR, radio
- resolution: 20 - 1.000.000 (solar spectra)
- linear and non-linear/irregular sampling
- WCS information
- stellar parameters: e.g.: Teff, log(g), [Fe/H]
- noise
- masks (coded aperture mask data, INTEGRAL)
- observation dependent transfer matrix to convert counts to physical units
- spectral features (SLOAN): emission, absorption lines and parameters, emission redshift params. cross-correlation parameters
- time resolved spectral information (XTE, BATSE, Swift, HETE)

Columns/data items:

- wavelength (better fravergy: frequency, wavelength, energy)
- flux (magnitude, flux, flux density, counts)
- flux error
- quality flags

- variance arrays
- photon events
- antenna temperature vs. frequency (e.g. SWAS spectra)

2.3. Current Storage Format

Currently there is no widely used spectral data format, and each data collection tends to store spectra in a unique format. The range of formats used include the following:

- ASCII tables (catalogs, often highly processed data)
- Database tables (e.g. Sloan)
- FITS binary tables
- FITS images (often used for raw data (CCD images) or stacks thereof; radio spectral image cubes)

2.4. Is the Data Available Online?

Yes, data are generally on-line, but proprietary periods and restricted access depending on nationality may apply.

3. Data Consumers

3.1. Applications, Packages

Information about the following tools was collected: aXe, MIDAS-MOS, NOAO IRAF-Specplot, Pleinpot, Specview, ISO Survey Products Display Tool. Some tools dealing with SEDs are: NED on-line service, AVO Prototype SED utility.

3.2. Software Capabilities

The spectral analysis tools examined by the survey include the following capabilities:

- reading various instrument dependent FITS and ASCII formats
- overplotting of multiple arrays
- various display options (labels, zoom, colors, log/linear scaling, ...)
- physical unit conversion
- model fitting capability
- classification against standards
- rebinning
- quality filters

3.3. Desired Characteristics of Input Data

Respondents were asked to comment on the desired characteristics of spectral data to be read by an analysis tool. The following items were mentioned:

- As a minimum, data should consist of (x,y) pairs of wavelength-flux values expressed in wavelength/frequency/energy units, and spectral flux density units respectively. The units information should be present as well, in header keywords or table column descriptors or any other suitable form.

- Additional information that could be included with the above, (if available) is a third value associated with each (x,y) pair representing the measurement error in flux density.
- Uncalibrated data (such as in counts/s units) can also be ingested but not much else can be done with it besides plain plotting.
- FITS WCS; useful for well sampled rasters

3.4. Desired Input Data Formats

FITS (images, tables), XML, and plain ASCII are all supported by various tools. Some tools (e.g., Specview) support multiple input formats. It is fairly easy to deal with multiple input formats so long as they implement a similar data model. For quicklook purposes also graphic formats are used, like the ISO IDC poststamp service.

Acknowledgments. Many thanks to all who contributed to the survey. Special thanks to Ivo Busko, Mike Fitzpatrick, Satoshi Honda, Stephen Kent, Tom McGlynn, Pedro Osuna Alcalaya, Benoit Pirenne, Ray Plante, Phillipe Prugniel, Enrique Solano, Alex Szalay, Francisco Valdes and Andreas Wicenec.