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

Spectral Data Models for the Virtual Observatory

Jonathan C. McDowell, Stephen Lowe, Mark Cresitello-Dittmar, Janet DePonte Evans, Ian Evans, Arnold Rots, Michael Harris

Smithsonian Astrophysical Observatory, 60 Garden St, Cambridge, MA 02138

Abstract. There is no standard way in astronomy to represent digital spectroscopic data. We present requirements for a standardized 1-dimensional spectral data model for use in the Virtual Observatory. We discuss the different kinds of spectra and the different observables used, as well as the appropriate instrumental calibrations. Our model is intended as a special case of an n-dimensional model for image and spectral data, and would incorporate the FITS spectral WCS proposal for coordinate descriptions. The problem of describing spectral data is closely related to the problems of defining bandpasses and photometric calibrations, as well as of abstract instrument descriptions such as spectral responses and efficiencies.

1. Introduction

A significant fraction of the public data available to the astronomical community is in the form of spectra. Although most archives store image metadata in fairly standard ways the history of archiving spectral data is much less successful. A recent survey of spectral archives (see Tody on www.ivoa.net/forum/dal) which revealed a heterogeneous collection of formats, many in ASCII tables, FITS tables, or FITS images. This is in contrast to simple sky images which, despite problems with how to represent mosaics, are mostly in some variation of FITS image extensions. The current FITS WCS proposal (Greisen et al.) for wavelength transformations is only one of the steps needed to use archived spectra interoperably. The VO will need to specify a uniform way to describe spectra.

This study attempts to isolate the metadata needed for representing spectra to the VO, and proposes ways to structure this metadata. Our model separates metadata needed by applications using the idealized, generalized spectrum (pixel values, coordinates, errors, units, resolution) from metadata describing the idealized observation (sky region, observation date) and from metadata which are needed by specialized applications which deal with particular observational strategies (e.g. spectral extraction details).

2. What is a spectrum?

A spectrum is the value of an observable as a function of a spectral coordinate, corrected or not for various instrumental effects. The spectral survey confirms that existing public data use the full range of possible parameters for the electromagnetic spectrum; (see Greisen et al. 2004 for spectral coordinates in FITS).

We distinguish between a theoretical spectrum sense, the energy output versus e.g. frequency, and a spectral dataset in the observer's sense, which maps such a spectrum onto an instrument. Spectral datasets often have degeneracy when two celestial axes and one spectral coordinate are projected onto two instrument coordinates. Here we describe spectra (the idealized $F(\nu)$) rather than spectral datasets.

The 1-D spectrum is clearly a special case of a 1-D histogram, and our final VO scheme should unify common metadata with other 1-D histograms (e.g. lightcurves) and with n-dimensional generalizations such as the 2-D image. This case study will be used to ensure that the n-D observation model can encompass everything we need to represent a spectrum.

We can identify several other kinds of 'spectrum':

- Other observables as a function of wavelength: percentage polarization, extinction coefficient. These can use the present model.

- Arrays of spectra such as spectral-spatial data cubes. They are a simple extension if we model spatial images compatibly.

- Spectral coordinates for particles other than photons: massless (gravitational waves) or massive (electron energies in a jet, cosmic ray spectrum).

- Spectral coordinates not a particle property: power spectra of source variability or CMB anisotropies, Fourier transforms generally. Needs a slightly different model.

Observable	Typical unit
Energy flux Density vs λ	erg cm^{-2} s^{-1} Å^{-1}
Energy flux Density vs ν	Jy
Energy flux Density vs $\log \nu$ (for SED)	Jy Hz
Photon flux density vs Energy	photon cm^{-2} s^{-1} keV^{-1}
Luminosity (at source)	erg s^{-1} Å^{-1}
Luminosity per decade	$L_\odot$
Radiation energy density	erg cm^{-3} Hz^{-1}
Flux per solid angle (e.g. at source surface)	erg cm^{-2} s^{-1} Å^{-1} sr^{-1}
Antenna temperature	K
Brightness temperature	K
Magnitude in given band	mag
AB magnitude	mag
Surface brightness flux density	Jy / arcsec^2
Flux per resolution element	Jy / beam
Surface brightness mag.	mag / arcsec^2
Instrumental reading	ADU, count
Ratio of two spectra	Dimensionless

Table 1. An incomplete list of spectral observables

3. Observables

A crucial task for the VO is to standardize how data providers describe the observable. What do the pixel values represent? At the moment, if you are lucky there is a BUNIT keyword in a FITS image to at least tell you the unit, but that is not really sufficient. The VO will use tags such as Uniform Content Descriptors

(UCD2, Derriere et al. 2004) to unambiguously characterize the physical concept being measured. Our spectral data model must define a standard place to store this metadata. Examples of spectral observables are listed in table 1.

4. A Partial Model

There are three main parts of our model: the dataset description, (Fig. 1), the data container description (Fig. 2), and the observation coverage description (not presented here). The complete Dataset object, a simplified version of the one presented in Cresitello-Dittmar et al. (2004), contains curation and coverage objects as well as several Data Container objects. The dataset will have at least one Data Container for the main data, and may have additional ones for a background spectrum, an exposure array, and a sensitivity array.

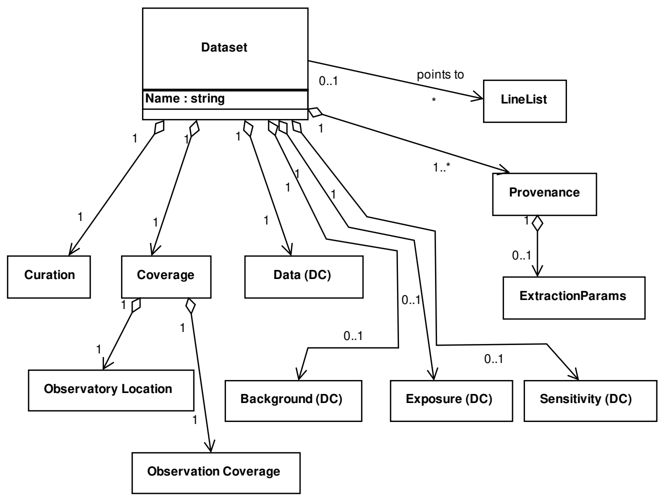

Figure 1. UML class diagram for the Dataset model.

The Data Container has a Data Storage object containing Value, Error, Quality and Resolution sub-objects. Our abstraction is that the data consists of an ordered array of values (accessed by the Index object) which may be coupled to one or more PixelMap objects locating each value in a coordinate system (Cresitello-Dittmar et al. 2004). In the spectral case, the PixelMap would provide a bijection between pixel number and the spectral coordinate.

A simple case of such a map is a set of regularly spaced, contiguous wavelength bins. However, our abstraction also supports irregular or sparse arrays.

One may in general obtain value, error, quality and resolution numbers for each pixel, although in many cases things like the resolution may be constant for all pixels; the four separate objects, accessed using the Index, hide this implementation detail.

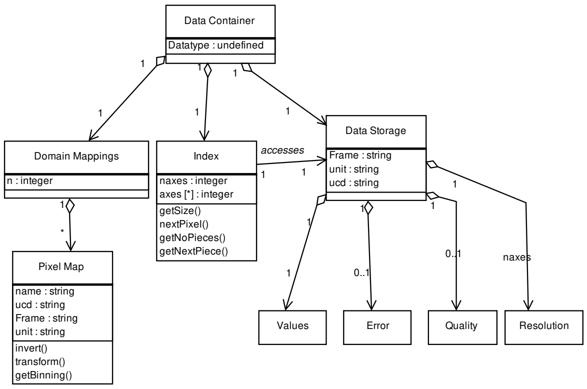

Figure 2. UML class diagram for the Data Container model.

5. Remaining Design Issues

The observable is declared with a UCD, this needs to be eloborated to fully model a Photometric System object.

The resolution is grouped within the Data Container together with values and errors, but the resolution object should ideally be a line spread function at each pixel. In contrast, the sensitivity, exposure and background are treated as separate data containers for two reasons: firstly, their effects are considered to be calibrated out, and accounted for in the error object; and secondly, they often have their own error, quality and resolution information different from the main data - although we should require them to have compatible pixel maps in some sense. Alternative choices would be to include all these arrays in a single Data Storage object, or at the other extreme to consider them as separate but associated Dataset objects and replicate all the observation information.

The sensitivity and exposure require care when we extend the model to a 3D energy-position cube, where practical implementations are likely to express things separably as, e.g., an on-axis energy sensitivity and a spatial sensitivity map.

We need to add appropriate UCDs in the observation description for specifying that a spectrum is in the rest frame and corrected for Milky Way but not intergalactic absorption, or corrected for detector QE but not vignetting.

Acknowledgments. We acknowledge support from NSF grant no. AST 0121296 and Cooperative Agreement AST 0122449, as well as the Chandra X-ray Center under NASA contract NAS8-39073.

References

Cresitello-Dittmar, M. et al. 2004, this volume, 277

Derriere et al. 2004, this volume, 315

Greisen et al., 2004, in prep.