The Simple Spectrum Access Protocol

an Activity of the IVOA Data Access Layer (DAL) Working Group

Markus Dolensky[1], Doug Tody[2]

[1]European Southern Observatory
[2]National Radio Astronomy Observatory

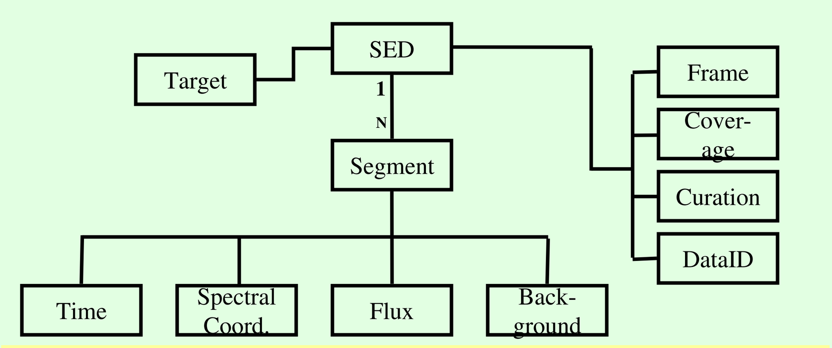

Figure 1. Simplified data model describing the structure of spectrophotometric datasets with spectral and temporal coordinates and associated metadata. It was developed by the data model working group of IVOA by McDowell et al.

2. Data Model

The model for spectral energy distributions (Fig. 1) defines a set of spectra or time series, some of which may have only one or few data points (photometry) and each of which may have different contextual metadata like aperture, position, etc.. Specifically, a spectrum has arrays for the spectral coordinate like the wavelength, time coordinate, flux value, background, and spectral resolution. There are provisions for expressing the accuracy and errors in a number of additional structures which are not described in this text.

An *SED* object has a number of global attributes indicating the number of SED segments and *curation* information. An *SED* object has exactly one *target* which identifies either an astronomical source or something different like a calibration or the fact that this is a model rather than a real observation.

A *segment* can either be a spectrum or a time series. Each *segment* has a *frame*, *coverage*, *curation* and *data identifier* object.

The *frame* object is a simplified instance of the space-time coordinate system object (A. Rots). It describes the spectral coordinate and time system. The *coverage* object holds info about the observed region on the sky, the time range and spectral range. The *curation* object carries information about the creator and publisher. The *data identification* model gives a dataset ID and its membership of larger collections.

The *time coordinate* contains elapsed times relative to a reference time. The *spectral coordinate* can be expressed as a wavelength, frequency or energy plus velocity. Similarly there is a range of ways to express *flux*: flux density, surface brightness, luminosity, photometric magnitude etc. The *background* represents a model for the expected flux values if the target had zero flux. Often, it will be generated by taking a flux measurement at another location and rescaling it for any difference in exposure. The *time*, *spectral coordinate*, *flux* and *background* are vectors which can form the columns a single output table.

1. Abstract

The goal of the Simple Spectral Access (SSA) specification is to define a uniform interface to spectral data including spectral energy distributions (SEDs), 1D spectra, and time series data. In contrast to 2D images, spectra are stored in a wide variety of formats and there is no widely used standard in astronomy for representing spectral data. Hence part of the challenge of specifying SSA was defining a general spectrophotometric data model as well as definitions of standard serializations in a variety of data formats including XML and FITS. Access is provided to both "atlas" (pre-computed) data and to "virtual" data which is computed on demand. The term simple in Simple Spectrum Access refers to the design goal of simplicity in both implementing spectral data services and in retrieving spectroscopic data from distributed data collections. SSA is a product of the data access layer (DAL) working group of the International Virtual Observatory Alliance (IVOA). The requirements were derived from a survey among spectral data providers and data consumers and were further refined in a broad discussion in meetings and electronic forums as well as by prototyping efforts within the European Astrophysical Virtual Observatory (AVO, fig. 4) and the US National Virtual Observatory (NVO).

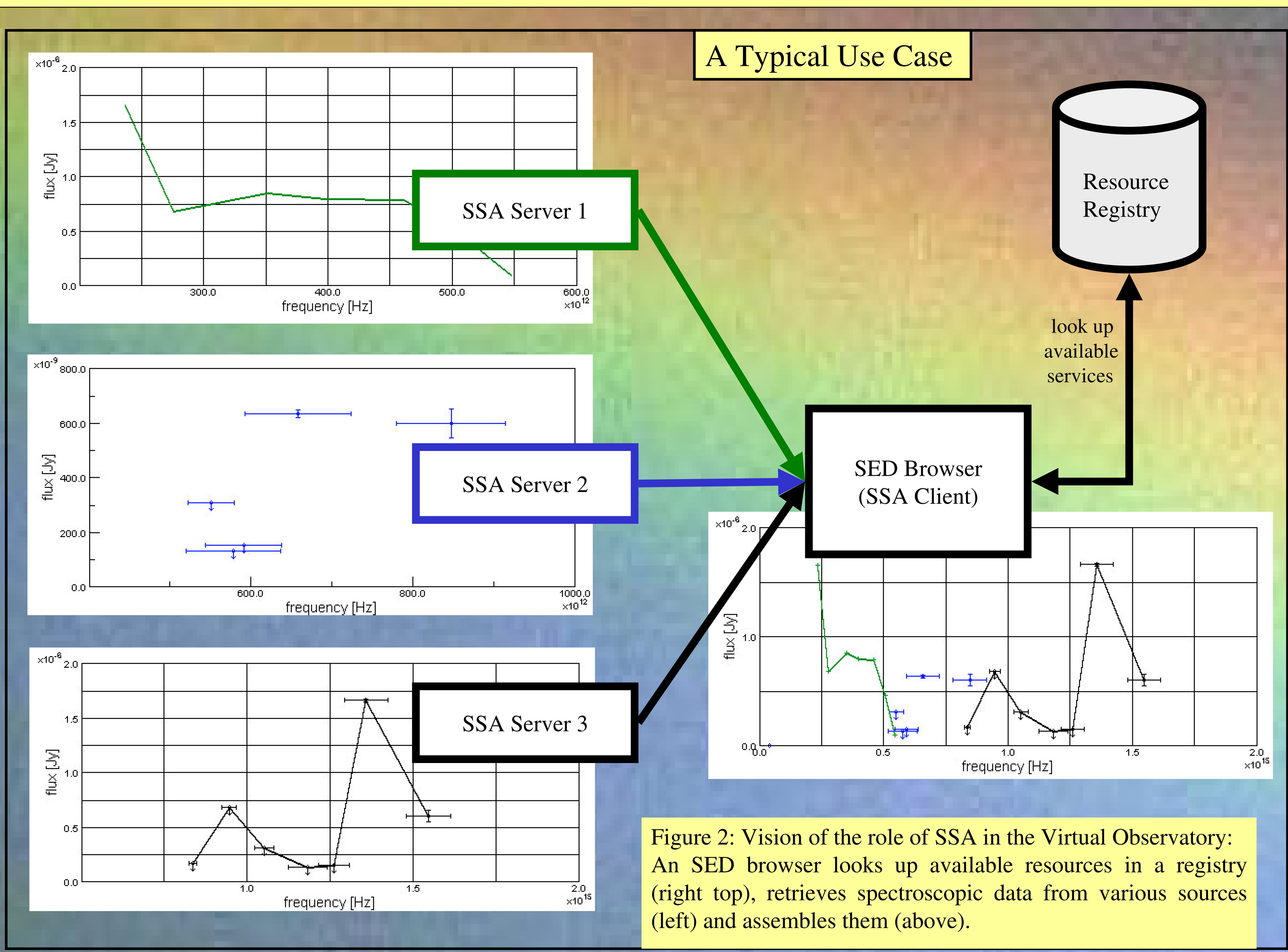

Figure 2: Vision of the role of SSA in the Virtual Observatory: An SED browser looks up available resources in a registry (right top), retrieves spectroscopic data from various sources (left) and assembles them (above).

4. Data Retrieval

The retrieval mode allows a client to retrieve an SED consisting of chunks of spectra or time series data. It does that with the given access reference as returned in the query mode. The access reference can be a simple URL but for instance a SOAP based implementation is possible as well. The format of the data returned in the retrieval mode corresponds to the mime-type previously specified in the query mode. It could be a VOTable, FITS, native XML, CVS, a graphic file or some *foreign* format used by a data provider.

5. Service Metadata & Registry

Compliant SSA services describe themselves in two ways. Firstly, services are registered with a VO registry. A registry stores service metadata which characterize it and any associated data collections. Secondly, capabilities are described through support of the special metadata query mode signaled through the *FORMAT=METADATA* parameter. When the service provider registers an SSA service, the registry can execute the metadata query to collect the capability metadata. The gathering of service and capability metadata from all such services enables a client to use the registry to discover the services most appropriate for a particular computation. Such registries are currently set up and are - to some extent - already functional.

6. Outlook

The SSA specification will stabilize within the coming months and is expected to become an IVOA recommendation later this year. It will allow VO projects to implement demonstrator tools and services by the end of 2004. The goal is to use those prototypes for complex scientific scenarios which various projects are planning for early 2005. In the longer term all DAL services including SSA will be integrated in the astronomical data query language (ADQL) which is currently under development within the IVOA.

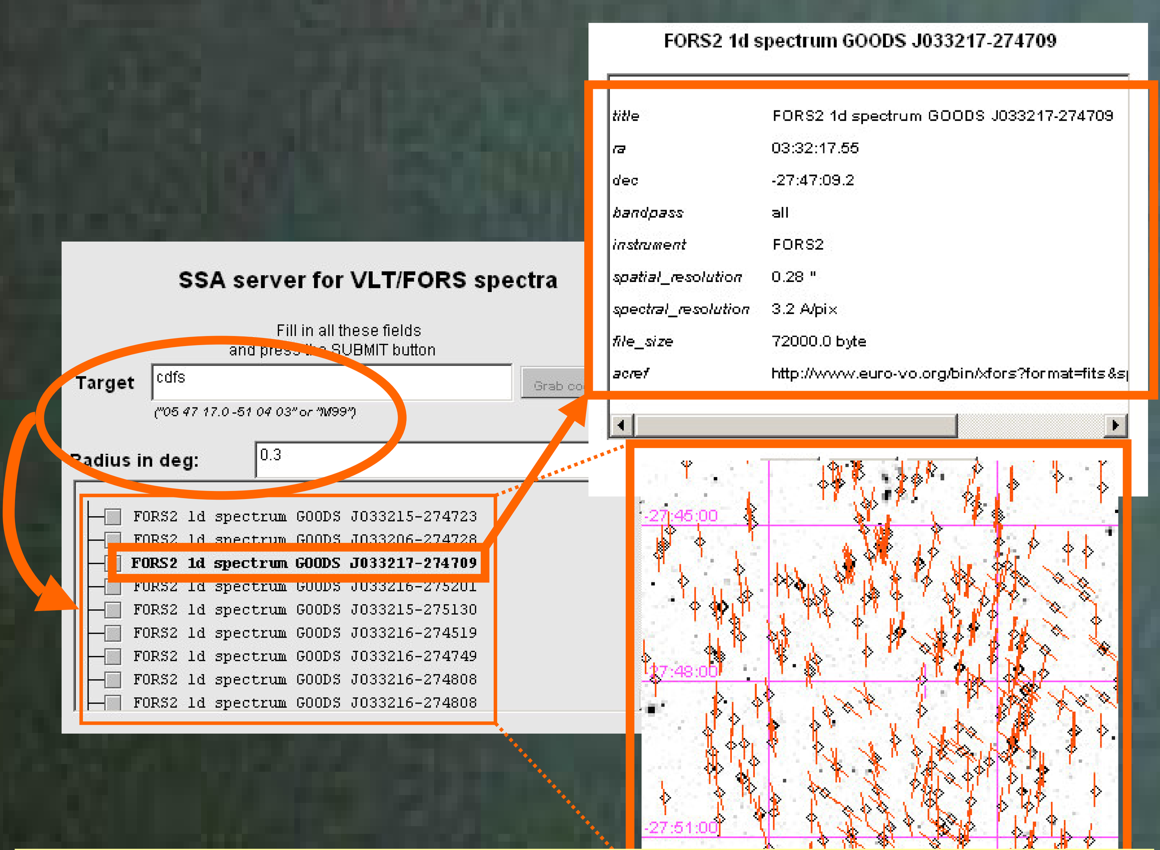

Figure 4. An SSA query interface is embedded in the AVO prototype toolset. In this example a search region is specified (top left), and the matching spectra are listed in tabular form. Detailed information of a single record is displayed (top right) as well as a graphical representation of matching spectra, their location and orientation (bottom right).

Acknowledgments

The Astrophysical Virtual Observatory has been awarded financial support by the European Commission through contract HPRI-CT-2001-50030 under the 5th Framework Programme for research, technological development and demonstration activities. The National Virtual Observatory is supported by the National Science Foundation's Information Technology Research.

Many thanks to all who contributed to the DAL specification and to the survey among spectral data providers and consumers. We thank Ivo Busko from STScI for adapting SpecView for its use as a graphical front end to the SSA prototype services and the ESA team of the ISO data archive for building an SSA prototype interface for the whole ISO archive.

3. Query Interface

The purpose of a spectrum query is to determine the availability and characterization of data satisfying the constraints. The result is encoded as a VOTable document. Queries can be restricted to certain types of data using the keywords *findSED*, *findSpectrum*, *findTimeSeries*. *FindSED* can find anything; the others only find Spectra or TimeSeries. This specification does not recommend a particular protocol. An implementation may utilize HTTP GET, POST or SOAP; all with the same parameters.

3.1 Required Query Parameters: A service must support these

- •pos ra,dec; IRCS
- •size diameter of aperture or search region
- •format all; graphic; metadata; provider; text/plain, application/x-votable+xml, application/fits, etc.

3.2 Optional Query Parameters: A service may ignore them

- •time lower, upper (ISO notation)
- •bandpass ID or numerical bandpass in meters
- •spectral resolution FWHM of line spread function
- •collection e.g., survey name
- •id dataset ID
- •aperture aperture size for computed or virtual spectra in degrees (default: size)
- •verbosity query response verbosity
- •score all (default), N top items

Further parameters are under consideration, including a generic parameter for specifying query parameters by Uniform Content Descriptor (UCD) or name, and a general keyword search capability. It is possible to extend a server by adding service-specific parameters. Parameter names, however, must not match any of the reserved parameter names defined above.

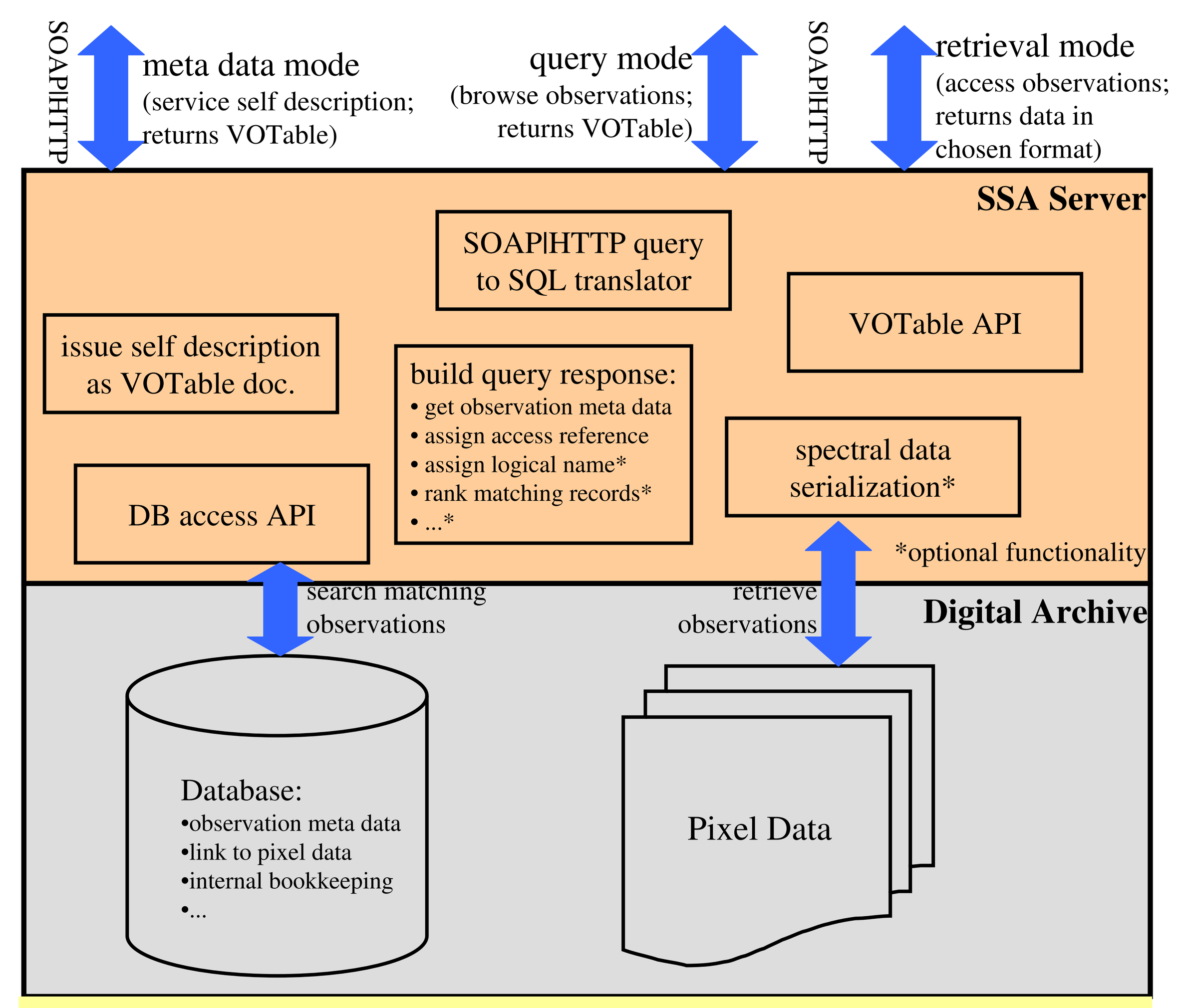

Figure 3: Example of a possible implementation of an SSA server. Its functional components (top) are sitting on top of a digital collection of spectroscopic data (bottom).

3.3 Query Response (cont.)

The response to an SSA query is a VOTable. Objects appear in the VOTable query response as grouped sets of fields with a special attribute that links it to a specific interface or element of the SSA data model.

The structure of the returned data consists of SED segments. Each segment is encoded as a table. Each segment has a type which is either spectrum or time series. Some meta information is global for all segments, like the target object. Other information is local to each segment like, for instance, the description of the coverage of the contained observation or curation info.

3.4 Mandatory Response Columns

A service must populate the following columns in the output:

- •Format dataset format, mime-type (e.g. application/fits, image/jpeg)
- •Acref access reference URL
- •SED object* SED object metadata
- •Dataset object* various fields giving details about a dataset in a collection
- •Coverage object* various fields describing the spatial, temporal and spectral coverage

*Objects are not further described in this poster. Any attributes of these objects can either be serialized as a single value or a vector of values.

3.5 Optional Response Columns

- •Score highest ranking candidates are best match
- •Target object target metadata
- •Npts number of points in spectrum or time series
- •Spatial.error, Spatial.resolution, Time.resolution, SpectralCoord.type, SpectralCoord.resolution, Flux.type