The VO-compatible Spectra Archives for Small Observatories

Petr Škoda
Astronomical Institute of the Academy of Sciences of the Czech Republic
Fričova 298, 25165 Ondřejov
skoda@sunstel.asu.cas.cz

Abstract
We present a survey of several server applications supporting preliminary versions of SSAP and give the overview of their capabilities and particular data format requirements, stressing their strengths and deficiencies. Finally we describe a small archive used for local SSAP access of spectra secured by Ondřejov Observatory 2m telescope.

Introduction

Although the VO Registry already contains hundreds of VO-accessible data resources, there are only few archives with optical spectra from ground-based telescopes. The majority of spectra available in Virtual Observatory comes from giant dedicated surveys like SDSS or archives of astronomical satellites. It is so because the design of VO-compatible data archive is an integral part of planning every large astronomical facility since the beginning and the project staff includes experienced SW developers keeping pace with current VO technologies and protocols. Small observatories, however, have still a very low awareness of VO principles and goals and they do not see any benefit of publishing their spectra in public accessible archives. Quite often they impose unreasonably long proprietary period on their spectra or do not want to publish the spectra at all. Unfortunately those, who want to publish their spectra in VO do not usually have enough developing capacity and VO-aware SW developers. The VO-enabled tools can increase the research efficiency by allowing an easy preview and visual comparison of a large number of spectra, without the need of tedious archive searching, downloading, transformation of various data formats and conversion to common physical units. Most evolved clients like VOSpec, SpecView or SPLAT support basic spectra analysis like continuum normalization, measurement of radial velocities and equivalent widths or identification of spectral lines. The effective utilization of these clients requires the server supporting Simple Spectra Access Protocol (SSAP) to be installed in the institution at the close proximity of source datafiles, getting the metadata mostly from local SQL database of FITS headers. The possibility of restricted access to their own spectra only for the local staff will be welcome by many institutions, who are afraid of releasing their spectra into public use.

An Ideal VO-compatible Server of Ground-based Optical Spectra

The majority of current astronomical spectra servers (based on Perl, PHP or Python CGI technology) present a HTML form translated into corresponding SQL query of the database of FITS header keywords. The result is presented as a list of matching objects and usually there are HTML links to more detailed information and to spectra preview (statical or generated on-the-fly).
The VO-compatible server is constructed in a different way. Its basic function is only to deliver the VOTable with embedded pointers to spectra selected using the SSAP (Simple Spectra Access Protocol) query given the target coordinates and searching radius.
Although VO supports the spectra in various output formats, the most typical requirement of ground-based spectra archive will be the delivery of spectra in original FITS files as obtained after the reduction.
The rest of spectra handling is done by the VO-compatible client. It should download the spectrum, display it and allow basic analysis on it (e.g. fitting the continuum, measurement of central line position and equivalent widths). Moreover, the positions of spectral lines of selected element can be overplotted in displayed region using the queries to theoretical databases by SLAP (Simple Line Access Protocol).

Survey of SSAP Servers

There were several server tools presented at EURO-VO publishing workshop at ESAC in June 2007, some of them supporting the preliminary versions of SSAP and thus allowing to build a simple VO-compatible spectra archive. All of them are using open-source database management system (MySQL or PostgreSQL) to store the data and metadata in database tables including the whole path to the spectra itself (mostly in FITS files). The system consists of at least two modules. The **Database Ingestor** extracts the values of relevant keywords from FITS headers and stores them into database tables as well as URL where the spectrum file can be downloaded. The **SSAP Server** (usually combined with SIAP and catalogue server) is then accepting SSAP requests in part of service URL and returns the VOTable with embedded links. Below we give a short overview of every presented server toolkit as well as the PLEINPOT system which was decided to become a heart of a small archive of spectra secured by coude spectrograph of Ondřejov Observatory 2m Telescope.

Metadata Extractor (MEX)

Developed at ESO as the base of future VO archive of data acquired by ESO instruments. It consists of interactive Mapping Editor and Metadata Extractor itself.
Mapping Editor uses the interactive web form (Tomcat servlet) for mapping required metadata from Data Model to corresponding FITS keywords. The mapping may be even quite complex – e.g. values of metadata can be calculated from existing FITS keywords. It requires the text file describing all Data Model items and their usual FITS keywords and isolated FITS header representing the particular dataset.
The result is a mapping file according to which the Metadata Extractor itself ingests values of FITS keywords and computed metadata into database system.
SSA server is driven by Python CGI script *retrieveSpectrum*.
The system is a combination of Python, Java 1.4. and 1.5 tools, scripts and SQL commands. Requires both Tomcat and Apache servers and MySQL database.

SAADA

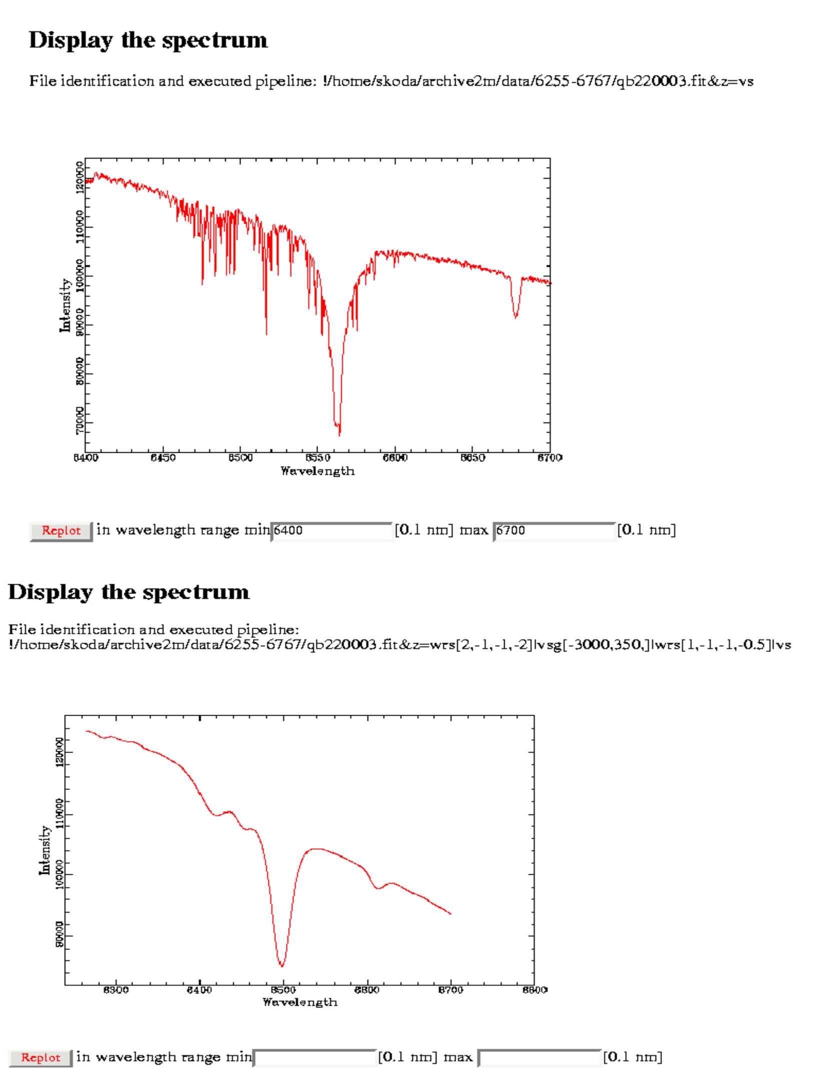
Developed in Strasbourg as a general system for creation of web-based databases of a collection of astronomical data files, including spectra. It is called the Astronomical Database Generator. For its efficient usage no programming is needed.
It consist of several jar files. *Saada.jar* is the installer. In addition there are two main java programs with clickable GUI. One is for creating new database (*newDB*) and the second (*saadadbgui*) extracts FITS keywords (including some form of mapping and deriving virtual keywords). Deployed Tomcat servlet has nice interface for data access (web form for querying and browsing databases).
Saada has its own query language SaadaQL, can create different sets of collections (images, catalogues, spectra of different origin).
It uses PosgreSQL database and requires java and Tomcat
The advantage is a detailed documentation with examples and ready-to-run demo showing all capabilities.
For SSA and SIA requires the externally edited configuration XML file.
It does not have spectra previewer but external tools may be configured to provide this capability (Aladin, VOSpec, SPLAT).

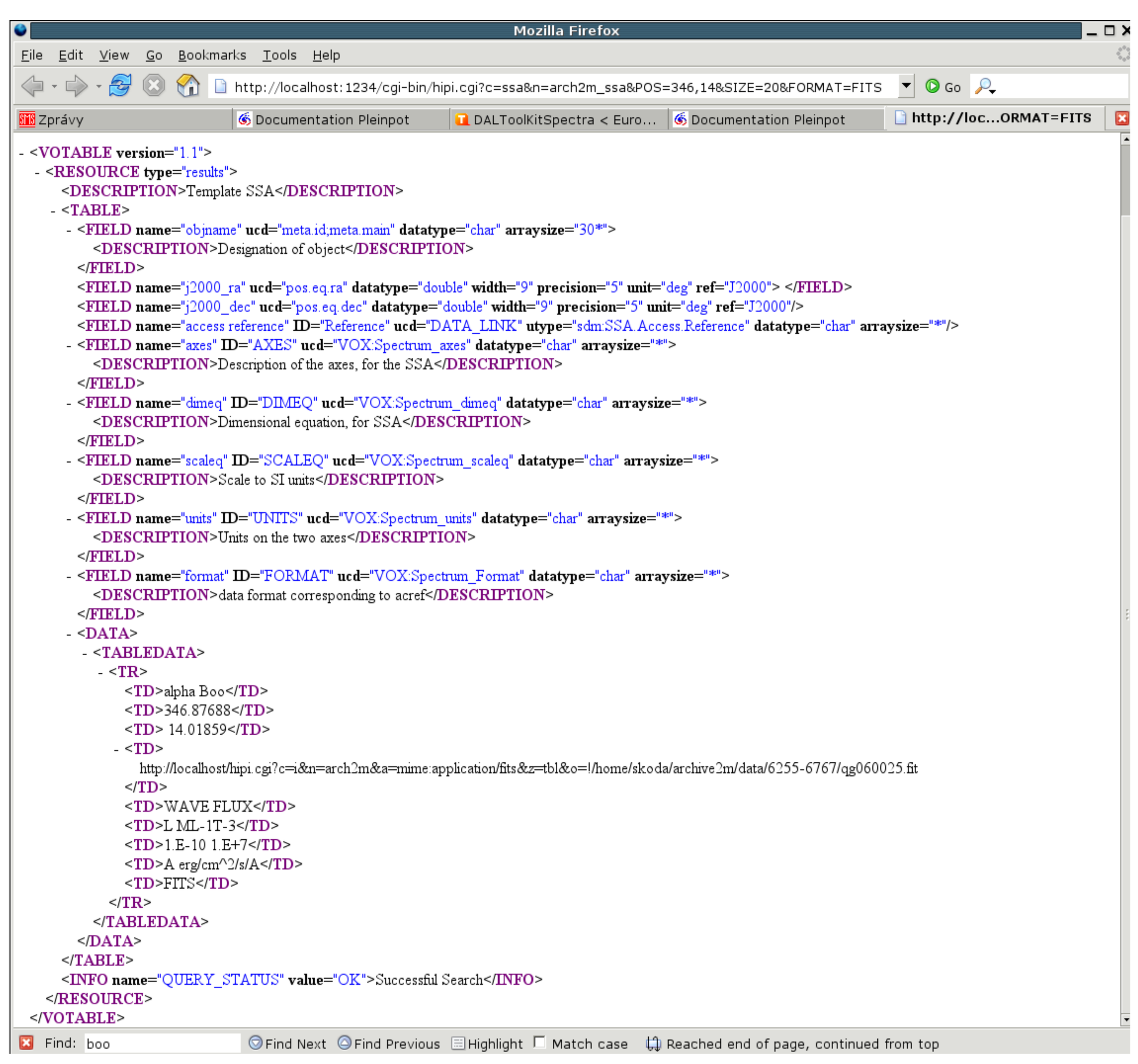
DALTOOLKIT

Developed in ESAC as the basis for current ESA ISO and XMM-NEWTON Satellite archive. It is programmed strictly according to VO standards in JAVA, using Tomcat servlet and MySQL database.
It consists of a simple database ingestor called *DALIngestor* and the SSA server (currently called *XMSSAToolKit*)
Its serious limitation is the acceptation of only FITS files with binary tables. The mapping process strictly requires names of binary table columns to be given.
It does not have spectra previewer and has only simple web-based parameter input form.
Because of its binary table FITS requirements it is currently unusable for most spectra archives of ground-based spectrographs, which are mostly in FITS 1D image format using CRVAL1 and CDELT1 keywords.

PLEINPOT

Developed in CRAL (Lyon) by Philippe Prugniel for HYPERLEDA galaxy database. It is in heart of ELODIE and SOPHIE archives.
Pleinpot represents a comprehensive set of general astronomical libraries, reduction and processing pipelines, CGI scripts, graphics libraries and database scripts. Mostly written in Fortran77 and C. Communicates with the PostgreSQL database.
The whole archive is accessed by a single CGI script called *hipi.cgi* . Most parameters are given in CGI URL (HTTP GET) string. There is a sophisticated (but rather cryptic) system of single-letter keywords triggering different pipelines like spectrum display with interactive form or even spectra processing like RV shift, rotational broadening or telluric lines removal. The results of query may be delivered in many output types (e.g. simple text, HTML document or VOTable).
Very interesting feature is that the system does not need data to be stored locally, but using general URL can access any files in the Internet and build the database for them. So the PLEINPOT may be used to create alternative database-driven archive for the simple collection of files exposed on Internet. But it also means that PLEINPOT allows easily to build a VO-compatible archive (using its SSAP and SIAP metadata tables) calling existing classical web archives and its content without the need of their modification. So it may be work as a kind of VO-compatible middleware.
There is also a mechanism of dynamically loadable modules useful for seamless addition of new protocols, data formats (e.g. directly serving MIDAS bdf files) and processing capabilities (new filters, on-the-fly reprocessing of data etc.).
The flexibility of the systems may be demonstrated on the ELODIE archive. It is a archive of separate extracted echelle orders with different dispersion in each. The archive data may be then either accessed as it is , or the hidden pipeline may be called to make correction of residual blaze function, rebin in wavelength and finally merge the individual orders into one single FITS file that is finally downloaded.
The FITS header ingestion for SSAP access is made according to the manually edited mapping file the SSA query (currently only 0.9 using several simple VOX_ keywords) is created using the metadata table (so can be easily adapted).
The system has nice documentation describing all functions, parameters etc with examples and tutorials. Installation is base on autoconf and make. There is a number of other features needed for HYPERLEDA (coordinates conversion, image transformations, cutouts, profiles, etc .) but this is not of great interest for building archive of spectra.
The example images show the output of the previewer before and after transforming the data (RV shift of -3000 km/s followed by rotational broadening of 350 km/s) and the output output VOTable as a response of a SSA query. The original spectrum is in addition zoomed in wavelength to given range.





Conclusions

All the VO-compatible spectra servers described have different intended usage driven by the requirements of particular community in which the SW has been developed. The DAL toolkit is (in current form) considered to be a reference implementation of VO standards. Its capabilities are thus quite limited. The best tool for building the general purpose spectra archive seems to be PLEINPOT.

Acknowledgements
This work was supported by grant GACR 205/06/0584 and EURO-VO DCA WP6..
The Astronomical Institute Ondřejov is supported by project Z10030501.
The author thanks to Phillipe Prugniel for the in-depth explanation of Pleinpot and all the lecturers of ESA VO publishing workshop for the patience and help with installation of described products on author's computer..

Presented at ADASS 2007, Kensington Town Hall, London, UK, 23-26 September 2007