The Virtual Observatory
Data, Standards and Tools
a technical - user point of view

Christian Surace and the CeSAM team

CeSAM, Laboratoire Astrophysique de Marseille, OAMP,CNRS,
38 rue Joliot Curie, F-13388 Marseille cedex 13, France
email: `christian.surace@oamp.fr`

Abstract. In 2002 the International Virtual Observatory Alliance (IVOA) has been created in order to gather efforts on data standardization and dissemination. Since then, the virtual Observatory allowed to spread validated data all over the world and to use data from everywhere from earth. From the standards definitions to development of tools, developers have set up a technical infrastructure used by astronomers to easily search for data and make science with all available products, more tools and more confidence on the quality of data. The goal of this review is to present the state of the art of the VO data, standards and tools. This review focuses on basic astronomer's questions : what kind of data are accessible, how to deal with these data and how to use them.

Keywords. methods: data analysis, astronomical data bases: miscellaneous, catalogs, surveys, Virtual Observatory

1. Introduction

During the past years the Virtual Observatory (VO) has focused its efforts on definitions and set up of the overall infrastructure to work easily with distributed astrophysical data. Based on this infrastructure, the data are now more accessible, and usable thanks to specific tools that have been developed. The Virtual Observatory is entering a new era, and is more focused on the scientific aspects and the interests of the astronomers. It is really a new age with new discovery technics for astronomy. Indeed, while some years ago, it was easy to analyze individual spectrum, Large detectors and Multi Object Spectroscopy brought thousands of data to analyse, and future surveys will provide Petabytes of data to be analyzed throughout automatic algorithms. Due to this still growing amount of data, it is very difficult to perform deep study of individual objects. Statistical approaches, group selection and analysis became the necessary steps for data analysis. Although some VO tools are widely used by the astronomers, the power of the Virtual Observatory is not yet well known. In 2010 only 15 scientific papers have been published showing data and results obtained with the Virtual Observatory environment (see for ex: Jimnez-Esteban *et al.*, Ptak*et al.*, Rauch *et al.*, Hatziminaoglou *et al.*). This paper reviews several aspects of the Virtual Observatory. The available data are shown is section 2, standards and technics are explained in section 3.1, tools are listed in section 4, and future evolutions will be discussed in the conclusion. Other fields of the Virtual Observatory, like the way to put data into the VO and grid computing will not be surveyed. Moreover refer to IVOA,US-VAO,EURO-VO web sites for complete and up to date information. This review want to show that 'the final goal of the VO is to facilitate and foster astronomical research and astronomers are its ultimate users and scientific requirements should now drive the IVOA process" (Padovani (2010))

2. VO Data

The Virtual Observatory makes available more and more data in several formats. From Solar system and stellar objects to galactic objects, and thoretical data, tabular data are covering several basic quantities like astrometric, photometric, spectroscopic data. Cross identifications have been performed and added values data like redshift, nature, velocity fields, velocity dispersion are available. Surveys are a really powerful source of data covering the whole wavelength range and in ten years time, data available throughout the VO will represent several billions of sources. Indeed undergoing and planned surveys have already plan to deliver their products in formats compatible with the Virtual Observatory standards. The following list show the most important surveys available or that will be available in the next future.

Radio : NVSS, FIRST, WENSS and more to come like LOFAR, SKA

InfraRed : IRAS, 2MASS, DENIS, UKIDSS and more to come like WISE, VISTA and probably EUCLID

Optical : SDSS, 2dF and more to come like PAN-STARRS, LSST, CoRoT

UV : GALEX

X-Ray : ROSAT, XMM and more to come like eROSITA

gamma-ray : EGRET, Fermi and probably SVOM

Other kinds of data are almost ready to be accessible throughout the Virtual Observatory.

numerical simulations : GalMer

3D Data : IFU, FabryPerot, Radio Cubes

transient events : Bursts, SuperNovae

Exoplanet data : CoRoT, Kepler

Moreover national and international data centers (like CADC, ESA, ESO, MAST, SLOAN, CDS) are able to join national efforts to make available data in Virtual Observatory formats. The willing is there, the capabilities also. The time has come to change the way to access data. But still some questions are open to astronomers : 'How to access data ?, How to use them ?". Obviously there was a need to standardize the data, to organize them, a need to standardize the exchanges protocols and a need to see prototypes working. I can say now that we are beyond this point, beyond the testing phases and ready for science exploitation. Nevertheless, the next session will introduce briefly the technical aspects of the Virtual Observatory environment.

3. VO Standards

To explain the strategy of the Virtual Observatory, one has to have an idea of the overall technology, the vocabulary and the techniques that are in constant evolution but are providing stable configuration for accessing and using the data. Next sections give a quick overview of this vocabulary.

3.1. *Data models*

First of all the exchanges and formats are based on the Extensible Markup Language (XML) format (see XML on w3). One of the first activities of the IVOA has been to define the data models to describe most, if not all, data that can be provided by the astronomical community.

Data models definitions include:

- VOTable : that is the main VO format to exchange data (explained in section 3.1.1)
- Space Time Coordinates : STC defines the coordinates of an event or an object

• Astronomical Dataset Characterisation : defines the overall characterisation of an observation
• Simple Spectral lines data model : defines environment and nature of a spectral line
• Spectral data model : defines an observational spectrum

More data models are still in development :

• Observation Provenance : describes the context of observation
• VOEVENT : describes a tansient event ; it is explained in 3.1.2. This standard is foreseen to replace GCN
• Theory : describes and access any numerical simulation

Finaly some data models refer to the disposal of data throughout description of resources (VOResource, Registry, VODataService,...). This part can be found on the IVOA web site. To better understand the models I will focus on 2 models:

3.1.1. *VOTable (Ochsenbein et al. (2009))*

VOTable is one of the first VO standards defined. The VOTable format is an XML standard for the interchange of data represented as a set of tables. It was initially designed to be close to the FITS Binary Table format and has been set as an exchange format for tabular data. It is structured as a set of metadata to describe the fields and a set of separated data. This format allows to better understand the nature of the data in the files, and provided easy access to metadata, and data separately. The metadata are described using unique keywords unified for the astronomical world (UCD,UTYPES).

3.1.2. *VOEVENT (Seaman et al. (2010))*

VOEvent defines the content of a standard information packet for publishing information about a transient event. The goal is to describe targets for robotic telescopes automatic follow-up. It mainly focuses on time events and is a mechanism for broadcasting discoveries that astronomers may wish to follow-up. As the VOTable, it is structured as an XML file with several packets defined respectively to identify the author, characterize the event, define the Space-Time coordinates of the event, define the Instrument configuration and the initial scientific assessment. This event will also list the follow-up observations, a description and references to the event.

3.2. *Data Access protocols*

In order to access data, protocols have been defined. From the basic ones to the most advanced ones, their purpose is to allow the user to acces the data whatever the tool the astronomer is using. Such protocols include:

SIA : Simple Image Access : to access images and part of images depending on coordinates and observational band
SLAP : Simple Line Access : to access spectral lines depending on atomic data and environment
SSA : Single Spectrum Access : to acces spectrum type depending on coordinates and spectral range
TAP : Table and catalog Access : to access data with selection possible on any criteria
ConeSearch : Position related search : to access any sources depending on its coordinates
VOQL : VO Query Languages : to setup queries with defined parameters linked to astronomical searches

4. VO - Tools

To get best advantages of the data disposal, tools have been developed. They can be split in six categories:

Portals and queries : Several portals exist to access data of the virtual observatory, like Datascope, CDS Portal)

Database Discovery Tool : Catalog exploration are available like VOCAT which is a catalog data interface tool to transform astronomical data in databases or SAADA

Tabular DATA Analysis : several tools offer plotting and analysis functionalities like VOPlot, designed for Large data sets, or TOPCAT which is the interface to STILTS Command-line tools for table/VOTable manipulation or VOSTAT - tool for statistical analysis of astronomical data

Discovery- Visualisation : Aladin is one of the tools for Image and Catalogue displaying, VisIVO as well offers a visualisation Interface like DS9 that already has interfaces to VO data. Mosaic Service from VOIndia offers a FITS Image Mosaic Service.

Data Mining : Some tools already offer data mining studies and analysis : MIRAGE - Bell Labs Mirage offers multidimensional visualization of data, VOSTAT is a VOIndia tool for statistical analysis

Spectrum visualisation : VOSpecESA VOSpec and SPECVIew from the STSCI, SPLAT (Spectral Analysis Tool) from UK are tools to handle observational and theoretical spectra

Such tools are really powerful as they hide the complexity of the VO infrastructure throughout easy interfaces and extend the VO capabilities with dedicated functionnalities. These tools provide basic access to VO formated data, FITS and ASCII data. Next session focuses on some of these tools.

4.1. *Aladin*

This tool was born a decade ago in the early ages of the Virtual Observatory. It has been the first interactive software sky atlas visualizer. It allows to visualize multi-wavelength digitized data images, to superimpose entries from astronomical catalogues or databases and access related data and information from distributed services and archives. Built as a java software, it includes plug-in interfaces to plug other software like well known S-extractor and any X-correlations web services. It allows several cross-plots to combine, tablular data and images. That makes of Aladin a really powerful discovery tool.

4.2. *TOPCAT*

Topcat is a table plotter, defined as an interface to the STILTS libraries. Topcat has extended its functionalities to include spherical representation of data and mutliple selection functionalities. Moreover this tool provides also multiple cross correlations with any loaded catalog. This tools is still under development and is one of the most used tool of the Virtual Observatory world.

4.3. *SAMP*

All these tools offer great functionalities to analyse any kind of data but when combined they form a really powerful software suite. The idea is to be able to exchange information from one software to another. This combination is possible using SAMP (Simple Application Messaging). SAMP is a messaging protocol for interoperability to enable individual tools to work together. As a successor of the former messaging protocol "PLASTIC", it is based on XML-RPC. Messages are standardised using keywords defined as standard for exchanges using "mtypes" (message types) and "params" (parameters) . As an extend of applications messaging protocol, a new WEBSAMP has been defined to connect
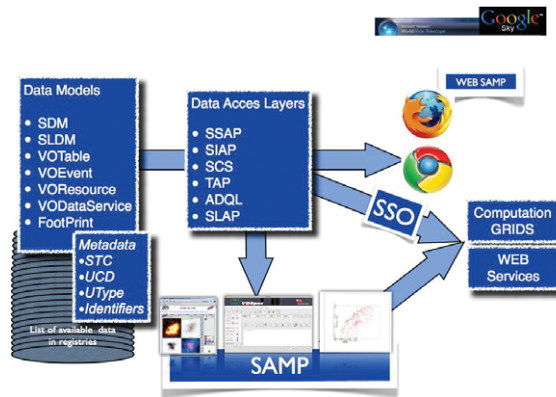
Figure 1. Example of implementation of VO standards in a scientific use case. Get data defined in Metadata,using data accesslayers and use WebServices or VO tools to analyse the data. Then processes can be executed on grids or use web services to get scientific results

WEB applications to the framework. Main realisations of SAMP providing hub and client libraries are available with Java (JSAMP) or python (PySAMP)
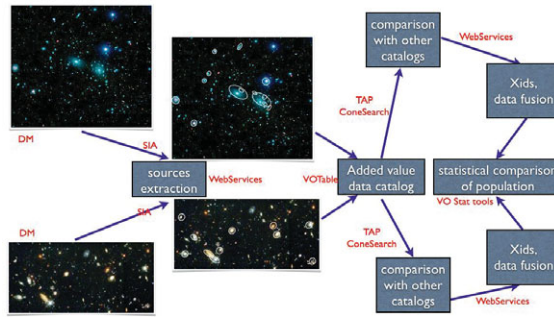
5. Using the VO

Starting from the technical implementation, the scientists can start using the multi-wavelength data with tools developped for the Virtual Observatory. The combination of the different VO standards for astrophysical purpose is shown in figure 1. It gives a basic idea on how the data can be used throughout standards and services. Fluxes are described with blue arrows while standards are shown in boxes. The goal is to use these standards and protocols in a transparent way to provide scientific outputs.

Trying to use the infrastructure, one can define its own use case. As seen in figure 2, the astrophysicist may, for example, try to compute the statistical properties of two populations of galaxies derived from different observational fields or using two different wavelength range. Starting from extracted images from the Virtual Observatory (using Data Models and SIA), sources catalog are created (using Web Services SExtractor), then VO tables are created, to be cross identified with other catalogs (using TAP, ConeSearch) and data fusion is operated (using tools TOPCAT,ALADIN, Web Services). Then a comparative study can be performed and statistical results can be derived.

6. Conclusion

The virtual Observatory environment becomes easier and easier to use with more and more data available. After focusing on technical parts, this is a new era where the scientific goals drive technical developments. Of course there is still some things to improve (check for data quality, the curation of data, new standard to finalize). Of course there is still som things to do (include ALL data, define even easier-to-use portals). Of course there is new area to work on (Data Mining, ObsTAP, cloud computing). But Science can be done easily, and the technical efforts put in the definitions and infrastructure of the Virtual Observatory start to be really fruitful to science. One last question: Who is going to try ?

Figure 2. Example of Astrophysical implementation : a comparison of cluster population, using the virtual observatory standards and protocols. From the initial observations to the statistical comparison, the protocols and standards (in red) and processing steps (boxes) are shown.

References

IVOA : http://www.ivoa.net

US-VAO : http://www.usvao.org/

EURO-VO : http://www.euro-vo.org/

CDS Portal : http://cdsportal.u-strasbg.fr/

Data Scope : http://heasarc.gsfc.nasa.gov/cgi-bin/vo/datascope/init.pl

VOCAT : http://vo.iucaa.ernet.in/ voi/vocat.htm

VOSTAT : http://vo.iucaa.ernet.in/ voi/VOStat.html

SAADA : http://amwdb.u-strasbg.fr/saada/

Aladin : http://aladin.u-strasbg.fr/

VisIVO : http://visivo.oact.inaf.it/index.php

Mosaic Service VOINDIA : https://sdss.iucaa.ernet.in:9191/MosaicApplication-war/

MIRAGE - Bell Labs : http://cm.bell-labs.com/who/tkh/mirage/

VOPLOT: http://vo.iucaa.ernet.in/ voi/voplot.htm

SAMP : Taylor *et al.* 2010 , http://www.ivoa.net/Documents/latest/SAMP.html

ESA VOSpec : http://www.sciops.esa.int/index.php?project=ESAVO&page=vospec

SPECVIew : http://www.stsci.edu/resources/software_hardware/specview

SPLAT : http://star-www.dur.ac.uk/ pdraper/splat/splat.html

Euro-VO ICE : http://www.eurovo-ice.eu/twiki/bin/view/EuroVOICE/ICEOverview

*Open Sky Query*http://openskyquery.net/Sky/skysite/OSQform/default.aspx

Taylor et al. 2010 topcat : http://www.starlink.ac.uk/topcat/

Ochsenbein F,Williams R, et al. http://www.ivoa.net/internal/IVOA/IvoaVOTable/votable.html

Seaman R., Williams R., et al. http://www.ivoa.net/internal/IVOA/IvoaVOEvent/voevent_v2.html

P. Padovani interop meeting (2010).

XML : http://www.w3.org/XML/

Jimnez-Esteban, F. M. *et al.* 2011, *A&A*, 525, A29

Ptak, Andrew; Krughoff, S.; Connolly, A, 2010, *HEAD*, 11, 0903

Rauch, Thomas *et al.* 2010,arXiv1011,3628R

Shirasaki, Yuji; Tanaka, Masahiro; Ohishi, Masatoshi; Mizumoto, Yoshihiko; Takata, Tadafumi; Yasuda, Naok, 2010, *IAUS*, 267, 267S

Hatziminaoglou, E., 2010, *ASPC*, 424, 411H

Yin, Q.-Z., Lee, C.-T. A., & Ott, U. 2006, *ApJ*, 647, 676