THE VIRTUAL OBSERVATORY

ENRIQUE SOLANO
Laboratorio de Astrofísica Espacial y Física Fundamental, Instituto Nacional de Técnica Aeroespacial, E-28080 Madrid, SPAIN

Abstract: The advances in technology (telescope design and fabrication, large-scale detector arrays, computing capabilities) are permitting for the first time to explore the Universe in a multi-parameter space. Although this situation should potentially lead to a more complete and less biased understanding of complex astrophysical phenomena, the reality is that the progress in the scientific exploitation is not keeping pace with the exponential growth of data. Two are the major limiting factors: The absence of a real interoperability among astronomical archives and the lack of scalability in the classical methods of retrieving and analyzing astronomical data to tackle the new vast datasets.

The Virtual Observatory is an international project aiming to solve this situation by: a) the creation of a federation of astronomical archives that, with the implementation of new technologies and standards, provides an easy and efficient access to the astronomical data ("data grid"). b) the development and implementation of analysis tools by the data centres holding the data ("service grid"). Data mining, as a way to perform an efficient and systematic study of the vast amount of information that will be available from the federation of astronomical archives, constitutes one of the key activities for the Virtual Observatory and where the greatest scientific benefits are expected to come from.

Although VO is still an emerging project, it is already considered both from the technical and scientific point of view a basic requirement for the astrophysical research and the framework where to settle in the short-term the astronomical archive-related activities. The biggest data providers (NASA, ESA, ESO, ...) have understood the importance of this initiative and are already displaying their contents under the VO requirements.

In this paper I will summarize some of the work areas, both from the technical and scientific point of view, that are being developed in the context of International Virtual Observatory Alliance (IVOA[1]). The last part of this contribution will focus on the Spanish Virtual Observatory and its main lines of research.

---

[1] http://www.ivoa.net

1 Introduction

The final goal of a scientific mission is to obtain new data that can give origin to new science. In order to optimize efficiency, it is necessary to develop systems that ensure an easy access to the expensive-to-obtain datasets. In Astronomy, all the space-based missions and most of the largest ground-based telescopes have developed friendly archive and data access systems. It is important to remark the difference between a safe storage system and a real scientific archive which needs to incorporate tools for data querying, visualization and retrieval.

It is undoubtedly that, nowadays, archives play an essential role in modern astronomical research as demonstrated by the intensive usage that the scientific community makes of these facilities. A well designed and properly implemented archive is a fundamental step toward the full exploitation of the data. INES[2], the archive and access system of the IUE (International Ultraviolet Explorer) satellite, with a sustained high level of usage by the community as demonstrated by the fact that the data volume transferred in the last four years has surpassed the number of observations made during its 18 years of lifetime ($\sim$104000 spectra), is an excellent example of how far-reaching science can be done 28 years after the launch of the satellite and ten years after the end of operations.

Several are the reasons to explain the importance of archives in Astronomy:

- *Efficiency*: Astronomers might find that the observations they want have already been made and that it is simply a matter of retrieving the relevant information from the archive. With archives a given dataset can serve for many different scientific purposes, including some not considered when the observations were made.
- *Time domain exploitation*: All astronomical objects vary on time-scale. Using archives is the only way to exploit the time domain and study both periodic (e.g. variable stars) and transient phenomena (supernovae, Gamma-Ray Burst,...).
- *Democratization*: Archives permit astronomers from countries with limited resources for research to have the opportunity of accessing high-quality data and produce top-level science.

The already existing archival information, together with the all-sky surveys foreseen in the coming years, will produce a coverage of large areas of the sky in most of the wavelengths ranges. An efficient coordination among the archive centres is, therefore, essential to optimize the scientific return of these projects. However, although most of the archive centres have developed users friendly interfaces allowing

---

[2] http://sdc.laeff.inta.es/ines/

an easy data retrieval, it is also true that the big heterogeneity among them (different formats, units, policies, retrieval protocols,...) makes it very inefficient the data compilation from more than one archive.

Added to this lack of interoperability is the management of the volume of data generated by the all-sky surveys. SDSS[3] and 2MASS[4], with millions of observations, are good examples of this. The classical approach of transferring data from the archive to the local desktop for their further analysis is not applicable. For instance, the transfer of the SDSS DR2 (Data Release 2) catalogue would take a week with a good connectivity (1 Mb/s) ... and two months to transfer the whole set of images which would occupy 1000 DVDs. The situation will be even worse at the end of the project as the expected number of data will be a factor of three greater. Moreover, the problem is not only constrained to the data transfer but it also affects the analysis techniques as most of the tools are not able to handle such amount of data.

The Virtual Observatory is an emerging research environment created with the aim of solving the problems that this situation creates to the astronomical research. The concept of the Virtual Observatory is that all the world's astronomical data should feel like it sits on the astronomer's desk top, analysable with a user selected workbench of tools and made available through a standard interface. Different phase-A projects began in 2001 (NVO[5], AVO[6], Astrogrid[7]). In June 2002, IVOA was created as the framework to coordinate these and other similar initiatives. At present, IVOA is formed by 16 projects. In Europe the Virtual Observatory initiatives are being built around EURO-VO. Based on the experience gained within the AVO project, EURO-VO is conceived as a Phase-B project with the objective of deploying a fully operational VO in Europe. Further information about EURO-VO can be found in [1]. Similar transition activities in other VO projects are given in [2].

2 Interoperability among astronomical services

2.1 The problem

Interoperability among astronomical archives will provide the scientific community with an integrated access to a wide variety of information of a very diverse nature (astronomical data in image or tabular form, catalogues, bibliographic information, ...). Interoperability is a key issue for multiwavelength astronomy where it is necessary to gather data in different regions of the electromagnetic spectra. Figure 1 is an

---

[3] http://skyserver.sdss.org/edr/en/sdss/

[4] http://www.ipac.caltech.edu/2mass/

[5] http://www.us-vo.org

[6] http://www.euro-vo.org

[7] http://www.astrogrid.org

example of how even straightforward results can be obtained only if multiwavelength data are available.

Although before the advent of the Virtual Observatory there were already projects that understood the importance of this issue and developed their systems to fulfill this requirement (the IUE project was pioneering in this field and the data exchange between the INES and the ADS[8] archives can be considered as one of the first examples of real interoperability), the situation was that an efficient communication among astronomical data services did not exist and, in practice, the simultaneous query to multiple databases was done, most of the times, slowly and painfully by hand. Let's see now some examples on how this situation was affecting the astronomical research.

- *Lack of interoperability among catalogues*: VizieR[9] is a service maintained by CDS[10] that gives access to more than 5000 astronomical catalogues. Although each column of the tables contained in the catalogues had associated a name and a unit, there were not fixed rules for the naming convection that was left up to the users. This led to such a situation where, despite the enormous wealth of information contained in VizieR, it was not possible to answer easy questions like "give me all tables containing the V magnitude in the Johnson system" simply because that concept, the V magnitude, had associated almost 150 different names.
- *Lack of interoperability among spectroscopic archives*: Building Spectral Energy Distributions (SEDs) is a clear example of how the lack of interoperability was affecting the astronomical research. Gathering all the spectrophotometric information available for a given object implied, firstly, to identify the services of interest, then, submit the same query for each one the services and, once all the information has been collected, tackle problems related to the unit conversions, flux calibrations and/or data formats.
- *Lack of interoperability among image archives*: Even if the same query (coordinates of the central position and search radius) was sent to two different image servers, in most of the times, it was not possible to directly compare and cross-match the results as the images showed different scales, orientations and astrometric accuracy.

2.2 The solution

Standardization is the basic issue to solve the interoperability problems. Although some steps were already done in this field before VO (e.g., a standard data format,

---

[8] http://adsabs.harvard.edu/

[9] http://vizier.u-strasbg.fr/viz-bin/VizieR

[10] http://cdsweb.u-strasbg.fr/

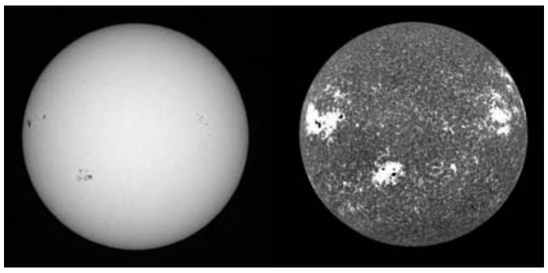


Figure 1: Interoperability: a basic requirement for multiwavelength astronomy. *Left:* Information from the solar photosphere, in particular about the sunspots, can be obtained in visible light. *Right:* Using the CaII-K filter it is possible to access the chromosphere and study supergranularity, a phenomenon produced by the vertical transport of hot gas to the solar surface through convective cells. Given that the images are simultaneous in time, a direct comparison between them comes to the conclusion that the strong magnetic fields associated to sunspots prevent the chromospheric supergranularity.

FITS, was developed in the late seventies to enable the exchange of astronomical data between computers and systems of different type and a standard reference code, *bibcode*, was set to describe a reference to a paper published in a journal or book within the astronomical community), it was clear that an accurate definition of standards was still missing. The Virtual Observatory is working in defining standards in the following fields:

- *Standard semantic*: The Unified Content Description (UCD[11]) is a controlled vocabulary in order to avoid proliferation of terms and to reduce ambiguity. It provides the basis for the semantic description of most of the concepts that can be measured or handled in Astronomy and constitutes a basic requirement not only to handle the heterogeneous catalogue information contained in Vizier but also to ensure interoperability among data archives with similar information but with different names.

- *Standard Access Protocols*: The Virtual Observatory is not interested in how archives are implemented or how data are stored. On the contrary, VO is interested in how data are exposed to the world through well-defined requests and responses. Conesearch, SIAP and SSAP are VO access protocols already in use. In short, they require the service responds to hhtp GET requests of the following format: <URL>?POS=&SIZE=, where POS and SIZE refer to

---

[11] http://www.ivoa.net/twiki/bin/view/IVOA/IvoaUCD

the coordinates of the center and the size of the search region respectively. For instance, if we are interested in retrieving all the spectra of Vega available from the INES data server, our query using the SSAP protocol should be like this:

http://sdc.laeff.esa.es/ines/jsp/ssap.jsp?POS=279.2347,38.7836&SIZE=0.0833

Simplicity is the main characteristic in common to these three protocols. Because of this, their performances are seriously limited if complex queries are required. ADQL[12](Astronomical Data Query Language) represents an alternative approach to tackle this problem. Briefly, ADQL is an XML language for constructing SQL-like queries that, in addition to the typical commands, also implements built-in functions that allow, among other possibilities, cross-matching between astronomical catalogues or defining sky regions.

- *Standard Data Format*: VOTable[13] is the XML standard proposed to exchange tabular data between agents acting in the framework of the VO. VOTable is more flexible and interoperable than FITS and allows, for instance, data and metadata to be stored separately with the remote data linked. Just as we are used to the idea of sending a pointer to a document (URL) in place of the document, so we can send metadata-rich pointers to data tables in place of the tables themselves. The VOTable structure is similar to that of a FITS file in the sense that it consists of a hierarchy of metadata (playing a similar role as the keywords in the header of a FITS file) and the associated data that, in the case of VOTable, can be represented by a link. The output from the query to the INES archive executed in the previous section provides a good example of VOTable.
- *Data models*: Astronomical data are stored in many different ways. It is necessary, therefore, to define abstract entities to ensure that the same kind of data are described in the same way. These entities are called data models and they are needed by data providers to know which metadata must be stored associated to a given type of datasets. A good example is the Observation Model, an abstract representation of an astronomical observation that contains information about the spatial and temporal coverage, the instrumental configuration, the scientific reasons for which the observations have been made, the scientific publications where the observations have been quoted, the reduction and calibration as well as the format of the final output products.
- *Registries*: An essential capability of the Virtual Observatory is to provide a mean for describing what data and computational facilities are available, where, and once identified, how to use them. For instance, if a user is interested in

---

[12] http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL

[13] http://www.ivoa.net/Documents/latest/VOT.html

finding X-ray images there is no point in querying the ISO archive. These services are called *registries*.

More information on these topics can be found at http://www.ivoa.net/Documents.

2.3 An example of interoperability: Intelligent agents in robotic telescopes

The advances in computer technologies and communications have allowed scientist to access remote facilities in the same way as they would do it locally, controlling and monitoring the system sufficiently fast as to allow them to react and make decisions in real time. Robotic telescopes constitute a step further as they do not require the, even remote, human presence. A robotic telescope executes the pre-scheduled tasks in an autonomous way using specific programming languages. Despite its small-size, robotic instrumentation is called to play an important role in modern observational astrophysics as they are particularly well-suited for projects that require a large number of observations (e.g. discovery of extra-solar planets using the method of transits) or follow-up/preparatory works of large projects and space missions (e.g. COROT[14]).

However, many astronomical events (e.g. supernova explosions) happen suddenly and unpredictably. Therefore, an automated system able to react to these unforeseen phenomena would constitute, without any doubt, a very valuable tool. *eStar*[15] is a UK project formed by a network of autonomous telescopes and associated rapid data reduction pipelines connected together using GLOBUS[16] middleware. The system implements intelligent agents, defined as computational entities that perform their actions with some level of proactivity and/or responsiveness. In this way, the system is able to think and react for itself, deciding whether something it has discovered is interesting enough to need more observations and, if so, going ahead and getting them.

A science case that nicely fits the potential offered by the eStar system is the discovery of long term dwarf novaes. In a Virtual Observatory framework, the system is able to cross-correlate (in real-time) the observations with the USNO-A2[17] to perform the source identification. After this, the system mines SIMBAD[18] to flag all the already known variables. If, according a pre-defined classification criteria, one of the unknown variables turns out to be a dwarf nova, the intelligent agent will request further follow-up observations from the network of telescopes.

---

[14] http://corot.oamp.fr/

[15] http://www.estar.org.uk/

[16] htp://www.globus.org

[17] http://vizier.u-strasbg.fr/viz-bin/Cat?I/252

[18] http://simbad.u-strasbg.fr/sim-fid.pl

3 Management of large volumes of data

3.1 The problem

For many years, Astronomy has been living in the world of the small numbers. The usual way of carrying out astronomical research was that of a single astronomer or a small group of astronomers performing observations (or retrieving these observations from archives if they already existed) of a small number of objects in one or a few narrow wavelength bands. In all cases, the data analysis was done locally. Although this methodology is adequate for many astronomical cases, it has demonstrated to be quite inefficient in other occasions, in particular when it is necessary to deal with large volumes of data, a scenario becoming more and more common in modern Astronomy due to the exponential growth of the information available from astronomical services.

Sky surveys like SDSS, 2MASS, DENIS,..., carrying out millions of observations and measuring tens of attributes for each one of them, are the main actors but not the solely responsibles for this situation. The ESO/ST-ECF Science Archive Facility is also a good instance of the dramatic growth of the astronomical archive holdings: with an increase of a factor of a hundred in the next seven years, it will reach the PetaByte level in 2012. Planned survey-dedicated telescopes like VISTA, VST or LSST, with a data rate a thousand times larger than the current surveys, will even make things worse, ending up in a situation in which the classical approach of transferring data from the archive to the local desktop for their further analysis will be, by far, not applicable.

3.2 The solution

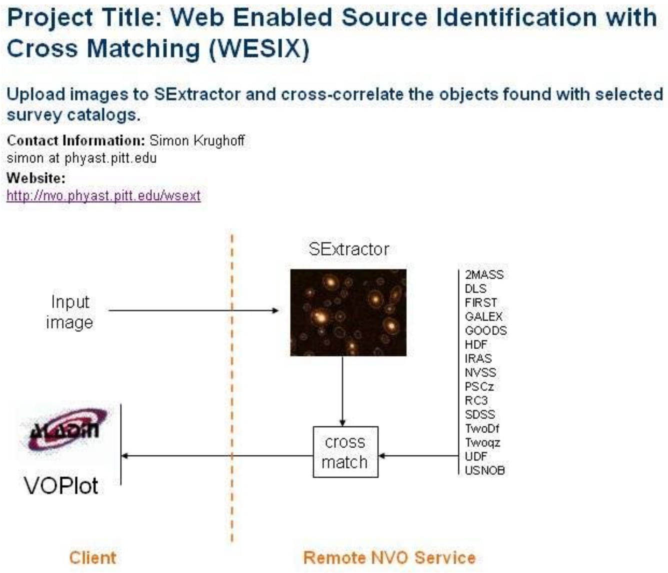
The volumes of data mentioned above clearly do not fit either the “classical” retrieval procedures or the research methodology astronomers are used to follow. Transferring of results instead of data is one of the solutions that the Virtual Observatory proposes to solve this situation. The data are left as they reside and operations on them are performed remotely on the data server. This new scenario puts strong computing requirements on data centres which must be able to manage intensive processing and analysis tasks using either local resources or distributed computing like the Grid. Moreover, on top of this infrastructure, it will be necessary to create an analysis tools layer that, working as web services, allow remote functionality as if they were local.

In astronomical analysis packages such as MIDAS, IDL, IRAF, astronomers select building blocks (commands, procedures, scripts,...) and build up analysis tools from them. In the Virtual Observatory, the methodology is very much the same except that the building blocks are web services running on remote computers. WESIX[19], a web service developed by the NVO, is an excellent example of this type of applications. In

---

[19] http://frank.phyast.pitt.edu:8080/wsext/

WESIX, the user is required to upload an image to the server. After this, SExtractor performs the source identification, sources that are subsequently cross matched with a pre-defined list of survey catalogs. The results are sent back to the user in VOTable format who can analyse them using VO tools like, for instance, VOPlot[20] or Aladin[21]. As we can see in Figure 2, all the analysis is performed on the server side and not on the user's side as normally occurs if the classical methodology is used.



Project Title: Web Enabled Source Identification with Cross Matching (WESIX)

Upload images to SExtractor and cross-correlate the objects found with selected survey catalogs.

Contact Information: Simon Krughoff
simon at phyast.pitt.edu
Website:
http://nvo.phyast.pitt.edu/wsext

Figure 2: The WESIX workflow (courtesy of NVO).

3.3 Analysis tools. Data mining

The potential for scientific discovery afforded by the mass of data available in the astronomical services is enormous and the Virtual Observatory provides an excellent mechanism to obtain new and unexpected scientific results of major significance emerging from the combined use of datasets, a type of science that would not be possible from such sets used singly. However, transforming vast masses of bits into a refined knowledge and understanding of the universe is a highly complex task that requires very specific analysis tools. *Data mining* can be defined like the non-trivial

---

[20] http://vo.iucaa.ernet.in/ voi/voplot.htm

[21] http://aladin.u-strasbg.fr/aladin.gml

extraction of previously unknown implicit information and represents the best way to perform an efficient and systematic study of the information available from the VO federation of archive centers. Automated classification based on observations, the discovery of the unknown using unsupervised methods, the search of rules and properties, ..., are some of the fields that clearly benefit from the use of data mining techniques. Let's see an example of the potential of this methodology in the framework of the Virtual Observatory:

Automated classification of ROSAT sources

It is estimated that only 10% of the sources observed by the ROSAT satellite have a reliable classification. In most cases the identification rests on cross-correlation with catalogues of already classified objects. Also, for some of them, detailed follow-up observations have been performed on a source-by-source basis, an extraordinarily time-consuming approach that does not scale at all with the total volume of data (100 000 unclassified sources). ClassX[22] [3] is an on-line system for automated classification of X-ray sources: given an input source, the system performs an automatic search for counterparts among the VO-compliant data services and retrieves from those archives all the attributes required for the classification into six class categories: stars, white dwarfs, X-ray binaries, galaxies, active galactic nuclei and clusters of galaxies.

4 The Role of Science in the VO Projects

Although technologically enabled, the Virtual Observatory must not be seen as a mere technological project since it is driven by science as its final goal is to produce better, new and more efficient science. This concept was clearly understood by the major VO projects which, since very early, set up Science Working Groups to provide scientific advice to the projects. One of the major tasks of these groups was the identification, with a clear emphasis on the definition of the science requirements, of cases that could benefit from the use of the VO technology. The ScienceProblem[23] or the TopTen[24] lists compiled by AstroGrid and the AVO Science Reference Mission, consisting of a number of key science cases that should be achieved in a fully operational VO framework, are examples of this.

It is a fact that a high percentage of the successfulness of the VO projects rests on the involvement of the research community. The Virtual Observatory is a new concept still poorly known and this could lead to a infrautilisation of its resources. IVOA is aware of this and, since very early, has emphasized the importance of annual

---

[22] http://heasarc.gsfc.nasa.gov/classx/

[23] http://wiki.astrogrid.org/bin/view/VO/ScienceProblemList

[24] http://wiki.astrogrid.org/bin/view/Astrogrid/ScienceProblems

demonstrations of the emerging VO capabilities as a way to engage the scientific community in the Virtual Observatory endeavour. In January 2003, 2004, and 2005, coordinated demonstrations of VO developments were held in the US and Europe which highlighted progressively more complex VO capabilities.

The Virtual Observatory has already reached a level of maturity enough to produce relevant science results. The first major discovery made with the Virtual Observatory is described in [4]: the AVO science team involved in this project discovered 31 previously undetected supermassive black holes in the GOODS (Great Observatories Origins Deep Survey) fields following a VO methodology. In addition to this, several new scientific findings through VO-based research have been already published ([5], for instance). Although a detailed description of these projects is beyond the scope of this paper I summarize below a couple of science cases which clearly show the advantages of adopting a VO methodology.

4.1 Searching for Rare Objects: New Candidates in the AGB - Post-AGB - PNe Phase

The transition from AGB stars to PNe is poorly understood due to the scarce number of objects presently identified. Two are the major reasons for this: On the one hand, this is a very rapid evolutionary phase with typical times ranging from 100 to 1000 years. On the other hand, many objects in this phase are heavily obscured in the optical by thick circumstellar envelopes consequence of strong mass loss (up to 0.001 $M_\odot$/year). To test current evolutionary models and to accurately characterize the properties of the objects in this phase is necessary, therefore, identify as many new candidates as possible by making massive and systematic search on infrared data.

This was the objective of the Stellar Science Case proposed for the AVO Scientific Demonstration in 2005. The search of new sources was made by combining the IRAS and MSX catalogues. The workflow consisted of a number of data manipulation steps that benefited a lot from a VO methodology: filtering (to discard bad quality data), cross-matching (to identify MSX counterparts to the IRAS sources and flag sources already known using SIMBAD) and column manipulation (to define MSX-IRAS colour indices) are some of the operations that take advantage of the VO tools and that would be far more difficult any other way.

As a result of this analysis, 100 new candidates in the transition phase were found. Even if it is necessary further spectroscopic confirmation of the intrinsic nature of these objects, the number is relevant and the advantages of using a VO methodology clearly remarked if it is compared with the 200 bona-fide members previously known or with the 21 well-determined transition sources obtained after years of analysis using classical methods [6].

4.2 Searching for Rare Objects: Discovery of Brown Dwarfs

A fundamental question in the area of star formation is the form of the stellar mass function at the lower end, in particular, the contribution of brown dwarfs to the mass budget. The discovery and census of the nearby brown dwarfs has been identified as a key VO-Science case both by AstroGrid (included in the "Top-Ten" cases), EURO-VO (through its Science Reference Mission) and NVO (who selected it for its 2003 Science Demonstration).

Brown dwarfs are objects occupying the gap between the least massive stars and the most massive planets. They are intrinsically faint and rare objects so their detection is not straightforward and, in fact, was almost impossible until the advent of global surveys at deep optical and near-infrared wavebands like 2MASS, SDSS or DENIS. This has led to the situation that, although the brown dwarf density should be, according to theoretical expectations, twice higher than that for Main Sequence stars [7], only a handful of brown dwarfs are known at less than 10 parsecs from the Sun compared to more than 300 stars within the same volume.

This problem can be alleviated mining the sky surveys previously mentioned (an others foreseen in the near future like UKIDSS) through an appropriate combination of colours in the optical and the infrared and/or proper motion information (since any detectable brown dwarf must be nearby and, so, it will have relatively high proper motion). [8] and [9] described the discovery of new substellar objects using this methodology. Nevertheless, building a census of substellar objects implies the discovery of a statistically significant number of them through queries that combine attributes available from different archives, an approach out of the scope of the "classical" methodology but that perfectly fits into the Virtual Observatory. The discovery of a new L-type brown dwarf by cross-matching the SDSS and 2MASS databases using VO tools during the 2003 NVO Science Demonstration clearly shows the potential of carrying out this type of analysis in the VO framework.

5 The Spanish Virtual Observatory

The Spanish Virtual Observatory (SVO[25]) is a project funded at national level that officially joined IVOA in June 2004. It was born with the goal of coordinating in an efficient way the work done by the Spanish astronomical community in the framework of the Virtual Observatory. The SVO project, presently formed by 5 institutes being LAEFF the main actor, is divided in the following work areas: uptake of VO standards, development of analysis tools for their further application in VO Science Cases of interest for the SVO scientific community and compliance of theoretical models with the Virtual Observatory requirements. The work that is being done in these

---

[25] http://svo.laeff.esa.es

fields is outlined here.

5.1 Work lines

Astronomical Archives and services in the VO framework

The archive activities at LAEFF started in the nineties. Since then, LAEFF has been responsible for the development and maintenance of the following archives and services:

- *The INES Archive:* In June 2000 LAEFF took over the responsibilities of the INES[26] system acting as Principal Centre and coordinating the user support and data access with a number of National Hosts distributed world-wide. Loyal to its pioneering role in the world of astronomical archives, INES has been present in the VO endeavour since the kickoff meeting at European level held in Strasbourg in 2001. Shortly after this meeting, a project to adapt the INES archive to the Virtual Observatory requirements and standards was initiated. The way how the INES project envisaged the evolution of astronomical archives nicely fits the IVOA concept allowing a smooth transition to the VO world and making INES one of the first VO-compliant archives.

- *The COROT Archives:* Much of the experience gained with INES was applied to the development of GAUDI[27], a preparatory archive for the COROT mission. GAUDI contains spectroscopic and photometric data as well as information of physical parameters of 2 000 objects of the Asteroseismology Programme. The GAUDI spectroscopic information can be retrieved in VOTable format using the SSAP protocol.

  In March 2003 LAEFF was selected, together with CDS, as responsible for the long-term maintenance of the COROT Final Archive, which will contain all COROT data processed in a homogeneous and uniform way. To ensure the optimum scientific exploitation of these data, LAEFF will guarantee the compliance of the archive with the VO standards.

- *The OMC Archive:* OMC[28] is the acronym for Optical Monitoring Camera, an optical telescope on-board the INTEGRAL satellite, an ESA gamma-ray mission. The OMC data server is open to the astronomical community since November 2003 and presently contains light curves for more than 100 000 objects. At the time of writing, the SSAP access protocol and the time-series data model are being implemented in the OMC archive.

---

[26] http://sdc.laeff.inta.es/ines/

[27] http://sdc.laeff.inta.es/gaudi/

[28] http://sdc.laeff.inta.es/omc/

- *The GTC Archive:* The Gran Telescopio Canarias (GTC[29]) is a 10.4 meter telescope which is being built at the Roque de los Muchachos Observatory (La Palma, Spain). Its first light is foreseen for the end of 2006 and LAEFF is in charge of the design, implementation and maintenance of the Scientific Data Centre. The integration of a infrastructure like the GTC Archive into the VO Data Centre federation will be a step of enormous importance toward the full exploitation of its unique characteristics.

- *The DARWIN preparatory data server:* Due to the large influence of the radiation of the parent star on the planet atmosphere, a careful target selection is a critical issue for the DARWIN[30] mission. Consequently, a precise knowledge of the observational properties of the parent stars as a way to estimate their fundamental parameters is required. LAEFF is in charge of developing a system whose goal is to get all the relevant information already available in astronomical archives and catalogues of the potential DARWIN candidates taking advantage of the Virtual Observatory standards and tools. The system is also able to include new observations performed in the framework of the preparatory observational roadmap.

Data mining tools

As pointed out in previous sections, in the framework of the Virtual Observatory an archive cannot be a simply data repository but must implement analysis tools to optimize the scientific exploitation of its contents. SVO is playing a pioneering role in this field and has already implemented the following tools in their archives:

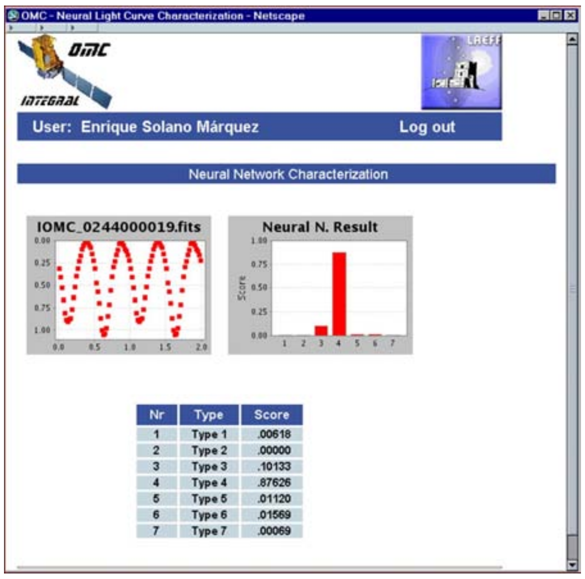
- *Light curve classification:* The OMC archive incorporates a tool that performs a supervised analysis of its contents (Figure 3). The system, a Bayesian network classifier originally focused on the classification of eclipsing binaries according to their physical parameters and light curve morphologies [10], has been recently improved to tackle the classification of other types of variability. In the near future the system will also be able to handle transient phenomena like flares using wavelet techniques. Similar tools are being developed for other missions like COROT or GAIA.

- *Spectral Energy Distribution fitting:* The study of protoplanetary disks is currently undergoing an exciting stage partly propelled by the discovery of extrasolar planetary systems following the detection of 51 PegB by [11]. The possible discovery of telluric planets in the near future, in addition to the Jovian-like

---

[29] http://gtc.iac.es/

[30] http://sci.esa.int/science-e/www/area/index.cfm?fareaid=28



Nr	Type	Score
1	Type 1	.00618
2	Type 2	.00000
3	Type 3	.10133
4	Type 4	.87626
5	Type 5	.01120
6	Type 6	.01569
7	Type 7	.00069

Figure 3: The OMC automated classification system.

planets already detected, will pose interesting questions on the formation of extrasolar planetary systems. Knowledge of the properties of protoplanetary disks and how they evolve from the dense and turbulent disks around PMS stars to the tenuous debris disks around MS objects is, therefore, a crucial step in such a process. The study of this evolutionary process was the SVO science case proposed for the AVO Science Reference Mission. This analysis requires studying a statistically significant number of stars covering a wide range of masses and evolutionary stages (HAeBes, ETTs, CTTs, PTTs, A-shell and Vega-type stars) to track and characterize from an observational point of view the evolution of protoplanetary disks.

Theoretical modelling of Spectral Energy Distributions (SEDs) has proven to constitute an invaluable tool for understanding the structure and properties of protoplanetary disks. However, SEDs building requires accessing to a variety of astronomical services providing, in most of the cases, heterogeneous information. Moreover, model fitting and the associated detailed analysis demand a tremendous amount of work and time which make it very inefficient even for a modest dataset.

In the framework of the Spanish Virtual Observatory we have developed VOSED[31], a Web application to build spectral energy distributions taking advantage of the already existing VO standards and tools. VOSED queries the VO registry to gather all the spectroscopic information available for the object of interest, information that is later complemented with queries to other VO services like, for instance, VizieR. The objective is, given an object, to compare its observed SED with a grid of synthetic models (including both stellar and disk contribution) in order to derive the physical parameters of the star and the disk, an approach that has proven to be highly successful in understanding the structure and properties of protoplanetary disks.

Model fitting constitutes, therefore, a fundamental step in this process. VOSED includes a tool to estimate the model parameters (both stellar and disk) following a Bayesian method. The main aim of the tool is to quantitatively analyse the data in terms of the evidence of models of different complexity, evaluate what other alternative models can compete with the most a posteriori probable one and what are the most discriminant observations to discard alternatives. Preliminary analysis have proven the ability of the system to reproduce the results obtained by [12] in a much more efficient and homogenous way.

Theory in the context of the Virtual Observatory

There are many fields in Astrophysics with a strong need of direct and rigorous comparison between theoretical models or simulations and real data. However, the different architectures, programming codes, formats, etc, make it, most of the times, extremely difficult and inefficient.

In order to define the requirements needed for a full interoperability between observations and models, the IVOA Theory Group[32] was created in January 2004. One of the first issues undertaken by this group was the definition of query protocols adequate for theoretical data: the VO protocols already defined for observational data (ConeSearch, SSAP, SIAP,...) are based on absolute positions on the sky and, in most cases, cannot be applied to theoretical models. Moreover, for theoretical models, there is not a common set of input parameters that could play the role of the coordinates for observational data as each model has its own collection of input parameters that normally differ from one model to another.

To handle this problem, SVO and ESA-VO have defined the TSAP (Theoretical Spectra Access Protocol) protocol as an attempt to incorporate theoretical models to the VO. With TSAP, a client-server dialog is established in such a way that the client can ask the server for the collection of the allowed input parameters as well as the valid values for each one of them. In this way, the client can build an adequate form

---

[31] http://sdc.laeff.inta.es/vosed/

[32] http://www.ivoa.net/twiki/bin/view/IVOA/IvoaTheory

for a given model and shows it to the user, who will select the preferred parameter values or ranges. Afterwards, the client will build a valid query and will send it to the data server which will give back the results in VOTable format.

Using this approach, SVO has implemented TSAP services for two theoretical datasets: the Kurucz's ATLAS9 models of stellar atmospheres as described by [13] and the models of irradiated accretion disks around pre-Main Sequence stars by [14]. Both datasets are also accessible through a web page[33] from where it is possible to consult the on-line library of models. Likewise, SVO is collaborating with INAOE in the development of PGos3[34], a federative project to study the star formation history in galaxies. A use case[35] based on the comparison using TSAP among different grids of evolutionary synthesis codes with observational data was presented in the IVOA interoperability meeting held at ESAC in October 2005.

The task of implementing a TSAP and/or a web page for theoretical models can be a laborious work in many cases. To simplify it, SVO has developed DB2VO, a php application that, taking a collection of ASCII files (the grid of models) and a table with the associated description as input parameters, is able to build both the TSAP service and the web interface in an automated way.

5.2 The Spanish Virtual Observatory Thematic Network

Shortly after the beginning of the SVO project it was clear that the interest of the Spanish astronomical community in VO went far beyond the boundaries of the project. To solve this situation and to create a broader framework comprising all the initiatives, the SVO project applied for a Thematic Network to the Spanish Ministry of Education and Research. The primary goal of the network (approved in December 2005 for a period of three years) is to enhance the collaboration among the Spanish groups with interest in VO through the organization of meetings, tutorials and workshops. At the time of writing, the network is formed by more than 60 researches from 17 institutes and universities and it is open to everybody interested in the field. The kick-off meeting was held in Madrid in April 2006. The meeting was divided in five sections: VO Services (with contributions from LAEFF, IAA, Univ. Barcelona and Univ. País Vasco), Theoretical Models (LAEFF, IAA), Data Mining (UNED, Univ. La Laguna, Univ. Coruña, Univ. Alicante), Information Systems (ROA, UCM, LAEFF) and Science with Virtual Observatories (LAEFF, IAC). More information about the meeting and the network can be found in the SVO web page.

---

[33] http://svo.laeff.esa.es/theory/db2vo/html/

[34] http://ov.inaoep.mx

[35] http://www.ivoa.net/twiki/bin/view/IVOA/VOtheorySynthesis

6 Conclusions

The main conclusions to be extracted from this paper can be summarized as follows:

- The final goal of the Virtual Observatory is to facilitate the astronomical research making it easier, faster and more efficient thanks to the implementation of standard procedures. This new methodology will enhance the use of archives for astronomical research opening new fields. Its therefore compulsory to adequate the already existing astronomical archives to the VO requirements and design the new ones according to them.

- The Virtual Observatory represents a new research environment that will produce entirely new and unexpected scientific results of major significance. The Virtual Observatory is not replacing the scientist but making them much more efficient. It will allow astronomers to concentrate on what it is really important: ask scientific questions, interpret the data and publish the results rather than hunting and gathering data.

- Real VO science is already being produced. Following the pioneering work by [4], almost 300 papers have been published related to "Virtual Observatories" indicating the very high potential that the VO has to enable new astronomical research.

- The Virtual Observatory is not only for people interested in big surveys but also for all projects requiring an efficient interoperability and seamless access to several archives and services. Suppose, for instance, an astronomer who wants to analyze all the high resolution spectra in a given wavelength range for a given object. Most likely, the number of observations will not be high but the process of retrieving all the information without VO techniques would be very inefficient and time-consuming.

- The successfulness of VO from the scientific point of view strongly depends on the interaction with the scientific community. The participation of the community at different levels must be encouraged as much as possible.

References

[1] Padovani, P. 2006, in ASP Conf. Ser., Vol. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, E. Solano (San Francisco: ASP), 771
[2] Hanisch, P. 2006, in ASP Conf. Ser., Vol. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, E. Solano (San Francisco: ASP), 765

[3] McGlynn, T. A., Suchkov, A. A., Winter, E. L., Hanisch, R. J., White, R. L., Ochsenbein, F., Derriere, S., Voges, W., Corcoran, M. F., Drake, S. A., Donahue, M. 2004, ApJ 616, 1284

[4] Padovani, P., Allen, M. G., Rosati, P., Walton, NÅ. 2004, A&A 424, 545

[5] Tsalmantza, P., Kontizas, E., Cambrsy, L., Genova, F., Dapergolas, A., Kontizas, M. 2006, A&A 447, 89

[6] Suárez, O., García-Lario, P., Manchado, A., Manteiga, M., Ulla, A., Pottasch, S. R. 2006, A&A 458, 173

[7] Reid, I. N., Kirkpatrick, J. D., Liebert, J., Burrows, A., Gizis, J. E., Burgasser, A., Dahn, C. C., Monet, D., Cutri, R., Beichman, C. A., Skrutskie, M. 1999, ApJ 521, 613

[8] Burgasser, A. J., Kirkpatrick, J. D., McElwain, M. W., Cutri, R. M., Burgasser, A. J., Skrutskie, M. F. 2003, AJ 125, 850

[9] Scholz, R. D., Lodieu, N., McCaughrean, M. J. 2004, A&A 428, L25

[10] Sarro, L. M., Sánchez-Fernández, C., Giménez, A. 2006, A&A 446, 395

[11] Mayor, M., Queloz, D. 1995, Nature 378, 355

[12] Merín, B., Montesinos, B., Eiroa, C., Solano, E., Mora, A., D'Alessio, P., Calvet, N.,Oudmaijer, R. D., de Winter, D., Davies, J. K., et al. 2004, A&A 419, 301

[13] Castelli, F. Gratton, R. L. Kurucz, R. L. 1997, A&A 318, 841

[14] D'Alessio, P. Calvet, N. Hartmann, L. 2001, ApJ 553, 321