Scientific Requirements for Extremely Large Telescopes
Proceedings IAU Symposium No. 232, 2005
P. Whitelock, M. Dennefeld & B. Leibundgut, eds.

doi:10.1017/S1743921306001013

The Virtual Observatories: a major new facility for astronomy: linking ELTs, great observatories and the science community

Nicholas A. Walton[1], **Anita M. S. Richards**[2], **Paolo Padovani**[3] **and Mark G. Allen**[4]

[1]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
email: naw@ast.cam.ac.uk

[2]Jodrell Bank Observatory, University of Manchester, Macclesfield, SK11 9DL, UK
email: amsr@jb.man.ac.uk

[3]European Southern Observatory, D-85748 Garching bei München, Germany
email: ppadovan@eso.org

[4]CDS, ULP/CNRS, 67000, Strasbourg, France
email: allen@newb6.u-strasbg.fr

Abstract. We describe how the Virtual Observatory (VO) projects in Europe, the USA, Japan, and elsewhere are meeting the challenge of providing simple and efficient access to the data from the world's observational facilities, together with applications and computational resources required to support the analysis of this data. We note the pan-European Euro-VO project and its technological development VO-TECH project which are now in the process of designing the framework for comprehensive access to emerging high data volume facilities such as ESO's VISTA infrared survey telescope.

Science drivers from major new astronomical missions are helping to define the development of the VO. Scientifically this is in terms of developing systems able to meet the demands of the main science programmes shaping the ELTs. VOs must be able to handle the large data streams from the complex multiplexed instruments on the ELTs, and provide access to applications required to analyse/interpret the data. VOs must enable the effective distribution of ELT data to the global community.

Conversely, the rapid development of the Virtual Observatory, offers opportunities for major new projects such at the ELTs. This could be: in the design of their down stream data-flow systems; in terms of opening up access to 'real-time' availability of ancillary data flows; in multi-wavelength observational programmes.

We highlight these areas, and give some specific current examples of early VO usage in delivering science from, e.g. the mining of deep multi-wavelength surveys to study the high redshift universe.

Keywords. methods: data analysis, astronomical data bases

---

1. Introduction

Astronomy is currently in a productive epoch, with an explosion in the rate of new discoveries. These are made possible, in part, by the availability of high quality observations of the cosmos, from a range of major new facilities. These telescopes and missions have have opened up the sky at all wavelengths, from radio, through gamma ray. Key examples include the Hubble Space Telescope, the European Southern Observatory's (ESO) VLTs, the Keck Telescope, the VLA, WMAP, XMM-Newton, Chandra and recently Spitzer and Swift.

Supporting these facilities are the systematic surveys of the whole sky at a number of wavelengths, e.g. 2MASS in the near infrared, the Sloan Digital Sky Survey (SDSS) in the optical, The NRAO VLA Sky Survey in the radio. Additionally our understanding of the observed Universe is aided by comparison with increasingly sophisticated and physically accurate simulations (e.g. the Hubble Volume Simulations†).

However, the observational facilities that we now possess, and plan to construct (e.g. ALMA, Extremely Large Telescopes (ELTs), XEUS) present astronomy with a range of challenges. The key science programmes for these facilities currently imply that astronomers will use data from many of these facilities to tackle any particular problem, from life on distant planets, to the dark ages of the Universe. There is more data (at the peta scale), more complex data (objects can routinely be described by many hundreds of parameters), and data from more multiple sources. Further, there is increasing pressure to ensure maximum access to data from these high value facilities, and to allow maximum re-use of their data, to allow for the extraction of the maximum amount of science.

In this short paper we briefly introduce the Virtual Observatories (VO), describe the importance in the development of interoperability standards, note how individual projects are building systems to conform to these standards, and show examples of early scientific use of these systems. We close by noting how VO systems will be important in the era of ELTs in providing solutions to the data challenges that these new missions will pose.

2. Virtual Observatories and science drivers

There are a number of major Virtual Observatory initiatives currently underway, with most formal work in this area beginning in 2000. These include projects in the USA, the National Virtual Observatory (NVO), the Japanese Virtual Observatory (JVO), the Russian Virtual Observatory (RVO) and many others.

The VOs generally have a key goal to provide easier and more cost effective access to the increasingly large heterogeneous data resources available to astronomers. In this fashion, large scale astronomical science will be made easier for the astronomer. The UK AstroGrid VO project (Walton 2005), for instance, has defined its key goal as making research projects which are currently impeded by the complexity or volume of data, possible, via improved access to the data and computational systems to handle it. In this way, new science is enabled, which, whilst possible before, was in practice often not carried out due to these computational barriers.

The VO projects have generally taken exemplar science cases, which have been subsequently used to define the scope of the VO systems. For instance the Euro-VO (Sec 4) developed a Science Reference Mission‡, whilst AstroGrid drew up the 'AstroGrid Ten' science drivers (Walton *et al.* 2003). These cases invariably require multi-wavelength datasets. For instance, in the study of the high redshift universe utilising deep legacy datasets such as that from GOODS (Giavalisco *et al.* 2004) where data is available in the X-ray, Optical, Infrared and so forth.

3. Standards for a Virtual Observatory

The VO projects recognised early on that science problems in astronomy, require universal access to data and applications. Because of the distributed nature of data holdings, it would not be effective for any national VO project to develop a system which accessed

† http://www.mpa-garching.mpg.de/Virgo/
‡ http://www.euro-vo.org/internal/Avo/AvoSRM/srm.pdf

https://doi.org/10.1017/S1743921306001013 Published online by Cambridge University Press

data using interfaces which would preclude the visibility of that data to other projects. Thus it was seen that, at the low level, standards must be developed to allow for common interfaces to data. This means that an archive could make its data holdings accessible through a standard set of protocols, and then be assured that this would be visible to any VO system (assuming the compliance of that VO system to the agreed standards).

The International Virtual Observatory Alliance (IVOA, see http://www.ivoa.net) was thus formed in 2001 by the national VO projects, to provide a working forum in which these low level interoperability standards could be developed, formulated and agreed.

The main work of the IVOA is carried out in a number of Working Groups (WG). These address the key standards areas, these being derived from the IVOA's analysis of a standard VO architecture (Williams *et al.* 2004), which decomposes the system into its component building blocks, and which also shows the technical complexity underpinning any VO system.

There are IVOA WGs currently active in areas such as: *registries*, the index of resources; *content descriptors*, the astronomy dictionary to describe data; *data models*, providing described mechanisms to access astronomical data; *data access* and *VO query language*, one query language to enable querying of heterogeneous databases. Additionally there is significant work under the general heading of *grid and web services* developing standards to handle more generic issues such as how to authorise access to data using standard protocols, distributed storage and so forth. This last point ensures that proprietary data can be made securely available to only those with the required access rights. The various IVOA working groups maintain details of their activities on the IVOA wiki pages linked from http://www.ivoa.net/twiki/bin/view/IVOA/WebHome.

By adhering to the agreed IVOA interoperability standards in creating their VO systems, the user of one (e.g. the UK's AstroGrid) gains full access to the resources of the others.

4. The Euro-VO

A number of partners (ESO, CDS-Strasbourg, AstroGrid, Jodrell Bank Observatory, and Terapix-Paris) in Europe formed the successful Astrophysical Virtual Observatory Project (AVO) (http://www.euro-vo.org). This three year (2001–2004) programme produced the outline of the design for a Virtual Observatory in Europe. A key success of the AVO was the production of the AVO Science Reference Mission (SRM) document, which lays out a set of exemplar science cases which will benefit from, and in many cases require use of, VO capabilities, and thus impact on the design of the VO. The AVO SRM was produced by the AVO Science Working Group and is available on-line at http://www.euro-vo.org/internal/Avo/AvoSRM/srm.pdf.

With the completion of the AVO project, a more ambitious European Virtual Observatory (Euro-VO) initiative is underway. This is composed of a range of European consortia representing major astronomical resource providers: ESO, ESA, AstroGrid (UK), NOVA (NL), INAF (Italy), CNRS (France), LAEFF (Spain) and GAVO (Germany). It is anticipated that it will be composed of three inter-operating distributed structures:

(*a*) **Euro-VO Facility Centre:** charged with providing user support and ensuring provision of registry services

(*b*) **Euro-VO Data Centre Alliance:** providing a focus for the interface of major resource providers to the Euro-VO

(*c*) **Euro-VO Technology Centre:** developing the infrastructure and technical components creating the Euro-VO. This final element is now funded through the **VO-TECH** project, led by AstroGrid. It is organised into topics addressing:

https://doi.org/10.1017/S1743921306001013 Published online by Cambridge University Press

(*a*) Infrastructure: the VO middleware, e.g. Workflows, job execution, security, transport layer etc
(*b*) New Tools: applications for the VO, e.g. Footprint, best fitting, SED builder, etc
(*c*) Resource Discovery: finding the needle in the haystack, e.g. Building ontologies, dictionaries, resource browsers, etc
(*d*) Data Mining and Visualisation: mass scale analysis, Large scale compute, multi dimensional visualisation, etc

5. AstroGrid: the UK's Virtual Observatory initiative

In the UK, the AstroGrid (http://www.astrogrid.org) consortium, consisting of major UK institutes with expertise in data handling and computational science. was formed and begun activities late 2001. After an initial definition and requirements analysis phase, it moved into a major development phase as described in Walton *et al.* (2004). The focus from 2005 through to 2007 is now on the transition from a developmental system to a more operationally focused system, one robust and capable enough to support significant science usage.

The AstroGrid system, which additionally underpins the Euro-VO's Technology Centre infrastructure design, currently (2005, see http://software.astrogrid.org) provides a number of key capabilities:

(*a*) **Data Set Access** modules are run at each participating data provider. These are configured to allow visibility of the underlying data resources, through the AstroGrid framework. As an astronomer, one is able to perform database queries of remotely held data, utilising a standard query language (Astronomical Data Query Language, ADQL, see http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL). Alternatively they can access data held in flat files (e.g. FITS files through a standard 'Simple Image Access Protocol' (http://www.ivoa.net/Documents/latest/SIA.html)

(*b*) **Registry** of each resource, be that data, application, hardware platform (e.g. CPU, disk, network). This registry entry describes that resource, following the IVOA's interoperability standard. The registry entry is sufficiently detailed to give the system the level of knowledge about that resource, to make sensible assessments of whether it is relevant or not.

(*c*) **Common Execution Architecture** provides a standardised framework in which applications can be run within the AstroGrid system. Each application is 'registered' in the **Registry**, thus is discoverable by the astronomer.

(*d*) **Workflow** gives the ability to construct chains of processes, allowing user defined processing sequences to be carried out. This could be very simple, perhaps do a one step query of a database, or extremely complicated, with many steps. Work flows can be saved, and thus re-run at will. The **Job Execution Scheduler** takes the work flows submitted by the user, and sends the instructions to the various AstroGrid resources.

(*e*) **MySpace** is the users own secure virtual storage, accessible by them, and available to store their work, thus data, especially intermediate data, work flows, results of catalogue queries and so forth. Because MySpace servers are located on fast networks, it is ideal for storing large data sets, as these can be quickly uploaded from MySpace to applications that may be needed to work on them.

These services are available to the end user through either a web based portal or a client 'workbench' application. The system allows for the creation of communities (thus groups, access control and so forth), thus facilitating group working on shared data sets.

A range of client side applications configured to interface to the server services, such as the MySpace file storage system, are available to the astronomer to allow for

https://doi.org/10.1017/S1743921306001013 Published online by Cambridge University Press

visualisation and end stage analysis of results generated by for instance server-side large scale processing jobs. These clients include: *Aladin* (Boch *et al.* 2004), a powerful image visualisation and catalogue access tool developed at the CDS (Strasbourg) and *Topcat* (see http://www.star.bristol.ac.uk/mbt/topcat/) which enables visualisation and analysis of tabular datasets.

6. Current use of VOs

VO tools are beginning to be used to enable science. Use of the AVO demonstrator tools (see Sec 4) was recently shown, where deep survey data from the GOODS survey (Giavalisco *et al.* 2004) were mined using VO tools to discover a significantly enlarged population of obscured, high luminosity Type 2 QSOs. These indicated the presence of a larger population of massive black holes and their host active galaxies than was previously predicted. This work was enabled via the access to key X-ray data including those from Chandra, with objects from there being cross matched with deep optical survey data. The work is fully described in Padovani *et al.* (2004). The AstroGrid system is now being used to mine infrared and optical survey data to discover QSOs at redshifts greater than $z=6$ (Gonzalez-Solares *et al.* 2006). It is expected that 2006 will see many new examples of science facilitated by the use of VO systems.

7. Conclusions and relevance in the ELT era

The science community will shortly be able to benefit from fully functional VO systems to aid in their science programmes. With the approach of major new ELTs, it can be expected that VOs will be an operational reality. During the early design phases of the ELTs and their instrumentation suites, it will be necessary to ensure compatibility of data flow systems with the VO. Thus, for instance, the ELT data archives should be accessible through appropriate VO interfaces. Further to this, designers of new instruments can anticipate that VOs will provide access to mass scale distributed compute and storage systems, along with access to a wide range of applications and powerful workflow execution subsystems. Thus the end user, working perhaps in distributed communities, will be able to rapidly access the new data flows, and perform computationally intensive analysis of the new large data sets. This will ensure maximum scientific return from the ELTs.

Acknowledgements

VO-TECH is supported by the EU under the FP6 Research Infrastructures programme, contract no. 011892-DS. AstroGrid is an eScience project funded by the UK's Particle Physics and Astronomy Research Council.

References

Boch, T., Fernique, P. & Bonnarel, F. 2004, in: F. Ochsenbein, M.G. Allen & D. Egret (eds.) *Astronomical Data Analysis Software and Systems (ADASS) XIII*, ASP Conf. Proc. 314, 221

Giavalisco, M., *et al.* 2004, *ApJ* 600, L93

Gonzalez-Solares, E., *et al.* 2006, in prep.

Padovani, P., Allen, M.G., Rosati, P. & Walton, N.A. 2004, *A&A* 424, 545

Walton, N.A., Lawrence, A. & Linde, A.E. 2003, in: H.E. Payne, R.I. Jedrzejewski & R.N. Hook (eds.) *Astronomical Data Analysis Software and Systems XII*, ASP Conf. Proc. 295, 25

Walton, N.A., Lawrence, A. & Linde, A.E. 2004, *Proc. SPIE*, 5493, 146

Walton, N.A. 2005, *A&G* 46, 3.23
Williams, R., *et al.* 2004, *IVOA Note 2004-06-14*
http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch-20040615.html

Discussion

BAADE: Do other research domains have similar data analysis and integration problems? If so, are there any techniques that they may have developed which can be used by astronomers?

WALTON: Problems of data access and handling are important issues to many research disciplines. Thus, in the area of Bioinformatics, projects exist to provide standard interfaces to data (e.g. the MyGrid project in the UK (http://www.mygrid.org.uk). VOs have taken up externally developed generic software, such as Grid and Web services from the computer science community. The VO projects have only developed software to meet the specific demands of the astronomy community, or where alternatives are not appropriate.