Astronomy for the developing world
IAU Special Session no. 5, 2006
J.B. Hearnshaw and P. Martinez, eds.

doi:10.1017/S1743921307007211

Virtual observatories and developing countries

Ajit Kembhavi

Inter-University Centre for Astronomy and Astrophysics, Pune, India
email:akk@iucaa.ernet.in

Abstract. I will discuss in this article the emerging concept of virtual observatories, the efforts being made in various countries to set up these structures, and the relevance of the concept to astronomy in developing countries.

Keywords. Virtual observatories, astronomical data, International Virtual Observatory Alliance

1. Introduction

A virtual observatory is a platform for launching astronomical investigations: it provides access to huge data banks, software systems with user-friendly interfaces for data processing, analysis and visualization, and even access to computers on which the work can be carried out. Virtual observatories the world over are seamlessly networked, and their resources can be accessed over the internet by astronomers regardless of their location, expertise and the level of access to their own advanced computing facilities. Due to their nature, virtual observatories can make an immense impact on the way astronomy is done in the developing world. I will consider some of the facilities that virtual observatories provide, discuss their possible use by astronomers, and how even small groups in the developing world can contribute to the setting up of virtual observatories, and benefit from their use.

2. Astronomical data

Astronomers carry out their observations using a variety of telescopes, based on the ground or on space platforms. They also use a variety of different detectors like photographic plates, CCD cameras, radio receivers, X-ray detectors, etc. With any of these instruments, there are two basic kinds of observing strategies which are followed: (1) observations of specific targets which are of interest to a specific research project and (2) observations which survey large portions of the sky and which can be used in a variety of scientific projects over the years.

Over the last two decades there has been great progress in telescope and detector technology, and it has been possible to build many large telescopes and increasingly sensitive detectors. The large installations are extremely expensive, and the trend has been to build telescopes and detectors through collaborative efforts and to make them available to a wide community. Astronomers all over the world can therefore use advanced facilities to which they may earlier not have had access. The data obtained using these facilities is generally archived and eventually made available to the entire community, regardless of who obtained it in the first place.

2.1. *Data volumes*

Data volumes generated in ongoing and planned modern surveys can range from several hundreds of gigabytes to many tens of terabytes. It is expected that some of the surveys which will be initiated over the next few years will generate terabytes of data per day. Storing, retrieving and scientifically using these vast databases is a formidable task. It requires the joint effort of astronomers and computer professionals to adapt existing hardware and software technology, and to develop it further to meet the challenging task of making the data accessible to all potential users.

At the present time devices which can usefully store many terabytes of data are expensive and difficult to manage, and it is expected that this situation will prevail for some time to come. It is therefore not practical to store all available data in many locations in the world for it to be locally available to astronomers. Moreover, maintaining mirror copies requires regular data transfers, which cannot always be done incrementally using the available bandwidth. Maintaining large data volumes also requires constant attention by computer professionals, and not all places would have access to such expertise. It is therefore necessary to store data in strategic locations and to make just what is needed available using the internet as well as other means for data transfer. This of course brings forth the issue of providing engines and interfaces to enable users to obtain and combine data from a number of locations.

2.2. *Data variety*

Data obtained in different parts of the electromagnetic spectrum, like the optical, radio or X-ray, requires vastly different kinds of processing before it is brought to a scientifically usable form. The data are also stored using quite different hardware and software systems, and techniques have to be developed for bringing together the different structures. Even for data in the same region of the spectrum, different observers use different notations, conventions and units, and comparing data from different sources can be an exacting task. The difficulty here becomes more pronounced when much of the processing is to be carried out with computers, avoiding human intervention. The solution is to provide extensive universal standard descriptors for the data which make automated analysis feasible.

3. Virtual observatory

A virtual observatory (VO) seeks to facilitate the tasks mentioned above, but it has to go beyond simply making large amounts of data available: it has to provide query tools for the required data to be accessed from the vast store, for analysis and visualization of the data, and for data mining, which will enable new scientific discoveries to be made. The queries needed to generate the data required by a user, and the subsequent analysis, can require computing facilities which may not be available at the user's establishment. A VO would seek to provide computing resources as well, either on its own site, or through a grid linking computers located at different sites.

Attempts to establish such structures have been made in the past, with moderate success, by different observatories and institutes. But the huge increase in the volume of data now available, and the need to carry out research simultaneously in many different parts of the electromagnetic spectrum, has made it necessary to make collaborative efforts, much in the manner of joint work undertaken to develop major new ground and space based telescopes. Using the internet, even a remotely located user can access the facilities offered by a VO, and can use the data just as well as anyone else. This is a development of far reaching consequence, particularly for astronomers in developing countries.

Astronomical data are available in the form of catalogues, spectra and images. The VO will enable astronomers to use these different kinds of data simultaneously, irrespective of their location and basic nature, for a full multi-wavelength analysis. The vast quantities of data will make it possible to look for very rare objects, patterns and relationships which remained totally inaccessible when only very limited data were available. Searching for these rare features will require the development of highly sophisticated data mining techniques for the search to be completed in a reasonable time. The features found will have to be subject to analysis, and to be compared with the results of numerical simulations. The VO seeks to provide hardware and software platforms on which all these operations can be carried out.

3.1. *VO projects in the world*

A virtual observatory, as the name implies, is a comprehensive concept which embodies computer hardware and software, data, and human expertise for providing services. A VO can occupy a small space in a single university department or research institution, or can be distributed over several locations. In fact all the VOs in the world can be considered to be meshed together to be parts of a single VO, providing data and related services to anyone who may need them.

There are several VO projects in operation in different countries in the world. Some projects are large, have huge budgets, and many people working on them. Other projects have just a small number of people engaged on very specific programmes, and work with modest resources; they have nevertheless managed to provide products which have proved to be very useful to astronomers all over the world. The bigger projects include the National Virtual Observatory (NVO) of the USA, the European Virtual Observatory (EURO-VO) which brings together many European countries, some of which also have their own individual VO projects (including those in France, Germany, Italy and Spain), and ASTROGRID, which is a VO project based in Britain. The smaller projects include those based in Armenia, Australia, Canada, China, Hungary, India, Japan, Korea and Russia. Links to various VO sites can be found on the Virtual Observatory - India (VO-I) homepage (`http://vo.iucaa.ernet.in/∼voi/`) and the IVOA homepage (`http://www.ivoa.net`).

4. The International Virtual Observatory Alliance (IVOA)

As its name suggests, this is an alliance of VO projects based in various countries. It provides a forum at which astronomers, computer professionals and others who are engaged in VO activities in different countries can come together to share their experiences and resources, set up collaborations, and most importantly, develop common standards and infrastructure for data exchange and inter-operability. The VO concept involves sharing of data and resources, and the IVOA is most effective in bringing this about through continuous discussion and collaboration.

The IVOA has defined six major programmes which have to be undertaken to make progress towards building up of virtual Observatories. These are (1) REGISTRIES: These collect metadata about data resources and information services into a queryable database. The registry is distributed. A variety of industry standards is being investigated. (2) DATA MODELS: This initiative aims to define the common elements of astronomical data structures and to provide a framework to describe their relationships. (3) UNIFORM CONTENT DESCRIPTORS: These will provide the common language for metadata definitions for the VO. (4) DATA ACCESS LAYER: This provides standardized access mechanisms to distributed data objects. Initial prototypes are a Cone Search Protocol

https://doi.org/10.1017/S1743921307007211 Published online by Cambridge University Press

and a simple Image Access Protocol. (5) VO QUERY LANGUAGE: This will provide a standard query language which will go beyond the limitations of SQL. (6) VOTable: This is an XML mark-up standard for astronomical tables. Over the last few years, much progress has been made in these directions.

The IVOA has a carefully set up procedure for producing standards through a cycle of Working Drafts, Proposed Recommendations, and finally Recommendations to the international community as represented by the International Astronomical Union (IAU). The IVOA has several Working Groups and Interest Groups in various aspects of the VO, and these groups have produced an overall architectural plan for an operational VO that identifies the critical areas for current and future development of standards and technologies. Various VO groups meet at Inter-operability Meetings organized in different countries, and the community also regularly participates in important astronomical conferences to familiarize the community of the work done, and to demonstrate through real science applications the standards and tools which have been developed.

5. Some VO projects

The developmental work undertaken by different national VO projects depends on the needs of the community they serve, and the facilities and expertise available at the particular VO. A few examples are given below:

5.1. *Hungarian Virtual Observatory (HVO)*

The HVO (`http://hvo.elte.hu/en/`) has been developing various VO services, and archives of large databases. An important HVO project has been the Spectrum Service, through which spectra of galaxies and other astronomical objects are accessible through Web services, and several manipulations and transformations can be made on them. All of these services are very useful for astronomers. The HVO is creating the first dynamical synthetic spectrum service to generate spectra on the fly, using inputs from the calling service. In another HVO project, photometric red shifts have been obtained for over 100 million objects, and these have been made available through Data Release 5 of the Sloane Digital Sky Survey (SDSS), in close collaboration with the SDSS group at Johns Hopkins University.

5.2. *Russian Virtual Observatory (RVO)*

The RVO (`http://www.inasan.rssi.ru/eng/rvo/`) provides Russian astronomers effective access to international resources, and integrates Russian astronomical resources into the international VO structure. It also aims at providing observational resources to its user base when the required data are not in the archives, and to develop electronic educational resources. Several Russian astronomical observatories have unique collections of astronomical photographic plates obtained since the beginning of the past century. An important project for the RVO is to make these data available as a digital archive.

The RVO has undertaken various scientific projects to exploit the VO resources. These include creating a three-dimensional map of interstellar extinction in the Galaxy, in collaboration with NVO; compilation of the fundamental stellar parameters and and evolutionary status of close binary systems, in collaboration with the Observatory of Besançon in France; investigations of open clusters (in collaboration with German and Indian astronomers), and work on the MIGALE project together with French scientists.

https://doi.org/10.1017/S1743921307007211 Published online by Cambridge University Press

5.3. *China - Virtual Observatory (China-VO)*

China-VO (`http://www.china-vo.org/`) is a consortium initiated by the National Astronomical Observatory of China (NAOC) and the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) project, and has several partner institutions.

LAMOST is a meridian reflecting Schmidt telescope, with a large focal plane which can accommodate up to 4000 fibres, and can accumulate up to 10,000 spectra per night of objects as faint as 20.5 magnitude. China-VO will develop tools to process this huge volume of spectroscopic data automatically, and to make it available online as a VO-compatible archive.

China-VO has also been developing various services like VOfilter to transform tabular data files into OpenDocument format, to bridge the VO with current desktop applications; SkyMouse, which is an intelligent information collector which uses online astronomical resources, including VO services, and traditional web applications; and VO-IMPAT, which is an interactive imaging tool that allows the users to visualize digitized images of any part of the sky and interactively access related data and information from the Beijing Astronomical Data Center (BADC). VO-IMPAT is available as a stand alone version, which is to be installed on the user's machine.

5.4. *German Astrophysical Virtual Observatory (GAVO)*

The GAVO (`http://www.g-vo.org/portal/`) project aims at providing fast and easy access to astronomical data archives and related documentation, as well as a capability to use highly sophisticated software tools for new studies. In GAVO's pilot phase, work is concentrating on four main areas: archive technology and publication, data mining and knowledge discovery in federated astronomical archives, theory in the virtual observatory, and Grid-computing. A special area of interest to the GAVO is theory in the virtual observatory. This involves the publication of theoretical datasets, obtained through extensive numerical simulations, in ways similar to the publication of observational data, and the creation of services appropriate for the use of the theoretical data. The goal is to create an environment in which theoretical results can be used for the interpretation of observations, and observations can be used to constrain theoretical models. GAVO is pursuing a number of concrete projects and through collaborations is exploring techniques for publishing theoretical datasets. GAVO is also actively involved in the IVOA theory interest group, which aims at channelling the requirements from the theory community into the IVOA standards process.

6. Virtual Observatory - India (VO-I)

This project is based at the Inter-University Centre for Astronomy and Astrophysics located in the city of Pune in India, and its novel feature is that it is a collaboration between an astronomical institute and a major computer software company, Persistent Systems Pvt. Ltd. (PSPL), Pune, which has expertise in data management related products. The project, which is funded by the Ministry of Communications and Information Technology, as well as by the two participating organizations, was begun in early 2002 for an initial period of three years. VO-I brings together astronomers from IUCAA and other institutions and University departments in India, and software experts from PSPL.

VO-I has successfully completed a series of projects which include: (1) Development of a parser for data in VOTable format, which consists of a library of programs in C++ format which helps the user handle data in VOTable files. Versions are available which act on data in the non-streaming as well as streaming format; (2) Writers for converting data in other formats to the VOTable format; (3) A Web-based FITS file manager and (4)

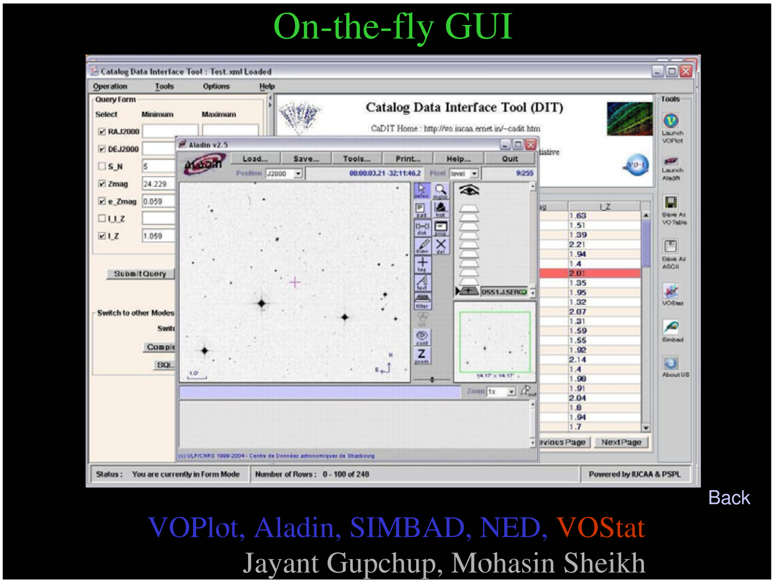


Figure 1. An example of using VOCat to study a catalogue. A specific source is selected, and the corresponding image of that source, shown as an inset, is obtained using ALADIN.

VOCat, which is a generic user interface for large data sets, and (5) VOPlot and VOStat which are major packages described below. As a part of VO-I activity, many useful large databases like the Sloane Digital Sky Survey (DR3), 2MASS, CHANDRA, etc. have been installed at IUCAA on RAID arrays, and access to these through appropriate user interfaces is being provided. An example of the use of VOCat is shown in Figure 1.

6.1. *VOPlot*

The VOPlot tool was developed in collaboration between VO-I and CDS, France, and is a menu and button driven tool for graphically visualizing data available in the form of catalogues. It is available as a web-based version fully integrated with VIZIER, which is an on-line catalogue service provided by CDS. VOPlot is also available in a stand alone version to be installed on the user's machine. VOPlot uses data in the VOTable format. It can be used to display the distribution, as a histogram, of any data field, or to plot two data fields against each other. Simple transformations can be applied to the extracted data. The visualized portion can be just a dynamically selected subset of extracted data. Some simple statistical information about the selected data can be obtained.

VOPlot goes far beyond being just a graphics tool. Each plotted point in a graph is active, and all the data related to it are available for further processing. If the data includes sky coordinates of the listed sources, it is possible to obtain the distribution of the sources in the sky in different projections. Selecting a number of points on a plot leads to the selection of the same points on the sky plot, which proves to be very useful in investigating outliers in a plot. An important feature of VOPlot is that it enables the user to pass from specific data points to images of the corresponding objects using ALADIN, which is an interactive software sky atlas developed by CDS. VOPlot is under continuous development, and is being used by the international astronomical community as a very useful new tool compatible with VO objectives. An example of the use of VOPlot is shown in Figure 2.

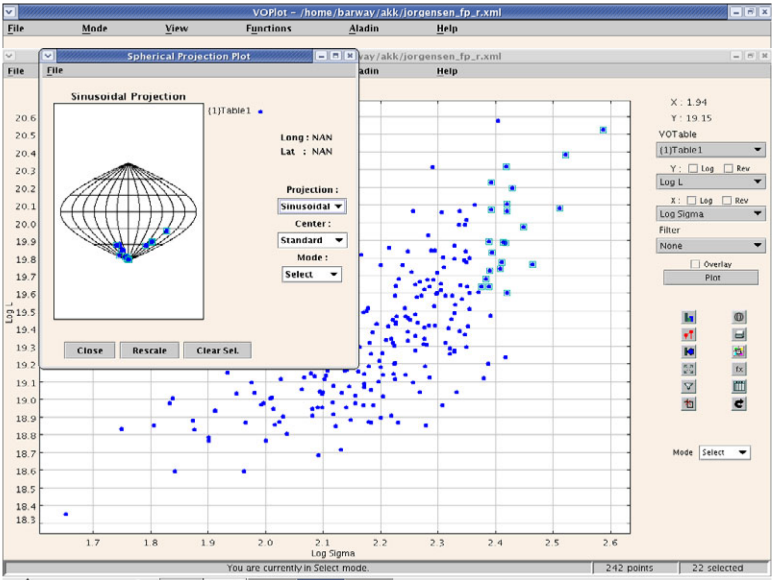


Figure 2. An example of a plot of galaxy properties made using VOPlot. The positions of the clusters from which the galaxies are drawn is shown as an inset.

Packages which are related to VOPlot, and have specialized capabilities, are VOPlot3D and VOMegaplot VOPLot3D is used in making three-dimensional plots, which can be viewed from different dynamically selected lines of sight. This allows three-dimensional correlations, like the fundamental plane of elliptical galaxies to be spotted easily. An example of the use of VOPlot3D is shown in Figure 3. VOMegaplot is used for plotting points numbering in the millions. This process would generally take substantial time to execute, and any further dynamical interaction with the data through the plot would be very slow. VOMegaplot gets over this limitation by creating a number of index files from the large data set, which are automatically used in all further interactions with that particular data set.

6.2. *VOStat*

This is a VO-compatible web-based statistical tool which was originally developed in collaboration between VO groups at Penn State University and the California Institute of Technology. A comprehensive new version of VOStat has been developed by VO-I, in collaboration with the original developers, as well as with statisticians from the University of Calcutta and the Indian Statistical Institute at Calcutta. VOStat uses the publicly available statistical package called *R*, and through an elegant interface provides access to a large number of statistical tests, which can be easily applied to data sets which are present with the user, or on remote data bases. VOStat will be made available on various sites, and will be integrated with other tools like VOPlot, so that data visualization and statistical analysis can be carried out simultaneously.

7. Science with the VO

The main aim of a VO is to facilitate astronomical research in such a manner that projects which depend on very large data sets and multi-wavelength studies can be easily undertaken. Mining large datasets will enable the discovery of rare objects, events and interesting new processes which remain obscured in the noise in small data sets.

https://doi.org/10.1017/S1743921307007211 Published online by Cambridge University Press

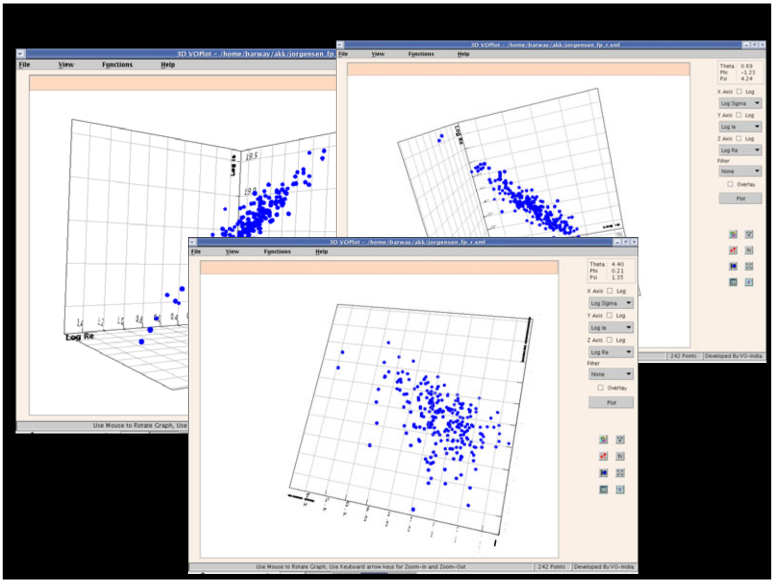


Figure 3. An example of a three-dimensional plot of galaxy properties made using VOPlot3D. Views along different lines of sight are shown.

Multi-wavelength studies will allow the form of the radiation spectrum to be better established for different classes of objects, and for correlation studies to be made across different regions of the spectrum. Most interestingly, with large samples it will be possible to examine how various properties change as a function of the types of objects being considered. For example, a sample of thousands of galaxies can be used to examine whether bulges of all types of galaxies show the same scatter from the fundamental and photometric planes.

Many scientific projects based on various VO facilities and tools are in progress at at the present time, and such projects are growing at a rapid rate. Many interesting results still remain unpublished at the time of writing, but excellent summaries of some of the work can be found in the presentations made at the Special Session on *The virtual observatory in action: new science, new technology, and next generation facilities* at the IAU XXVI General Assembly in Prague in 2006, which are available at `http://www.ivoa.net/pub/VOScienceIAUPrague/programme/index.html`.

The range of topics discussed in the special session include VO studies of (1) SDSS AGN with X-ray emission from ROSAT pointed observations, (2) the origin of soft X-ray emission in obscured AGN, (3) determination of radio spectra from catalogues and identification of GHz peaked sources, (4) the environment around QSOs with redshift of 1.3, (5) galaxy formation and evolution using multi-wavelength, multi-resolution imaging, and (6) super star clusters in nearby galaxies, (7) mapping galactic spiral arm structure, (8) discovery and characterization of brown dwarfs, (9) spectral classification of cool stars using high resolution spectra, (10) parameters for stellar population synthesis from SDSS galaxy spectra using evolution strategies, (11) photometric identification of quasars and AGN from the Sloane survey, (12) multi-wavelength properties of a sample of Texas Radio Survey steep spectrum sources and (13) near-IR properties of Spitzer selected sources. The investigations described here could all have been carried out using traditional methods, but that would have involved manually combining data from different sources after time-consuming searches through the literature, extensive programming to collate the data and to analyse it, and the use of various systems for graphical representation and statistical analysis. Use of the facilities provided by the VO makes all this very quick, and leads in fact to work which would have been very difficult to do without the new tools, and to results which otherwise would have remained obscure for a long time. The range of topics reported shows that people working in every discipline would benefit from using the VO, and that soon it will become an indispensable tool. It is very easy to gain access to these tools by navigating through various web-sites and it is also easy to learn to use these tools proficiently.

8. Virtual observatories and the developing world

Virtual observatory facilities are open to everyone, and even limited internet access is sufficient to enable any astronomer to use data and tools provided through various VO portals. Astronomers in the developing world, who have only limited facilities in their own centres can therefore participate in front-line research, and can use the data in the same manner that their colleagues in the most advanced centres do. In addition to enabling research, VOs also act as great teaching facilitators; projects at every level can be easily set up using their resources. VOs also provide excellent public outreach platforms for spreading awareness about astronomy in a very exciting way to young students as well as to the public.

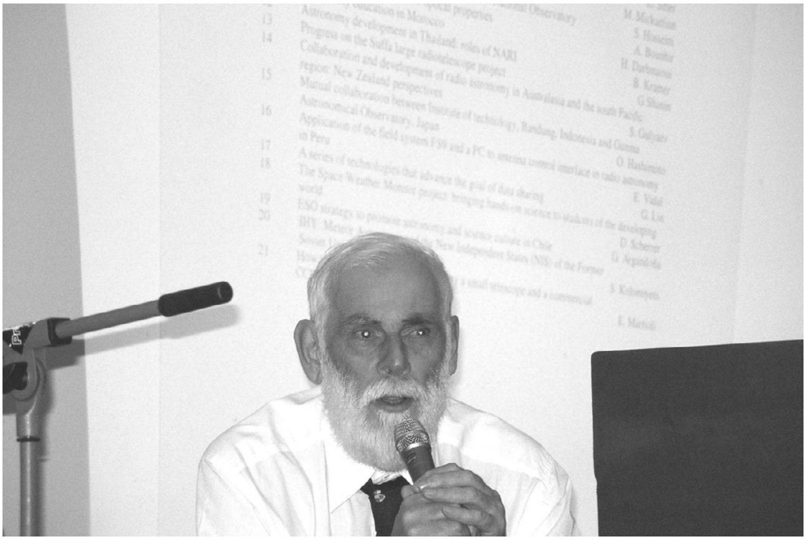
A very important aspect of VOs is that astronomers everywhere can contribute actively to their development. It is clear from the examples of some of the national projects described above that widely used VO tools have been developed by small groups of people working with limited resources. Many developing countries are now developing world class expertise in software technology, and collaboration between software specialists and astronomers in such countries can enable significant contributions to be made. Due to the highly collaborative nature of the enterprise, groups of interested persons can contribute in the manner most suited to their expertise, and in turn benefit from work of other groups, and from the data generated by advanced observing facilities the world over. Exciting new opportunities have therefore become available for astronomers everywhere, and they can both benefit from all the developments and actively contribute to sustaining the virtual observatory movement.

9. Further information

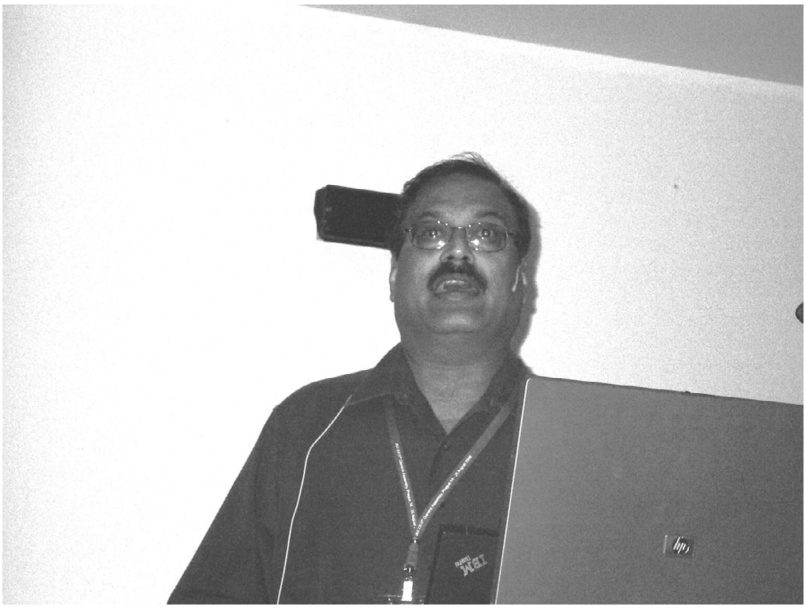
Due to the nature of this rapidly evolving field, the best resources for information and listing of the latest technical advances scientific work done using VO facilities are the web-sites of the various VO projects listed above, the links that they provide, and the papers that are regularly posted on them. Many VO sites also provide forums for discussion and users can join specialized mailing lists.

10. Acknowledgements

The author wishes to thank the Ministry of Communications and Information Technology of the Government of India for supporting the Virtual Observatory-India projects, and colleagues at IUCAA and Persistent Systems Pvt. Ltd, Pune, with whose help all the work was done. The author also wishes to thank the International Astronomical Union for providing partial support to him for attending the XXVI IAU General Assembly in Prague.



Alan Batten



Nat Gopalswamy

https://doi.org/10.1017/S1743921307007211 Published online by Cambridge University Press