
Volume 21 -
Issue 23,
Nov. 06 - 19,
2004

India's National
Magazine
from the
publishers of THE
HINDU



Frontline

[Home](#) •
[Contents](#)



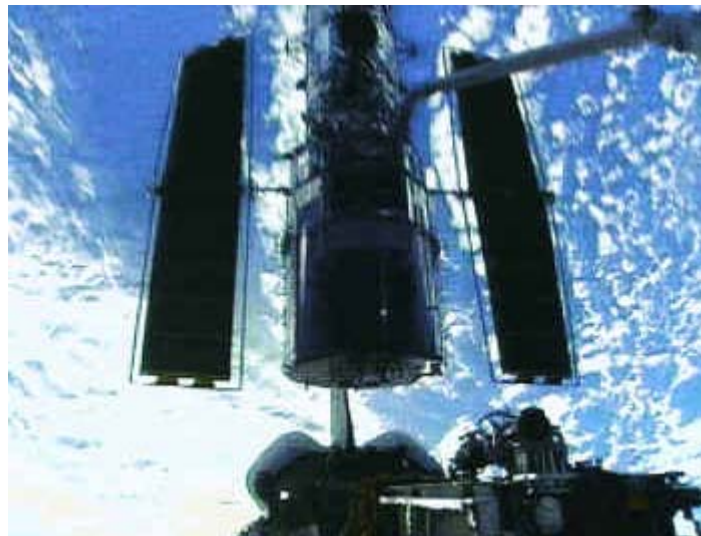
ASTRONOMY

[The universe online](#)

R. RAMACHANDRAN

Virtual Observatory, a new global endeavour, promises to bring about a major paradigm shift in the way large observational data sets in astronomy are accessed and processed.

AP



The Hubble Space Telescope above the cargo bay of the Columbia space shuttle. A file picture. The Virtual Observatory initiative is aimed at making fast-accumulating data from new technology telescopes and new astronomical platforms in space accessible to

astronomers the world over through a new software infrastructure.

MANY disciplines of science are facing the problem of data deluge from present-day experiments. This is a direct consequence of the stupendous advances in the technologies of instrumentation, detectors and sensors. For instance, in particle physics, data emerging from gigantic accelerators is of the order of petabytes (10^{15} bytes) per year. As a result, equally gigantic data processing and storage capacity is needed. Storing one petabyte per year on disk requires the computing power of thousands of personal computers.

Astronomy is no different. The universe is being "digitised" at a fantastic rate. For example, a single observation from the Hubble Space Telescope (HST) can be as much as several gigabytes (10^9 bytes). Images and related data produced by the ongoing ambitious Sloan Digital Sky Survey (SDSS), which aims to map a quarter of the sky (more than 100 million celestial objects) with unprecedented accuracy and depth using ground-based telescopes and digital cameras, will amount to a huge 40 terabytes (10^{12} bytes).

The future Large Synoptic Survey Telescope (LSST) will, in fact, outdo this by producing 5 terabytes of data each night. The data avalanche is not because the telescopes are bigger but because of the ever-increasing number of pixels in the charge coupled device (CCD) cameras that they use. Present-day telescopes deploy as many as a billion pixels in their detectors. Future telescopes will have several billion pixels.

As data streams from new technology telescopes and new astronomical platforms in space swell over the next few years, the amount of data in the archives of these experiments is doubling every year. But, as George Djorgovski, an astronomer at Caltech, has remarked: "Our understanding of the universe does not double at the same rate. There is a bottleneck somewhere. The old ways of dealing with data do not work anymore." In fact, the data doubling time is projected to come down to six months by 2007-08 as new astronomy projects begin to come on line, according to Peter Quinn of the European Southern Observatory (ESO), Garching, Germany.

The cumulative compressed data holdings of the ESO archive, according to him, will reach one petabyte by 2012. "This is

faster than the doubling time of 18 months in the performance of computer chips, the Moore's Law," he points out. More importantly, data access rates (megabytes/sec) are relatively static, presenting a major bottleneck in the transfer of large data sets. As a result, the gap between the end-user and the source of data (in terms of processing capabilities and download time) is widening. In a bid to solve this problem as regards astronomical data, a major paradigm shift in the way large observational data sets in astronomy are accessed and processed is under way.

THIS is the emerging concept of a Virtual Observatory (VO) as a global endeavour under the banner of the International Virtual Observatory Alliance (IVOA) to foster global access to astronomical data. IVOA comprises 15 separate national and international VO initiatives including that of India, called Virtual Observatory-India (VO-I).

In June 2002, at an international VO meeting in Munich, the three VO consortia of the United States, the United Kingdom and Europe came together and founded the IVOA. The directors of these three programmes - the National Virtual Observatory (NVO) of the U.S., the European Astrophysical Virtual Observatory (AVO) and the AstroGrid of the U.K. - stated IVOA's mission and also drew up a road map for tasks ahead up to 2005 towards realising the international virtual observatory. Soon, other smaller national initiatives joined the alliance. Today its membership - besides the founding three - includes VO initiatives of China, India, Japan, South Korea, Canada, France, Italy, Germany, Spain, Hungary, Russia and Australia.

The global initiative envisages large data sets and computational resources being concentrated at a number of data centres. A new software infrastructure will enable astronomers anywhere to access seamlessly and transparently these distributed (or "federated", as the astronomical community refers to it) resources analogous to the World Wide Web. However, unlike the Web, data will not be moved to the end-users, but rather accessed, processed and analysed remotely across the network of data centres and a 'computational grid'. The first steps towards realising this are to set rules for describing and manipulating both raw and processed data and to decide what kind of software tools are to be developed.

An important difference between astronomy and other areas like genomics, where a centralised repository of annotated database like GenBank is possible, is that it has to be necessarily distributed because of the manner in which scientists do astronomy. Each experiment looks at a particular region of the sky seeking a particular kind of information, with a particular spatial resolution and that too in a particular window of the electromagnetic spectrum - optical, infrared, ultraviolet, x-ray, gamma ray or radio.

Celestial objects radiate energy over an extremely wide range of wavelengths. Each of these carries important information about the nature of the object. The same physical object can appear entirely different in different windows. For instance, in the optical wavelengths, a young spiral galaxy appears smooth with spiral arms whereas in the ultraviolet it appears as many concentrated "blobs". A galaxy cluster looks like an aggregate of separate galaxies only in the optical wavelengths whereas in the x-ray the hot and diffuse gas in the intergalactic spaces also get revealed. So the sky looks very different in each of these wavelength windows.

The physical processes inside objects can only be understood by combining observations at several wavelengths, which are stored in different archives. Today we already have sky coverage in about 10 wavebands. Soon data in five more bands will become available.

Furthermore, each centre represents the data it holds in different formats and different software may be needed to read the data. Also each set of data needs to be read with additional information that is unique to the telescope and the particular experiment, like instrument calibration, atmospheric corrections and other conditions under which the observations were made. And also, unlike other disciplines, astronomical data remains "alive" for a much longer period, as with better understanding of the way in which a telescope responds, data get recalibrated and reprocessed over time.

Every experiment collects a lot more data than is needed for the particular research problem. There is, therefore, a wealth of information hidden in the unused data, which needs to be mined. Traditionally astronomers make their data public after a year and many major facilities have begun to archive these (which can be accessed on individual web sites). But more important is to know where exactly is the information that you

are looking for. This could mean endless searches in various archives over months. Even if one finds it, it is not user-friendly. It is not easy to move it around because of the sheer size. Analysis has to be done closer to the data.

The obvious solution is to make data in all the archives as well as the data that are being generated conform to the same format with whatever annotations that need to be made to describe the data so that they can be properly used and made available on different machines world-wide. This may seem an impossible task. But that is precisely what the VO initiative is attempting to do.

The VO is a system in which the vast astronomical archives and databases around the world, together with analysis tools and computational services are linked together into an integrated facility. The VO aims to achieve for astronomical data what the WWW has achieved for documents. Data from all the world's telescopes, both ground and space-based, will be available on the desktop PC to anyone, anywhere via the Internet - a World Wide Telescope that puts the universe in its multi-wavelength glory on-line, if you will.

In the framework that is being visualised, built-in software tools will enable users to search, query and mine data across archives. Astronomers will have access to a variety of tools that are currently being developed as part of the global initiative: a unified search engine to collect and aggregate data from several large archives simultaneously and a huge distributed computing resource to perform analyses close to the data.

A "Google-like search engine is not good enough," points out Andy Lawrence of the University of Edinburgh and director of the AstroGrid project in the U.K., which forms an important part of the AVO. "In Google, locating the exact page you need can take quite a while. But here we want the search facility to be much more efficient and do something more as well. You may be interested in the data on solar flares with certain characteristics. The search should be able to locate this for you. Essentially, what the web does for humans we want that done for programmes; that is, for the software to be able to pull out the correct data as well as do operations on the data the way you want."

"One of the major arguments in favour of the VO is the

efficiency of access," says Robert Hanisch of the Space Telescope Science Institute, U.S., who is also the project manager of the NVO and the Chair of IVOA. "Research publications in astronomy are now integrated into one on-line database and the astrophysical data system allows you to instantaneously search the literature. VO is the analogue for data in terms of efficiency of access. And the data you access is understandable to the software that you use on your desktop."

The VO will also serve as a grid computing network, giving astronomers, irrespective of location or resources, high performance computing capability on their desktops, for comparing billions of records from archives or running large-scale simulations using data from the archives. AstroGrid's aim is to demonstrate a prototype computing environment for a VO. "The emphasis here is not so much on achieving supercomputing capability but on an enabling environment to retrieve and mine simultaneously large data sets from different databases and overcome the I/O interface bottleneck," points out Andy Lawrence.

THE data deluge is not the only driver pushing the VO idea. It is the rise in research based on statistical studies and correlations across wavelengths. "For this kind of research, more and more astronomers are turning to online data," says Nicholas Walton of the Institute of Astronomy, Cambridge University. "But that is quite cumbersome, read about their data format... it is different every time. What VO attempts to do is to evolve standard rules for representing and archiving data. Each database must follow the same rules. There will be essentially one web page to look at all the databases. In a way VO is a response to market pressures," he adds.

Key to all this is the phrase "interoperability" of data between different holdings. This means all archives will understand the same query language, can be accessed through a uniform interface, and diverse data can be analysed by the same tools. But this is not to say that the VO will be a monolithic system. Like the Web, it will include a set of standards that make all the components of the system - data and metadata (data about data) standards, agreed protocols and methods for data exchange between archives, and standardised mix-and-match software elements - interoperable. To achieve this, however, data centres, archives, astronomy software developers and facility builders, all need to accept the new framework and

work within it.

But this is likely to evolve only slowly. According to Hanisch, nearly 70 per cent of the community is either unconvinced or indifferent to the idea at present. This is essentially because of the inherent tension between uniformity and autonomy of creativity and innovation.

Also, there is a section that feels that funding to VO might tempt governments to reduce funding for building major facilities, particularly space-based platforms. The apprehension on this count may be misplaced given that a new modern facility can cost several hundred million dollars whereas the combined budget of all the VO initiatives is a mere \$30 million.

The other fear is that VO will take the focus of research away from primary investigation and that it could end up breeding a generation of young people who sift through data without knowing about instrumentation.

"But doing actual observations with a telescope will never go away," says Hanisch. "VO does not replace the needs of doing observational astronomy. It complements these. People hopefully will realise its value when they come in, compare the data they have with other data or superpose them to derive new information.

The two will eventually find a balance," feels Hansich. Key to that will be successful demonstrations of the concept, in terms of new science that was not possible earlier, which are likely to catalyse the homogenisation of databases with time and convince an increasing number of astronomers the importance of an international VO.

"VO is dynamic, it is growing every time," points out Walton. "And that opens up the time domain in astronomy. Now we have telescopes that provide frequent all sky surveys. The trick now is to put these together, study and investigate online all these data and understanding how the sky is evolving in the time domain. It is an upcoming focus and big science may come out of such investigations," he adds.

SO, given the road map, how far has the concept been realised till now? In January 2003, the IVOA identified six major technical initiatives necessary to make progress towards the

scientific goals of VO. The first important international agreement reached by VO projects was what is called the VOTable, an eXtensible Mark-up Language (XML) standard for astronomical tables that has enabled considerable progress. The corresponding software libraries have also been developed, according to Quinn, who heads the AVO.

There has been an important contribution from India in this area for visual display of data conforming to the VOTable format as 2D and 3D plots. A software product called VOPlot has come out of a collaboration between the Inter-University Centre for Astronomy and Astrophysics (IUCAA), Pune, the nodal institution for VO-I initiative, and Persistent Systems Ltd., a Pune-based private software company.

The main pieces of a working international VO system are expected to be in place within two years, according to Jonathan McDowell of the Harvard-Smithsonian Centre of Astrophysics, which is pretty much according to the road map. "It has really gone very quickly," says Quinn.

"Standards and technologies that allow this VO concept to be realised are being developed and agreed upon on an international scale. Most importantly, interoperability has been shown through some well-defined demonstrations, including some which have yielded new science," Quinn adds.

Once VO becomes a reality, it is likely to change the sociology of astronomy. An important change will be the blurring of the now separate cultures of radio, optical, x-ray and gamma ray astronomers. More significantly, democratisation of access is at the heart of the concept and one of the stated goals of the global alliance.

By enabling access to the world's best astronomical data, the VO initiative will be particularly beneficial to astronomers from developing countries with limited resources. Even researchers from small institutions and colleges can hope to do frontline astronomy research using VO-derived data and images. More significantly, by producing new science at a fraction of cost and within months instead of several years, VO may even alter the very course of discovery about our universe.

Printer friendly [page](#)
Send this article to Friends by [E-Mail](#)

[Subscribe](#) | [Contact Us](#) | [Archives](#) | [Contents](#)
(Letters to the Editor should carry the full postal address)

[[Home](#) | [The Hindu](#) | [Business Line](#) | [Sportstar](#)]

Copyright © 2004, Frontline.

Republication or redissemination of the contents of this screen are expressly prohibited
without the written consent of Frontline
