

The heavens at your fingertips

A small but growing group of astronomers wants to put the night sky on the Internet. But will staring at a computer screen ever replace peering through a telescope? Geoff Brumfiel logs on.

Pune in India is no place for an observatory. Located 160 kilometres south-east of Mumbai, formerly Bombay, the city is pounded by monsoon rains for four months of every year. Besides heavy cloud cover and high levels of starlight-distorting humidity, a telescope in Pune would have to contend with dust and light pollution from the city's three million or so inhabitants.

But Ajit Kembhavi, an astronomer at Pune's Inter-University Centre for Astronomy and Astrophysics, sees potential for a new kind of observatory there. "We can't hope to have large telescopes and instruments," he says. "But it would be great to have access to the data they produce."

Kembhavi is one of many astronomers with a similar dream. They want to construct 'virtual observatories' (VOs) — gateways to images obtained by the world's ground- and space-based telescopes. Electronic archives of astronomical images are already available. If VOs can overcome the technical and cultural difficulties of sharing that vast amount of data, they could help to democratize astronomy. "All of a sudden, anybody with an Internet connection has access to an incredible amount of knowledge and information," says George Djorgovski, an astrophysicist at the California Institute of Technology and a founder of the National Virtual Observatory in the United States. "It's very empowering technology."

The idea of an electronic star catalogue is nothing new. The Strasbourg Astronomical Data Centre (CDS), based at the Strasbourg Astronomical Observatory in France, was created in 1972 to catalogue the increasing amount of electronic data that are produced by astronomers. Other archives followed. NASA's Extragalactic Database began

archiving information on objects outside our Galaxy in 1987, and is now linked to data from the Hubble Space Telescope.

In recent years, the amount of online data has been further boosted by a new generation of sky surveys, which use digital cameras and computer programs to photograph, identify and file astronomical objects in huge online archives. The Sloan Digital Sky Survey, for example, is using a New Mexico-based telescope to chart the position of around a million galaxies by 2005 (see *Nature* **407**, 557; 2000). Other surveys have produced images of the entire Milky Way (see above), and new telescope projects could soon be making even more data available (see "Brace yourself for the data deluge", page 264).

Searching for progress

Researchers at the CDS began by cataloguing objects that appeared in the published literature. But as time went on, they and other archivers developed tools for analysing their databases. By using these techniques to compare images from the Sloan archive for example, researchers have revealed details about how the Universe is structured, such as the way in which galaxies cluster together. The archives and search tools are now immensely popular. Françoise Genova, director of the CDS, says that her database receives 10,000 hits a day — a huge number for a site that is used only by professional astronomers.

Peter Quinn, head of data management for the European Southern Observatory (ESO), says that the primary role of VOs will be to bring together the flood of data produced by the dozen or so sky surveys such as the Sloan, as well as those in the archives of satellites and ground-based telescopes. Three extensive VOs

are being developed — Europe's Astrophysical Virtual Observatory, headed by Quinn, the US National Virtual Observatory, and Britain's AstroGrid — with several smaller projects also taking shape (see table, opposite).

All are still currently at the prototype stage. Most are likely to use a single, easy-to-use interface to comb existing archives. In many cases, the researcher might not even be aware of all the databases that the VO is searching until it sends back the data. Some, such as the German Astrophysical Virtual Observatory, will also run their own databases, in this case of images supplied by the country's researchers.

Although VOs are still a long way from producing results, survey databases offer a taste of what they can offer researchers who lack access to cutting-edge equipment, says Stephen Landy, a physics professor at the College of William and Mary in Williamsburg, Virginia. "Researchers at William and Mary have no access to large telescopes," he says. Landy is interested in whether the Universe is flat — whether it obeys the normal rules of geometry. Studies of the large-scale distribution of galaxies should shed light on this issue, and Landy can do this analysis without a telescope by downloading the positions of thousands of galaxies from the databases of the Sloan survey and the Australia-based Two-Degree Field Galaxy Redshift Survey.

VOs, Landy says, will open even more doors for researchers at small institutions because they will provide a single access point to numerous databases. It will be "a hundred times easier" than using multiple uncoordinated databases, he predicts.

Combining different databases would also increase the possibilities for astronomical investigations. "The sheer size and quality of



New vision: Ajit Kembhavi believes that Pune is the perfect place to host a virtual observatory.

these data sets lets you ask new kinds of questions," says Djorgovski. Searching for brown dwarfs, objects somewhere between a planet and a star in terms of size and mass, is one example. These dim objects emit optical and infrared radiation, so a simultaneous search of archives from optical and infrared telescopes would be a powerful tool for finding them.

In a similar way, trawls of X-ray, optical and radio databases will generate insights into galaxies with centres that produce huge amounts of radiation. These central regions emit radiation over a broader range of wavelengths than other objects, so many different types of telescope can be used to study them. Each wavelength has its own advantages and disadvantages. Radio waves, for example, are useful because they can pass through the dust that obscures some galactic cores, but they are emitted by only 10% of active galactic centres. Combining data from many wavelengths should give a much better picture of the properties of these galaxies, Djorgovski says.

Creating the tools needed to run such searches is one challenge faced by VO advocates. Software designed to pull information and correlations out of databases already

exists. "You can go on the web and buy software that does exactly this kind of thing now, except it's tuned to deal with hundreds or thousands of data points," Djorgovski points out. But to be effective, a VO must be able to sift through the spectral characteristics of billions of stars. "That is a major computer-engineering problem," he says.

Alex Szalay, an astronomer at Johns Hopkins University in Baltimore, Maryland, is working with other researchers to develop faster and smarter searching tools for the Sloan database, but admits that his collaboration has a way to go. "So far we haven't quite risen to the problem," he says. He hopes to improve matters by working with computer experts in public and commercial organizations, including Microsoft.

But developing search tools is not the biggest technical hurdle facing VOs. Because a patchwork of databases is involved, standardizing data quality is a significant challenge.

The resolution of an image depends upon the telescope used and the weather at the time it was taken. Likewise, different VO users will have different resolution requirements depending on their research. Accord-

ing to Quinn, every image in the VO will have to include detailed information on the conditions under which it was obtained, so that researchers can decide whether the data are accurate enough to be used in their work.

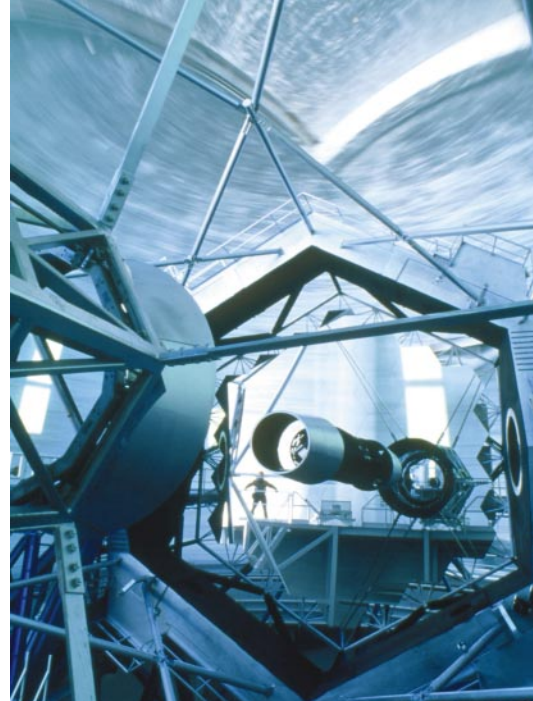
Data and time

Most sky surveys already attach such information to their images. But Sandra Faber, an astronomer at the University of California, Santa Cruz, says that it will be difficult to get astronomers into the habit of recording these 'metadata' when they add their data to databases maintained at telescope centres. Ground-based optical telescopes, for example, require lengthy calibrations, and recording the fine details of that process would be a drain on astronomers' often-limited observing time. "This is a very considerable overhead to the actual observers," she says. Some facilities, such as the ESO's Very Large Telescope at the Paranal Observatory in Atacama, Chile, already require users to record metadata. VO advocates are pinning their hopes on other observatories doing the same once VOs become more popular.

The problem of metadata underscores a much broader cultural challenge facing the VOs: astronomy has historically been a solitary activity. "In a typical ground-based observatory, if you get three nights on the telescope, those three nights are yours," says Andrew Lawrence, an astronomer at the University of Edinburgh and head of AstroGrid. So, adds Faber, are the data. "There is a general unwillingness to give away what you work hard to get," she says. Astronomers at

Starry line-up: the wave of virtual observatories being launched around the world

Project	Location	Funding (US\$ millions)	Grant duration (years)
National Virtual Observatory (VO)	United States	10	5
AstroGrid	United Kingdom	7.8	3
Astrophysical VO	Europe	3.9	3
Canadian VO	Canada	2.8	2
Indian VO	India	1.0	3
German Astrophysical VO	Germany	0.6	2.5
Australian VO	Australia	0.3	Not available
Japanese VO	Japan	0.3	1
Russian VO	Russia	0.01	3



Knowledge base: Alex Szalay hopes that facilities such as the Keck telescope (right), which currently has no archive, will develop searchable databases.

university-operated observatories are under no obligation to share their data with the wider community, for example. Those at public observatories are often required to share a copy of the raw data, but they are not necessarily required to provide the metadata that would make them useful to others.

In the United States, the situation is not helped by the fact that many of the largest and most modern observatories lack well-developed public databases. The Hawaii-based W. M. Keck Observatory has no archive, for example. And the publicly run facilities of the US National Radio Astronomy Observatory

store their data on digital data tapes that must be accessed manually upon request. Moves are being made towards creating online archives at both observatories. NASA, for example, plans to stipulate that all research at Keck sponsored by the agency should be deposited in a public archive. But progress is slow because of a lack of funding. "It really comes down to money," says Faber. "Money is needed to support the data-taking and the actual making of the archives, and then to make a usable user interface."

The cost of developing an archive should fall as they become more widely used, when the software can be copied or adapted from

existing databases. And as more centres sign up, facilities such as Keck may find themselves under pressure to set up archives. "Supposing that a whole bunch of these observatories really did get on board," says Faber. "What would happen to Keck? I think it would become a pariah."

But so far, the leaders of the VO projects say that the community's response has been lukewarm. "People tend to think, 'yes this looks good, tell me when it works,'" says Lawrence. To try to maximize their popularity, VO leaders will initially aim to incorporate the most popular databases and tools into their software. If they succeed in capturing the attention of researchers, innovative research should follow. And if other researchers can see good results flowing from scientists who use VOs, they are likely to want to get involved themselves. "I think its going to be a community-enlightenment process to get this running," says Robert Hanisch of the Space Telescope Science Institute in Baltimore, who heads the US National Virtual Observatory.

But for the hundreds of astronomers at universities throughout India, the promise is too great to ignore, says Kembhavi. These researchers may never be able to gather funds for a world-class telescope, but a VO will mean that a telephone line is just as good. "Most of them can afford to buy a PC with their own money and for a very small sum they can access the Internet," says Kembhavi. "A virtual observatory will empower them to conduct relevant, high-quality research." ■

Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.

US National Virtual Observatory

♦ www.us-vo.org

Astrophysical Virtual Observatory

♦ www.euro-vo.org

AstroGrid

♦ www.astrogrid.org

Brace yourself for the data deluge

The mass of data already produced by sky surveys could be dwarfed by the output of a proposed new telescope.

A consortium of US astronomers is currently developing plans for the 8.4-metre Large Scale Synoptic Survey Telescope (LSST). The \$120-million project will use a special three-mirror design to reduce distortion around the edge of its images, creating an exceptionally wide field of view. Together with an advanced digital camera, this will allow the telescope to catalogue 18 terabytes (1.8×10^{13} bytes) of data each evening, enough to survey the entire sky in only three nights, says project spokesman Tony Tyson, an astronomer at Bell Laboratories in Murray Hill, New Jersey.

The speed with which the LSST could scan the sky would give astronomers an almost real-time picture of the heavens, allowing them to spot unusual changes such as supernovae and near-Earth asteroids.

"There's a huge laundry list of exciting astrophysical science to do," says Tyson.

Detecting matter that does not emit light — 'dark matter' — is another item on the list. Dark matter is intrinsically difficult to study because it does not show up directly in an image of the sky. But dense clusters of dark matter exert a gravitational pull that can distort the light from galaxies behind them. By measuring these distortions, Tyson believes that the LSST can shed light on dark matter's

distribution throughout the Universe. In doing so, it could also provide a measure of 'dark energy', the mysterious force that seems to be pushing the Universe apart.

Within a decade, Tyson says, the LSST could catalogue 15 petabytes (1.5×10^{16} bytes) of data, enough to fill about 1.5 million compact discs. That will be far too much for LSST collaborators to hoard, he says. So the project calls for the group to put all of their data online almost as quickly as they can record them — a promise of openness that VO advocates hope will soon be common.

Tyson says that the LSST team plans to raise funds from private and public sources, and hopes to complete the project within ten years.