

ISSUES AND EVENTS

Astronomers Envision Linking World Data Archives

The National Virtual Observatory is neither national, nor virtual, nor an observatory," says George Djorgovski, a Caltech astronomy professor who is involved in developing the NVO, one of a half dozen or so such projects worldwide. The emerging virtual observatories aim to handle the humongous datasets that are proliferating in astronomy and will have built-in software tools for people to mine and query data across archives. Eventually, the various projects will be knit together so that ground- and space-based astronomy archives from around the world are linked and accessible to all.

In its 2000 decadal survey, the US astronomy community named the NVO its highest priority in the \$100 million-or-less category. Last November, NSF awarded the NVO \$10 million over five years. At about the same time, the European Commission granted 5 million euros (roughly \$4.5 million) over three years to the Astrophysical Virtual Observatory, a consortium in Europe. AstroGrid in the UK, an AVO partner, has another £5 million (\$7.3 million). Smaller projects are also getting started in Australia, Germany, India, and Japan. To win more money and broaden support within the astronomy community, the teams have to prove the virtual observatory concept is a good investment and will lead to new science.

The first steps are to set rules for describing and manipulating both raw and processed data and to decide what sorts of computational software to incorporate. The virtual observatories will consist of hardware and software frameworks that can cope with both spectroscopic data and images in all wavelengths. The various projects will develop mutually compatible systems, says AVO leader Peter Quinn of the European Southern Observatory (ESO). "Everybody realizes that it's a global virtual observatory. It's not in any way, shape, or form bounded by boundaries."

"The virtual observatory concept is

Virtual observatories could change the sociology of astronomy. But to bring them about, financial, political, technical, and sociological challenges must first be met.

a very ambitious project and a great challenge to astrophysicists, mathematicians, statisticians, and specialists in all areas of computer science to work together," says Wolfgang Voges of the Max Planck Institute for Extraterrestrial Physics in Garching. Voges heads up efforts for the German Astrophysical Virtual Observatory, which is intended to support astronomy research in Germany and coordinate that country's participation in creating a global virtual observatory. "There will be a fresh wind blowing through the graveyard of old and unused data. And there should be no newly formed graveyards in the future if we succeed in finding a globally accepted data format," he says.

Pixel glut

Data is doubling every year or two in astronomy. But, says Djorgovski, "our understanding of the universe doesn't double at the same rate. There is a bottleneck somewhere. The old ways of

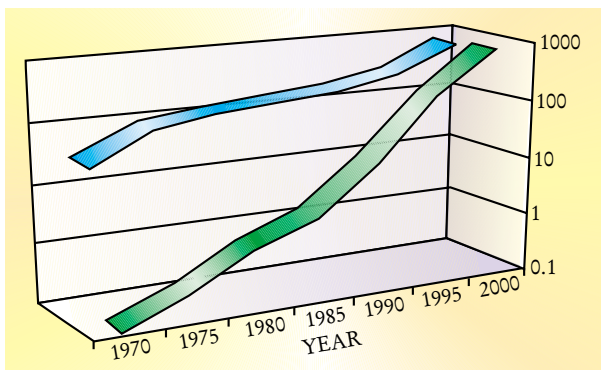
dealing with data don't work anymore."

"Even though the data is public, it's hard to get to it," says Alex Szalay of Johns Hopkins University, who oversees the Sloan Digital Sky Survey archive and is one of the leaders of the NVO, whose participants represent 17 institutions. "Every set of data has quirks. It's there. It's stored. It's archived. But it's not friendly. We can't shove it around at will. It's too much. We have to do analysis closer to the data." With the virtual observatories, he adds, researchers could search and analyze data from their desktops without having to download unwieldy datasets, which would stay in their home archives.

Sloan, currently the most prolific astronomy experiment, will produce 40 terabytes of data in five years as it maps a quarter of the sky and measures distances to more than a million galaxies and quasars. That figure will be dwarfed by future surveys. The Large Synoptic Survey Telescope, for example, is slated to make a movie of the visible sky every week starting in 2008, racking up about 5 terabytes each night. Says Szalay, "It's the CCD [charge-coupled device] cameras. It's the pixels. The data avalanche is not because telescopes are bigger."

When it comes to data avalanches, astronomy is not alone: The problem is shared by high-energy physics, genomics, neuroscience, and geophysics, to name a few (see the article on page 42). "There is synergy between us and the Grid crowd," says Szalay. "But there are differences. In particle physics, they tend to analyze individual events one by one, they don't want to cross-correlate, whereas in astronomy, a hierarchical data flow is not possible because different telescopes take different data in different formats."

Grid technology is just one of the developments paving the way for virtual observatories. In astronomy, unlike most fields, data headers have been standardized for more than two decades; space missions, at least, keep excellent, searchable records; and many archive centers link data from



THE NUMBER OF PIXELS in charge-coupled device cameras grew 3000-fold from 1975 to 2000 (green band, in megapixels). Over the same period, the total area of ground-based telescopes 3 meters in diameter or larger increased by a factor of 30 (blue band, in meters squared). (Courtesy of Alex Szalay, Johns Hopkins University.)

multiple experiments. And, last August, using Astrovirtel, a fledgling European virtual observatory and precursor to the AVO, scientists determined that a newly discovered asteroid, 2001 KX76, has a diameter of about 1200 km, making it the largest asteroid yet spotted in the Solar System.

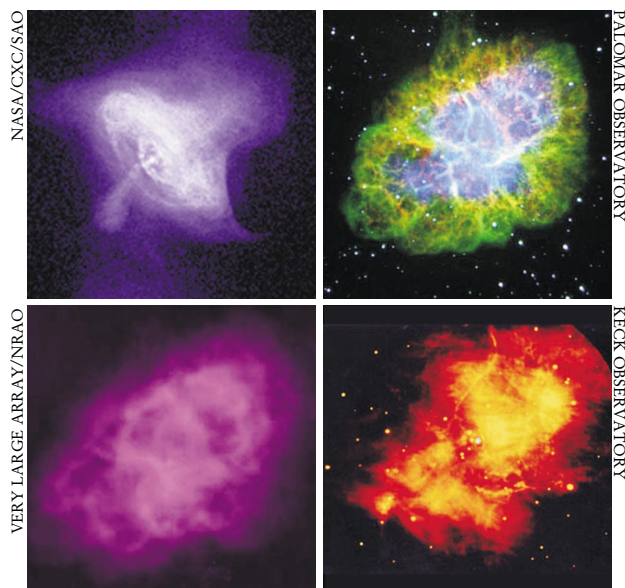
"I think the global virtual observatory is a fantastic idea," says UK Astronomer Royal Martin Rees of the University of Cambridge. "More than 99% of rare events—gamma-ray bursts, stars swallowed by black holes, and all different kinds of variable stars—are completely missed now."

The data glut is not the only driver for the virtual observatories. Statistical studies and correlations across wavelengths are both on the rise, and both stand to gain from combing through and comparing data. For example, only after astronomers took pictures in the infrared did they realize that most spiral galaxies have arms. "Our physical intuition about how these galaxies were made was influenced by the wavelengths of the images we took," says Quinn. "The universe doesn't know about our arbitrary way of dividing things by wavelength," adds Djorgovski. "That's the exciting part for me. This IT [information technology] revolution is enabling us to tackle the sky in a whole new organized way. It will enable science we couldn't do otherwise."

Common concerns

Not all astronomers embrace the idea of virtual observatories, however. Common concerns—even among supporters—include that too much money will be sunk into software; that archives at ground-based observatories should be brought up to snuff before thinking about linking them; and that users won't be able to trust the processed data they call up. Some people also worry that funders might view the virtual observatories as a replacement for actual facilities and that the know-how needed to use telescopes will disappear.

One of the more outspoken critics is Bob Fosbury of the European Space Agency—itsself an AVO member—who says, "a virtual observatory is a bad move for Europe. The danger here is that if the European AVO is funded and the observing projects are not, Europe ends up supporting the derivative proposals while others do the cutting edge science. Archive science



COMPARING IMAGES of the Crab Nebula in (clockwise from upper left, not to scale) x-ray, optical, infrared, and radio wavelengths reveals that the most energetic radiation comes from the pulsar, or rapidly spinning neutron star, at the center. As electrons from the pulsar move away and lose energy with time, they radiate longer wavelengths, with radio waves defining the full extent of the nebula.

is very valuable for supporting a scientific investigation. But archive science takes the eye off the ball. The ball is doing a primary investigation." What's more, says Fosbury, "we are likely to breed a generation of young people who sift through data without knowing about instruments. That's dangerous."

"The whole idea behind a national virtual observatory is to tie in different datasets," says Anne Kinney, who, as head of NASA's astronomy and physics division, will be key in obtaining funding for the NVO from that agency. "NASA says it's good that NSF is giving the first money, because they are behind on archiving," she says, referring to the National Optical Astronomical Observatory (NOAO) and most other NSF-funded ground-based facilities. "I think NSF should go back a step and give money for archiving—do right by science."

The astronomy community is split over archiving. "People give it lip service," says John Huchra of the Harvard-Smithsonian Center for Astrophysics. "But when it comes to choosing between archiving and a new instrument, they're not willing to pay for archiving." The notable exception to the typically lousy records—

often just backup tapes—at ground-based telescopes is ESO's Very Large Telescope in Chile, which allots time for calibration and keeps data records like a space-based experiment. The National Radio Astronomy Observatory is stepping up its archiving, and NOAO is just getting started; both are partners in the NVO. Meanwhile, at private telescopes such as the Kecks, the money problem is compounded by a culture of keeping data under wraps. Indeed, the emergence of virtual observatories is lending urgency to ongoing discussions about proprietary data (see PHYSICS TODAY, November 1998, page 52).

As for controlling data quality, Szalay says, "we've spent days and days already arguing about this. We don't want to be a police force." Adds Djorgovski, "The free market idea is the way it works. If some dataset is really no good, it will quickly become known. People don't operate in a vacuum. Part of the skill is to make judgments."

Sociological changes

"The NVO will entirely change the sociology of astronomy," predicts Szalay. Among the changes he expects are a blurring of the now separate cultures of radio, optical, x-ray, and gamma-ray astronomers and a democratization of astronomy, in that researchers from poorly funded colleges or countries would have easier access to fresh data. The virtual observatory, adds NOAO Associate Director for Science Stephen Strom, "will lead to two extreme changes. There will be more teamwork to create surveys and use them. Ironically, I also think it will change the way an individual looks at the world, because you can really pose questions at your desktop." The virtual observatory is essential, adds Strom, "but it will come about piecewise, and not without struggle, until people begin to clamor for making everything work together."

Virtual observatory enthusiasts want to start now with datasets that are already in good shape, and they hope that others will want to make their data compliant when they see the power of linking and querying archives. "The biggest challenge is psychological—getting people to want to join," says Caltech's Roy Williams, computer systems architect for the NVO. "You can think of it as

leveraging previous and future science investments."

In two years, says Williams, "I hope the NVO will be at a stage where 10 to 20 collections are federated, and

that one might have a dozen or so interesting projects going. In five years, the framework will have been used enough, with enough different datasets and computing tools, that it

will start to be part of the general infrastructure for astronomers. And in 10 years, nobody will mention it, people will take it for granted."

TONI FEDER

Federal R&D Tops \$103 Billion, but Physics Stays Flat

When President Bush released his 2002 budget request early last April, proposed science funding was so heavily weighted toward medical and defense research that most other R&D agencies considered themselves fortunate to just stay even with inflation. House Science Committee Chairman Sherwood Boehlert (R-N.Y.), said the science funding numbers were so bad that NSF should be read "not sufficient funds" instead of "National Science Foundation."

Then, thanks to what Office of Management and Budget Director Mitch Daniels described in a National Press Club speech in late November as "converging factors"—primarily the war on terrorism and the recession—the federal government returned to deficit spending. The rush to spend "had all of the characteristics of a jailbreak, with people running in all directions," he said. Those people, both in the administration and on Capitol Hill, were carrying supplemental spending proposals, and the result for science has been significant.

Federal R&D appropriations climbed to a record \$103.7 billion for fiscal year 2002, an increase of \$12.3 billion, or 13.5%, over 2001. The National Institutes of Health and the Department of Defense still lead the R&D spending parade, as they did in the original Bush budget proposal. NIH's appropriation of \$23.6 billion keeps it on track to double its budget over the five-year period ending in 2003. R&D at DOD jumped 17.3%, or \$7.4 billion—from \$42.7 billion in 2001 to \$50.1 billion in 2002. The DOD increase was \$1.7 billion more than Bush asked for.

But with a near tripling of counterterrorism spending, set at \$1.5 billion for 2002, and Social Security funds no longer viewed as sacrosanct, money has flowed, albeit less copiously, to almost every other R&D program as well. According to an American Association for the Advancement of Science analysis of the final 2002 appropriations, "all federal [science] agencies received increases for their research portfolios."

While science in general did better than most observers believed it would early on, physics funding at the

Four years of federal budget surpluses are giving way to deficit spending, thanks in large part to terrorists, war, and recession. As the unanticipated spending began late in the budget process, overall science appropriations increased significantly.

Department of Energy (DOE) remained fairly flat. High-energy physics, nuclear physics, basic energy sciences, and fusion-energy research programs are all held at about their 2001 levels.

Congress passed the last of the appropriations bills on 20 December, nearly three months after the fiscal year began. The 2002 budget can be seen as a transition budget, one that begins to reflect the new priorities triggered by the 11 September attacks and the ongoing war on terrorism. The following are R&D highlights from key science funding agencies.

▷ **NSF.** After receiving a record increase of about 13.0% in 2001, NSF's proposed 2002 funding increase was a meager 1.3%. Even that small increase was deceptive, because the administration was emphasizing education and human resource programs at NSF. The funding for R&D programs at NSF would have declined 1.6% under the Bush proposal.

In the final funding bill, NSF instead received an overall increase of 8.4%, or \$373 million, compared to 2001. Research and related activities, the division that funds most of NSF's R&D, received an increase of 7.7%, or \$256 million, over 2001. This boost means that, with the exception of biosciences and social sciences, NSF's research divisions received increases greater than 8.0%. Mathematical and physical sciences received an 8.4% increase to more than \$922 million. That amount funds physics and astronomical sciences, and earmarks \$4 million for the Telescope Systems Instrumentation Program. The budget document said NSF is expected to make upgrading its astronomical facilities and equipment, including the Green Bank Observatory and Robert C. Byrd Telescope in West Virginia, and the Very Large Array radio telescope in New Mexico, a priority.

The US Polar Research Program received a 9.0%, or \$18.9 million, increase, along with a 9.3%, or \$5.8 million, increase in Antarctic logistical support programs.

Information technology and nanotechnology, two research areas often cited by administration officials as being very important, received major boosts from congressional appropriators. Information technology received \$180 million, compared to \$155 million in 2001, and nanotechnology increased from \$150 million to \$199 million. Presidential Science Adviser John Marburger has hinted that those two areas will receive significant increases again in the 2003 budget.

NSF's major research equipment account, which funds construction of large scientific facilities, received \$139 million, a 14.1% increase over 2001 and \$42 million more than Bush requested. Of that amount, \$12.5 million goes to the Atacama Large Millimeter Array radio telescope project. The High-Performance Instrumented Airborne Platform for Environmental Research gets \$35 million instead of being eliminated under the Bush budget proposal. There is also \$15 million for the IceCube Neutrino Detector project, an Antarctic facility that wasn't included in the administration's budget request. The final budget also includes \$875 million for NSF's education and human resources programs, an 11.4% boost from 2001. The new Math and Science Partnerships program, championed by the Bush administration, received \$160 million.

▷ **DOE.** Science at DOE was funded at \$3 billion, a 2.1% increase of \$63 million over last year's budget. Both fusion energy at \$245 million and nuclear physics at \$355 million received about the same funding as in 2001. The congressional conferees who put together the final numbers urged DOE to use the nuclear physics money to enhance the operation of the Relativistic Heavy Ion Collider at Brookhaven National Laboratory, as well as improve operations at the Jefferson National Accelerator Facility in Virginia.

High-energy physics funding was up only 0.6%, to \$706 million, and the