The International Virtual Observatory

Robert Hanisch, Project Manager, US National Virtual Observatory, and Chair, International Virtual Observatory Alliance; Space Telescope Science Institute, Baltimore, Maryland, USA

Peter Quinn, Director, Astrophysical Virtual Observatory; European Southern Observatory, Garching, Germany

Introduction

Astronomy faces a data avalanche. Breakthroughs in telescope, detector, and computer technology allow astronomical instruments to produce terabytes of images and catalogs (Figs. 1 and 2). These datasets will cover the sky in different wavebands, from gamma- and X-rays, optical, infrared, through to radio. In a few years it will be easier to "dial-up" a part of the sky than wait many months to access a telescope. With the advent of inexpensive storage technologies and the availability of high-speed networks, the concept of multi-terabyte on-line databases interoperating seamlessly is no longer outlandish. More and more catalogs will be interlinked, query engines will become more and more sophisticated, and the research results from on-line data will be just as rich as that from "real" telescopes. Moore's law is driving astronomy even further: new survey telescopes now being planned will image the entire sky every few days and yield data volumes measured in petabytes. These technological developments will fundamentally change the way astronomy is done. These changes will have dramatic effects on the sociology of astronomy itself.

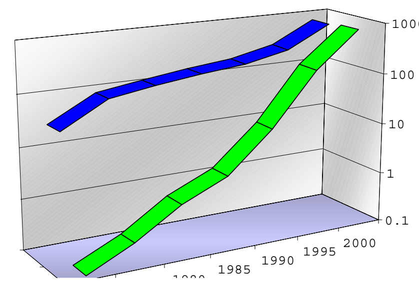

Figure 1. The total area of astronomical telescopes in m^2, and CCDs measured in Gigapixels, over the last 25 years. The number of pixels and the data double every year.

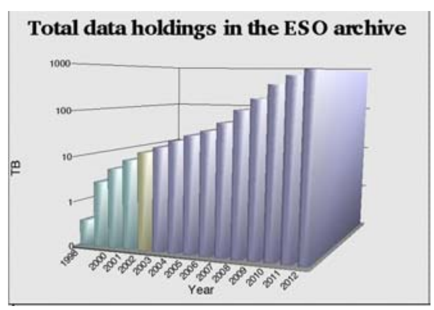

Figure 2. The cumulative compressed data holdings of the ESO archive will reach 1 PetaByte by 2012. The change is slope around 2007/8 is due to the simultaneous availability of 2nd generation instrumentation, the wide field VISTA IR camera and ALMA

Over the past two years the concept of the Virtual Observatory has emerged rapidly to address the data management, analysis, distribution and interoperability challenges. 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. Twelve VO projects are now funded through national and international programs (Table 1), and all projects work together under the International Virtual Observatory Alliance to share expertise and develop common standards and infrastructures for data exchange and interoperability. Astronomy stands at the edge of a new frontier for discovery, enabled by modern information technology (such as the Grid) and by political and technical collaboration among international partners.

International Virtual Observatory Alliance Partners May 2003
http://www.ivoa.net

Project	URL
AstroGrid (UK)	http://www.astrogrid.org
Australian Virtual Observatory	http://avo.atnf.csiro.au
Astrophysical Virtual Observatory (EU)	http://www.euro-vo.org
Virtual Observatory of China	http://www.china-vo.org
Canadian Virtual Observatory	http://services.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/cvo/
German Astrophysical Virtual Observatory	http://www.g-vo.org/
Italian Data Grid for Astronomical Research	http://wwwas.oat.ts.astro.it/idgar/IDGAR-home.htm
Japanese Virtual Observatory	http://jvo.nao.ac.jp/
Korean Virtual Observatory	http://kvo.kao.re.kr/
National Virtual Observatory (USA)	http://us-vo.org/
Russian Virtual Observatory	http://www.inasan.rssi.ru/eng/rvo/
Virtual Observatory of India	http://vo.iucaa.ernet.in/~voi/

Science Goals, Science Guidance

By providing the tools to assemble and explore massive data sets quickly, the VO will facilitate and enable a broad range of science. It will make practical studies which otherwise would require so much time and resources that they would be effectively impossible. Federating massive data sets over a broad range of wavelengths, spatial scales, and temporal intervals will be especially fruitful. This will minimize the selection effects that inevitably affect any given observation or survey and will reveal new knowledge that is present in the data but cannot be recognized in any individual data set. VO-based studies would include systematic explorations of the large-scale structure of the Universe, the structure of our Galaxy, AGN populations in the universe, variability on a range of time scales, wavelengths, and flux levels, and other, heretofore poorly known portions of the observable parameter space. The VO will also enable searches for rare, unusual, or even completely new types of astrophysical objects and phenomena. For the first time, we will be able to test the results of massive numerical simulations with equally voluminous and complex data sets. The VO-enabled studies will span the range from major, key project level efforts to supporting data and sample selection for new, focused studies of interesting types of targets, both for the space-based and major ground-based observatories.

Many of the IVOA projects have active Science Working Groups consisting of astronomers from a broad cross-section of the community representing optical, radio, high energy, space and ground-based astronomy. In some cases, IVOA projects have cross-membership of these groups. The common focus of SWGs is to form a clear picture of the scientific requirements for an operational virtual observatory. These requirements are a mix of new technologies and algorithmic capabilities as well as new standards that address fundamental issues of publishing data in the IVO (e.g., guidelines for describing all the aspects of data quality). Individual SWGs have identified the need for a design reference mission for the IVO which will capture the set of tools astronomers will need to do new science in the IVO as well as defining initial science cases and projects that can be run in the IVO to test and refine capabilities.

One of the main results of the SWG discussions to date has been a clarification of the role and nature of the IVO in modern astronomy. In the next few years, projects within the IVOA will build a new astronomical infrastructure with the guidance of the research community and utilizing emerging technologies. Once this fundamental infrastructure of standards and technologies is in place, the international astronomical community will be empowered to create new research programs and publish their data and results in a more pervasive and scientifically useful manner than was ever possible before. The IVO will not be an end in itself or an adjudicator, but a facilitator and an empowering agent in modern astronomy.

Demonstrations

A number of the VO projects are using science prototypes, or demonstration projects, to help guide technical developments and show the user community the benefits of the federated archives, catalogs, and computational services. The AVO's first demonstration, which debuted in January, provides astronomers with a powerful interface for exploring the GOODS (Great Observatories Origins Deep Survey) multi-spectral images, for obtaining spectral energy distributions for GOODS catalogued objects, and for measuring uncatalogued objects dynamically using a web-service front-end to Sextractor (Fig. 3.) The US NVO project showed three prototypes at the January meeting of the American Astronomical Society in Seattle: 1) a transient event follow-up service, 2) a brown dwarf candidate search, and 3) a galaxy morphology analysis. These demonstrations utilized new standard interfaces and protocols for accessing catalog and image data, and the galaxy morphology demo employed grid-based computing for doing parallel computations. The brown dwarf search was intended to validate previous results in a fraction of the time taken for earlier candidate identifications, but yielded the exciting result of a new brown dwarf discovery (Fig. 4). The Astro-Grid project contributed to the AVO demo but also has science use cases of its own, including analysis of deep field surveys, high redshift quasars, and searches for low surface brightness

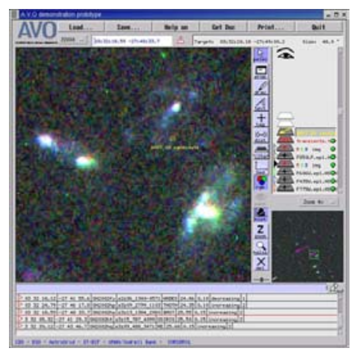

Figure 3. The AVO Demo interface to explore multiwavelength GOODS data http://www.euro-vo.org

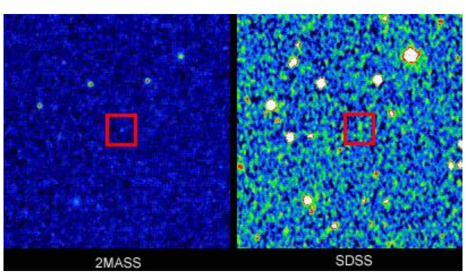

Figure 4. The NVO brown dwarf candidate search demonstration project yielded a new brown dwarf discovery.

galaxies. Recently the Canadian VO project released its first demonstration (available at the URL shown in Table 1).

Technology

In its January 2003 meeting the IVOA identified six major technical initiatives necessary to make progress toward the scientific goals of the VO.

Registries. Registries function as the "yellow pages" of the Virtual Observatory, collecting metadata about data resources and information services into a queryable database. But, like the VO resources and services themselves, the registry is also distributed. Replicas will exist around the network, both for redundancy and for more specialized collections. The VO projects are investigating a variety of industry standards for implementation of registries, including the Open Archive Initiative (OAI) developed in the digital library community. Registry metadata are using the Dublin Core definitions, also developed for the library community, wherever possible.

Data Models. Although the international astronomy community has long agreed on a common format for data, the FITS standard, there are many variations in which metadata can be encoded in FITS files, and many options for storing associated data objects (a spectrum, its wavelength scale, and its variances, for example). FITS is a syntactic standard, not a semantic standard. The VO data models initiative aims to define the common elements of astronomical data structures and to provide a framework for describing their relationships. Data models will allow software to be designed to operate on many data storage variants without needing to modify the source data structures.

Uniform Content Descriptors. The CDS in Strasbourg pioneered the development of UCDs in order to make semantic sense of its large collection of astronomical catalogs and tables. Among the tens of thousands of column names in its collection, they found that there were only about 1500 unique types of content. Astronomers are creative, having found some 250 labels for a Johnson V magnitude! UCDs will provide a lingua franca for metadata definitions throughout the VO.

Data Access Layer. Building on the VO data models and UCDs, the data access layer provides standardized access mechanisms to distributed data objects. Two initial prototypes for the DAL have been developed thus far: a ConeSearch protocol and a Simple Image Access Protocol. The former returns catalog entries for a specified location and search radius on the sky, and the latter returns pointers to sky images given similar selection criteria. Work is underway to extend the DAL to other data types and to enable legacy software systems to incorporate DAL interfaces.

VO Query Language. The many and distributed databases of the VO will need a standard query language. Although most, if not all, modern astronomical databases are queryable with SQL, SQL has limitations in areas fundamental to astronomical research, such as region specifications on the sky. The concept of a join based on spatial coordinates must be "fuzzy", allowing for uncertainties in the coordinates, differences in spatial resolution of detectors, and different physical scale sizes of objects at different wavelengths. Some groups are experimenting with very high-level query languages that would allow natural language query expressions.

VOTable. The first international agreement reached by the VO projects was VOTable, and XML mark-up standard for astronomical tables. The heritage of VOTable comes from FITS, the CDS Astrores format, and the industry-standard eXtensible Mark-up Language. The data access layer ConeSearch and SIAP services return results in VOTable. VOTable software libraries have been developed in perl, Java, and C++, and VO India has developed a general purpose VOPlot program in Java for data display. VOTable has been in use for just over a year, and the IVOA is now looking into what enhancements or extensions might be necessary.

IAU Joint Discussion and IVOA Displays

The IAU General Assembly in Sydney, Australia (July 2003) will be a focal point for IVOA activities. Joint Discussion 8 will include four sessions devoted to the Virtual Observatory, and a special session on Future Large Facilities will feature talks by the co-authors of this article on the impact of the VO on new missions and telescopes. The IVOA will also have a large display featuring demonstrations from participating organizations. We invite all participants at the IAU to stop by and see the VO prototypes in action.

Participation

The IVOA welcomes new VO projects to its membership. Please contact either of the authors of this article for more information via http://www.ivoa.net.