Library and Information Services in Astronomy IV
July 2-5, 2002, Prague, Czech Republic
B. Corbin, E. Bryson, and M. Wolf (eds)

Virtual Observatories: the Future of Astronomical Information

Peter J. Quinn

European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany
pquinn@eso.org

Abstract. There are currently several funded international efforts on the design and initial deployment of virtual observatories. The Astrophysical Virtual Observatory (AVO) project is an EC supported, three year Phase-A study of a VO for European astronomy. The UK eScience ASTROGRID project is a member of the AVO consortium. Both European projects are affiliated with the US National Virtual Observatory and other international VO projects. I will outline the goals and objectives of these projects and their current status as well as key European and global milestones for the next three years.

1. The Need for a New Research Infrastructure in Astronomy

1.1. The Winds of Change

> *"Internet computing and Grid technologies promise to change the way we tackle complex problems. They will enable large-scale aggregation and sharing of computational, data and other resources across institutional boundaries. And harnessing these new technologies effectively will transform scientific disciplines ranging from high-energy physics to the life sciences."*
> (Dr. Ian Foster, Co-leader of the GLOBUS Project)

Particle physics is a huge and booming science these days. Gigantic accelerators and simulators provide more and more bytes of information. The data available in this discipline comprise 8-10 PetaBytes per year on tape – a ten with twelve zeros: 10,000,000,000,000 bytes. As a consequence, equally gigantic data processing and storage capacity is needed; storing only one PetaByte per year on disk requires the computing power of approx. 300,000 personal computers.

What has this to do with astronomy?

Astronomy has become a BIG international science. The number of large telescope facilities has grown dramatically during the past few years, and there is no end in sight. Gemini on Hawaii, the multinational ALMA telescope, ESO's Very Large Telescope, the Square Kilometre Array (SKA) project, the New Generation Space Telescope and other extremely large telescopes are still on their way towards collecting as much data as possible. The peak has not yet been reached.

1.2. Strategies for Astronomy

To fully exploit the data explosion from these new observatories, astronomy projects require international coordinated research efforts. Projects are no longer limited to one or two wavelengths or data coming from one institute's archive, but involve distributed multi-wavelength teams, resources and data.

The speed of data increase is breathtaking. Astronomy service organizations need to provide their communities with access to software tools as well as high quality raw and processed data in the face of data volumes doubling in less than 12 months. For instance, the size of the ESO/ST-ECF Science Archive comprised approx. 10GB in 1992. In 2001, it had grown to remarkable 10,000 GB.

Slow CPU growth, limited storage capacities, limited bandwidth and data diversity are some of the most urgent problems in astronomy. Which strategies can be applied to solve them? The solutions lie in distributed computing and data storage, in information hierarchies that permit to move only those data which are indeed needed, and in interoperability. Using the GRID paradigm of distributed computing and resources, researchers will be able to conduct new and innovative programs.

Similar strategies are also applied in other fields that handle large datasets. For instance, CNN maintains data centers and content distributors around the world. A request for information launched on a web browser somewhere in Europe will collect the necessary information from the relevant data centers and combine them into one appropriate web page, tailored to the needs of the requester. Likewise, the GRID technology will enable researchers to access data specific to their information request.

2. New Science Through New Technology

2.1. Virtual Facilities

In order to monitor rare and detect unexpected events in astronomy, it is essential to join large data sets and mine them in an organized, structured way, providing researchers with an unbiased census of the sky at all wavelengths. Finding new examples of known objects or finding examples of new object types may be possible by probing the time domain: questions like "what has moved?", "what is new?" or "build me a light curve for objects of type x" can be answered if individual data entities are interconnected, making use of widely agreed-upon standards that guarantee interoperability.

Similar to the customized CNN page that pulls information from various data centers around the world and combines them into a personal "myCNN" page upon request, virtual observatories give access to specific data originating from a variety of information centers and archives. VOs consist of collections of observatories, each with its own observing proposals, raw and processed data and calibrations, with instruments and analysis tools, and with links to the resulting publications.

Like a “real” observatory, also the Virtual Observatory comprises various levels:

- the Observatory provides the infrastructure (GRID, SW and connection) and is in charge of the overall coordination and aggregation including securing necessary project resources
- the Telescope develops specialized resources and tools for seamless data access
- the Instrument offers software/inquiry tools for specific science projects; it may talk to multiple telescopes

2.2. The AVO Proposal, Status and Work Program

Following the OPTICON (Optical Infrared Coordination Network, www.astro-opticon.org) recommendations for a European Virtual Observatory effort, a proposal was submitted to the European Commission under its Research and Technical Development (RTD) scheme in February 2001. A similar effort in the US (National Virtual Observatory, NVO) was proposed by the latest Decadal Report.

The Astrophysical Virtual Observatory (AVO, www.euro-vo.org) will conduct a research and demonstration program on the scientific requirements and technologies necessary to build a VO for European astronomy. The AVO consists of six partner organizations, led by the European Southern Observatory (ESO, Principal Investigator: P. Quinn). The other partners are the Space Telescope European Coordinating Facility (ST-ECF, P. Benvenuti), the Centre de Données Astronomiques de Strasbourg (CDS, F. Genova), the CNRS-supported TERAPIX astronomical data center at the Institut d’Astrophysique de Paris (Y. Mellier), the ASTROGRID (UK) consortium (A. Lawrence) and the University of Manchester’s Jodrell Bank Observatory (P. Diamond), as well as affiliates from the NVO.

The AVO proposal asked for funds for a three year Phase-A study, totaling to 7.2 Million Euros, with 50% being awarded from the European Union and 50% from organizations. The Phase-A work program will focus on three work areas: (1) science: a detailed description of science requirements for the AVO, including multi-wavelength science case demonstrations based on the preceding ASTROVIRTEL project; (2) interoperability: data and archive interoperability in the light of new standards definitions for astronomical data, mainly carried out by CDS; and (3) technology: assessment of the GRID and database technologies necessary for use within a full AVO implementation (ASTROGRID/ESO).

3. The International Virtual Observatory Alliance (IVOA)

The need for virtual observatories is being recognized around the world. Astronomical communities are working towards one truly global virtual observatory that will enable researchers to carry out new science based on the wealth of astronomical data held in the growing number of first-class astronomical archives. The International Virtual Observatory Alliance (IVOA) is not just another project, but an alliance of existing and future national and international projects to

(a) define the common ground needed to make an operational and scientifically effective IVO, (b) reach agreement on standards and interoperability, and (c) allow the international scientific communities resident in the national projects, and elsewhere, a global channel for comment, criticism and guidance.

3.1. Future Milestones

The agreed roadmap for the years 2002–2005 is as follows:

January 2002	Initiate international dialog on interoperability. OPTICON Interoperability Working Group meeting, Strasbourg, including discussion on and revision of the draft VOTable standard
April 2002	Reach agreement on VOTable 1.0. Early progress has been made regarding the VOTable, a proposed XML format for astronomical tables. Version 1.0 was released on April 15, 2002; relevant documents can be found at cdsweb.u-strasbg.fr/doc/VOTable/ For comments, contact VOTable@us-vo.org
June 2002	Formation of IVOA
January 2003	Coordinated initial science demonstrations by IVOA members
January 2003	IVOA agreement on initial suite of interoperability standards and tools
May 2003	Working published web services
August 2003	Coordinated intermediate science demonstrations including international access at IAU General Assembly
October 2003	Astronomical Query Language
January 2004	Coordinated intermediate demonstrations + Grid
May 2004	Resource Discovery 1.0
July 2004	IVO 2005+ roadmap
October 2004	Compound Web Services and Ontology Service 1.0
January 2005	Coordinated complex science demonstrations

See www.ivoa.net for more info.

AVO Phase-A will continue from 2002 through 2004, followed by AVO Phase-B in the years 2005+. Phase-B envisions an alliance of all European data centers and data producers including observatories, theory centers, and libraries. An FP6 Expression of Interest for this project phase was submitted in June 2002.

Further information on the Astrophysical Virtual Observatory can be found on the AVO web pages at www.euro-vo.org

Acknowledgments. Special thanks to Uta Grothkopf for transcribing the presentation.