Towards the National Virtual Observatory

A Report Prepared by the NVO Science Definition Team, April 2002

Executive Summary

Astronomy has become an enormously data-rich science. The cumulative data volume, now measured in hundreds of Terabytes, is growing exponentially, with increases in data complexity and quality as well. The universe is now being explored systematically, in a panchromatic way, over a range of spatial and temporal scales that lead to a more complete and less biased understanding of its constituents, their evolution, their origins, and the physical processes governing them. Complex multi-dimensional astrophysical phenomena require complex multi-dimensional data for their understanding. The inherent limitations of individual data sets can be overcome by combining them and thus uncovering a new knowledge that cannot be gained from any one of them individually. This great richness of information poses substantial technical challenges, ranging from data access and manipulation to sophisticated data mining and statistical analysis needed for their scientific exploration. Our current ability to fully exploit scientifically this data avalanche is limited by the existing tools and resources, and the problem is growing rapidly.

The Virtual Observatory concept is the astronomy community's answer to these challenges. It represents an organized, coherent approach to the transition to a new, information-rich astronomy for the 21st century. The National Virtual Observatory (NVO) and its counterparts worldwide would represent *a new research environment for astronomy with massive data sets,* harnessing the power of the information technology and the expertise from applied computer science, statistics, and other fields to advance the progress of astronomy. This novel scientific organization will be a geographically distributed, web-based, open and inclusive environment, with a broad constituency of users and contributors that transcends the traditional divisions of wavelengths, ground vs. space based, etc.

The NVO will federate the currently disjoint set of digital sky surveys, observatory and mission archives, astronomy data and literature services, and it will greatly facilitate the inclusion of future ones. It will provide powerful tools for the effective and rapid scientific exploration of massive data sets. It will be *technology enabled, but science driven*. It will generate more efficient and cost-effective science, and even completely new science, by making practical those studies that today are too costly in terms of the efforts and resources required. It will empower scientists and students everywhere to do first-rate astronomy, and it will become an unprecedented venue for science and technology education and for public outreach.

The scientific and operational requirements of the NVO map into a set of technical capabilities needed in order to fulfill this vision. These include development of the following:

- Standards for accessing large astronomical data sets that can accommodate the full range of wavelengths, observational techniques, and application-specific needs for all types of astronomical data: catalogs, images, spectra, visibilities, and time series; including also standards for metadata, data formats, query language, data and service discovery, etc.

- Mechanisms for the federation of massive, distributed data sets, regardless of the wavelengths, resolution, and types of data (e.g., images, spectra, catalogs, annotations, numerical simulations, etc.).
- Provide new, effective mechanisms for publishing massive data sets and data products, including quality control and links to other published data and to the literature.
- Data analysis toolkits and services, including source extraction, parameter measurements, and classification from image data bases; data mining for image, spectra, and catalog domains, together with federations thereof; multivariate statistical tools; and multi-dimensional visualization techniques for novel and complex data sets.
- System wide gateways to provide access to these capabilities.
- EPO-oriented data access and capabilities.

These capabilities would build on existing foundations and would greatly expand their functionality. Their implementation would lead to a fully operating NVO. We envision and recommend three phases of development:

1. Conceptual design, expanded definition of science drivers, implied technical capabilities, general, management, and costing issues; early development work, including further development of prototype NVO services that are funded through the existing grants and programs. CY 2002 - 2003.
2. Definition of the NVO operational/management structure; detailed implementation plan; increased capabilities implemented within the existing data centers, surveys, and observatories; increased community input and involvement; initial development of archives for major ground-based observatories; dedicated NVO science funding. CY 2002 - 2005.
3. Implementation of the full-fledged NVO structure with international connections; commencement of major NVO-based science programs; start of routine operations. From CY 2006 onwards.

A preliminary analysis suggests that the total anticipated budget for the period of 10 years will be $ 90 M (in real year dollars) of new funds, about 30% of which will be in grants and fellowships to the community. We believe that an investment in the NVO will result in significant cost savings for future surveys, missions, and other data intensive projects. We recommend the following immediate action items:

A. Definition of a suitable organizational/management structure for the NVO that can accommodate the needs and constraints of both the NSF and NASA and be consistent with the vision, scope, and capabilities needed.

B. Pending the existence of the NVO as an organizational entity, creation of an advisory, planning, and coordinating body which would continue and expand the functions started by this Team.

C. Identification of the funding venues for the development and prototyping of NVO technical, scientific, and EPO demonstration projects, together with steps towards a more dedicated and extensive future funding.

In conclusion, it is the unanimous view of this Science Definition Team that the National Virtual Observatory is essential to the future well-being and competitiveness of U.S. Astronomy, and that this initiative should be fully funded as expeditiously as possible.