The National Virtual Observatory

An Information Technology Framework for Astronomy

Funded by the National Science Foundation

Project Development Roadmap

The National Virtual Observatory (NVO) is creating an environment for astronomical research that will enable the execution of research projects whose scale and scope have not been possible previously. This capability will lead to an era of unprecedented astronomical discoveries. The NVO will establish this environment through the use of high-performance computing, large-scale databases, and web and grid services. The NVO will establish standards for data representation and services, and it will integrate resources (catalogs, image archives, and processing pipelines) with standard services (image, spectrum, and catalog access protocols) to provide an environment of unprecedented power and simplicity for carrying out scientific research.

A series of science prototypes will guide the development of infrastructure and help define the core toolkit. We will work closely with the communities of end-users and developers. This interaction will help to define the science and technical demonstrations, and will engage a new community of VO developers--just as the development of the WWW ignited a new wave of creativity in information services. These services, such as a sophisticated global resource registry for the publication, discovery, and use of NVO-compliant resources, will be the basis for ground-breaking astronomical research. Advanced information technologies such as virtual data, distributed computing, grid services, and large-scale database access will facilitate the construction of complex data analysis and visualization tools.

The NVO is an excellent vehicle for education and public outreach, both in astronomy and in the underlying IT. We will increase the amount of data in NVO-compliant repositories that is useful for outreach activities, and develop the software tools and services that can readily identify and access suitable NVO datasets. We will help create portals to NVO datasets to that are useful to outreach professionals, educators, and/or the general public. We will show how modern IT infrastructure enables new research and vastly enriched information dissemination.

Year 1

Science

Objectives

Select initial science prototypes for implementation based on science interest and feasibility of implementation. Each prototype will have a design document, implementation schedule, key technical requirements, and project team. Demonstrate at the the January 2002 AAS meeting.

Accomplishments

Created and demonstrated Gamma-ray burst follow-up service (a “show me the sky” service), a brown dwarf candidate search (an SDSS/ 2MASS cross-correlation service, which discovered a new brown dwarf!), and a galaxy morhphology analysis (which used the Grid tools Chimera, Pegasus, and Condor to perform the analysis dynamically).

Technology

Objectives

Reach consensus on V1.0 of VOTable (XML) standard for exchange of tabular data, and develop associated I/O software libraries.

Define basic catalog and image access protocols, and utilize in science demonstrations.

Develop core metadata standards for describing astronomical datasets, and for collections of astronomical data. Plan and prototype resource directories (registries) that collect the metadata associated with data collections.

Peform initial experiments with Grid technologies to gain experience with capabilities. Evaluate and test distributed, Grid-based data storage mechanisms.

Accomplishments

VOTable 1.0 released in April 2002. At least four I/O libraries developed in VO collaborations.

“Cone search” (catalog access) and Simple Image Access Protocols developed and utilized in Gamma-ray burst and galaxy morphology prototypes.

“Unified Content Descriptors” were developed for semantic description of catalog content.

Definitions were drafted for Space-Time metadata standards. Wrote initial draft of resource metadata schema for international resource registry for astronomy.

Grid software (Chimera, Pegasus, Condor) was used to support the galaxy morphology algorithm on a Grid-enabled cluster. Gained experience with the Storage Resource Broker.

Education and Public Outreach

Objectives

Gather community inputs for EPO-based requirements for the NVO. Initial focus should be on metadata needed to properly identify and classify NVO resources appropriate for EPO use cases.

Accomplishments

An EPO workshop was held resulting in contacts with ~15 EPO organizations. Representative projects were identified, including the establishment of an amateur astronomy image archive with Sky & Telescope magazine. A requirements document was developed describing EPO-specific metadata.

Year 2

Science

Objectives

Incorporate requirements from the theoretical astrophysics community, developing a science demonstration based on theoretical simulations of globular clusters for the January 2004 AAS.

Extend Year 1 science demonstrations to usable services.

Incorporate international data access into science demonstrations.

Collect and integrate science requirements for access to spectral data. Expand data access capabilities with additional Cone Search and SIAP services.

Accomplishments

Created development plan for science demonstration based on GRAPE globular cluster simulations.

NVO Data Inventory Service released as first NVO public service.

Continued to collaborate with IVOA to assure interoperability of US and non-US facilities. Science demontrations for July 2003 IAU incorporated international data access.

Usage scenarios for spectral and simulation data requested and received. NED database has a NVO-compliant (SIAP) service.

Technology

Objectives

Refine VOTable standard, as agreed with IVOA partners.

Re-examine the mapping of UCD structures onto data models to understand how to provide access to data collections.

Work toward international consensus on registry of astronomical resources, and on creating persistent digital identifiers.

Develop a VO Query Language standard to access the registry.

Determine strategy for integrating web services with grid technology.

Implement an NVO testbed on the NSF Teragrid, including the replication of additional image archives onto Teragrid resources (SRB). Gain experience in implementation of Web Services, including a Registry Service.

Develop initial data models for spectral and time series data, and extend SIAP interface definitions accordingly.

Accomplishments

VOTable extensions will be considered by the IVO collaboration in Autumn 2003.

Full ontology extension of the UCD vocabulary being discussed.

Metadata attributes (VOResource) defined to support a distributed NVO registry, with two publication portals. Achieved international agreement on syntax for digital object identifiers.

VO Query Language development broken into three streams (SQL extensions, OpenSkyNode, and natural language). Role in registry queries remains under discussion.

Base current implementations on web services, and map to corresponding Grid services as they become stable.

Grid codes in development for mosaicing and multi-wavelength image federation (Atlasmaker/Montage).

Spectral data access will be through separate interface, SSAP, at least initially. Use-case scenarios being used to define protocol. Work on time series data deferred.

Education and Public Outreach

Objectives

Incorporate EPO metadata in NVO resource registry.

Update survey of important EPO resources.

Update survey of EPO-oriented access tools to NVO resources, and document for EPO users.

EPO metadata incorporated in NVO resource registry. An analysis of an interactive kiosk design for museum and planetarium partners, based on the Data Information Service. Interactions with IVOA partners will be pursued for development of common services.

Survey EPO Community projects that could most easily utilize and benefit from NVO data access.

Accomplishments

EPO-related metadata integrated in resource metadata for registry.

Year 3

By year 3, the expectation is that the Teragrid will become available for demonstrations at scale of representative NVO services. This requires both access to NVO collections and integration with grid services. The NVO service registry will be populated with services from each of the data centers.

Science

Objectives

Make direct comparisons of observed and simulated data, with focus on Globular clusters.

Execute multiple large science runs that analyze of the contents of entire image archives. The Atlasmaker/ Montage package will create atlases of SDSS, 2MASS, DPOSS, FIRST, etc, and provide NVO-compliant access services.

Create wide-area atlases (digital reference sets) from multiple, large sky surveys, allowing data-mining of multi-wavelength imagery.

Port multi-parameter analysis packages to ingest NVO-compliant data services (e.g., density estimation, N-point correlation). These packages will run as services on the Grid. Plan a scientifically significant cluster/outlier search. Integrate clustering/outlier software with other packages.

Deploy core science services for use by the research community.

Accomplishments

Technology

Objectives

Define and develop OpenSkyNode services to provide open database access among international VO partners. Evaluate OGSA-based implementations.

Implement generalized cross-correlation services for distributed catalogs, with web and grid service support.

Complete spectral data access protocol and deploy SSAP services.

Work closely with major data providers to facilitate development of VO-compliant data access services.

Use registry services routinely for publication, discovery, and utilization of VO resources.

Prototype a knowledge engineering (ontology) approach to astronomical knowledge by extending the UCD vocabulary.

Expand NVO test-bed and test scaleability of algorithms and data access methods. Incorporate support for virtual data products.

Accomplishments

Education and Public Outreach

Objectives

Promote interactions with the EPO community through the deployment of interactive kiosk, and the completion of an NVO portal for outreach.

Encourage and aid resource managers in converting data with EPO potential into EPO-ready data.

Produce a document that provides a broad overview of NVO resources for non-technical audience.

Produce a list and short description of datasets already in NVO resources that are suitable for general EPO usage without further processing.

Accomplishments

Year 4

The NVO will plan for longer term deployment and support. Large amounts of metadata will be searchable through the NVO registry, and large amounts of data will be available through NVO-compliant standard services. Emphasis will be on supporting publication of data at all stages of creation from personal libraries to journals. Grid computing and big databases will be accessible through the Virtual Data paradigm.

We will build the 10% of services that will make 90% of customers happy!

Science

Objectives	***Accomplishments***
Search for outliers and clusters in large federated datasets via datamining algorithms.	
Find faint, variable, and very extended objects in large federated datasets via datamining in the image domain.	
Compare large (TB-scale) theoretical simulations with observational data.	

Technology

Objectives	***Accomplishments***
Increase deployment of Grid services for operating on large, federated data collections.	
Provide large-scale derivation (virtual) data products, such as statistically qualified cross-matches between large surveys.	
Expand registry functional to a semantic web or concept space.	
Interface NVO web-based services to Grid-based counterparts.	
Provide NVO development toolkit and associated training to community of researchers. VO-enabled applications developed in the community should begin to exceed those developed within the core project.	

Education and Public Outreach

Objectives	***Accomplishments***
Provide, with NVO partners, EPO-oriented portals to NVO data and services.	
Use NVO services as a vehicle for outreach about IT itself.	
Repeat and update survey of EPO community requirements and formulate plan for inclusion in NVO infrastructure.	

Year 5

The NVO will transition to longer term deployment and support, including workflow systems for digital library processes for creating standard digital reference data sets and publication of scientific data. The NVO will promote full integration with Journal publishing, and wide-scale use of the NVO registry. The NVO registry will become an essential component of cyber-infrastructure for astronomy. Wide use of grid computing and big databases will be accessible through the Virtual Data paradigm.

Science

Objectives	***Accomplishments***
NVO-enabled science should be visible in the peer-reviewed literature.	
NVO-based research tools will be in routine use, and will become and essential part of the environment for doing astronomical research.	

Technology

Objectives	***Accomplishments***
Scale registry services to larger numbers of resources and perhaps finer level of detail, with support for many simultaneous users.	
Maintain core systems and improve capabilities in step with continuing evolution of underlying IT, digital library, and grid technology.	

Education and Public Outreach

Objectives	***Accomplishments***
Utilize NVO-compliant resources in many levels of education and outreach.	
Seek interaction with NSF NSDL initiative for development of curricula based upon NVO resources.	