Quarterly Report
January – March 2008

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

IVOA

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement AST0122449

Annual Report

Period covered by this report: 1 January – 31 March 2008

Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

The NVO Book—The National Virtual Observatory: Tools and Techniques for Astronomical Research"—was published as Volume 382 in the Astronomical Society of the Pacific Conference Series. The book is based on the materials developed for the three NVO Summer Schools. M. Graham (Caltech), M. Fitzpatrick (NOAO), and T. McGlynn (NASA GSFC/HEASARC) edited the book.

The NVO project had an exhibit at the January AAS Meeting in Austin, Texas. Pre-release components of the data discovery portal were demonstrated, and the theme "NVO Inside" was used to highlight how many organizations are making data and services available through VO protocols.

The first edition of the NVO Newsletter was distributed in March to a mailing list of nearly 400 people who have expressed interest in the VO. The Newsletter is also available on the NVO web site.

We are now accepting applications for the fourth NVO Summer School, which will be held in Santa Fe, New Mexico in early September. The Summer School faculty has been planning the program and assessing the need for new or updated software and tutorials.

Much progress was made on the deployment of the new registry schema. All IVOA registries worldwide were scheduled to be updated in late April. Also, initial implementations of VOSpace were under development, and interoperability testing between VOSpaces at Caltech, JHU, and in AstroGrid are planned for the coming quarter.

As the NVO development project comes to a close, we are undertaking a comprehensive review and assessment of all software (applications, tools, libraries) that have been developed in the past 6½ years. The goal is to understand the level of completeness and the long-term value so that we can identify those components that are most essential to support in the future.

Activities by WBS

1 Management

1.1 General (planning, reporting, communications, team meetings, etc.)

Regular weekly telecons of the Technical Working Group (TWG) continue. Similarly, the Executive Committee also continues to meet weekly by telecon.

The spring NVO team meeting was hosted by Caltech and held in the Millikan Library, March 19-20, with 37 people attending. J. d'Ignazio (Syracuse University), a graduate student in Information Sciences, attended as an observer. He is studying the dynamics and processes within distributed technology development projects. The summer team meeting will be held in Estes Park, Colorado on June 30/July 1.

1.2 Science

The NVO Executive Committee, and NVO Project Scientist D. De Young (NOAO), have been closely following the work on the data discovery portal (led by T. McGlynn, HEASARC). Progress on the portal was reviewed at the spring team meeting.

1.3 Technical (including standards, configuration management)

Members of the Executive have been following progress on several key, new standards: the Table Access Protocol (TAP) and the Astronomical Data Query Language (ADQL). Both will be discussed in May at the IVOA Interop workshop in Trieste.

1.4 Financial

Total project expenditures since inception now stand at $12,018,407 compared to a budget of $14,151,966, leaving a balance of $2,133,559. Quarterly expenditures have been running around $400k for the past year. We are preparing a detailed analysis and projection and will consult with NSF as soon as this analysis is complete.

1.5 International coordination/collaboration

Senior members of the NVO project participated in IVOA Executive telecons and discussions. D. De Young (NOAO) continues as chair of the IVOA Executive Committee, and R. Williams (Caltech) chairs the IVOA Technical Coordination Group. De Young attended a meeting of the EuroVO Science Working Group, and M. Graham (Caltech) served on the faculty for the Australia VO Summer School.

2 Science Requirements

2.1 Usage scenarios for all areas of astronomy research, including theoretical simulations

Previously defined usage scenarios were used to evaluate progress on the data discovery portal (see WBS 9.1). In particular, a use case based on comparing radio and x-ray properties of rich clusters of galaxies from the Abell catalog is proving to be a very valuable benchmark. The user discovers the location of the Abell catalog, selects a subset of the richest clusters, checks the inventory service to see which have radio and x-ray data, queries the relevant catalogs, and integrates the catalog information and imaging data together. This scenario exercises all standard interfaces and assures that the components of the data discovery portal can work together.

2.2 Requirements analysis
No activities to report this quarter.

2.3 Demonstration definition and review
Pre-release versions of components of the data discovery portal were demonstrated at the January AAS Meeting in Austin.

3 Operations, System Integration, and Testing

3.1 Quality assurance and software engineering design
A review of all software system developed at the NVO was initiated at the team meeting in March 2008. The software development process for approximately 50 software elements developed within the NVO was reviewed. The design, testing and configuration management processes for these systems were briefly discussed. Further more detailed review of software is planned to ensure that as the NVO effort is wound down, the follow-on project has a clear sense of the state and utility of the NVO software holdings.

While the NVO team will continue to use the CVS repository, discussion at the NVO team meeting recommended that new software tools more directly supporting distributed development may be appropriate for the successor project. T. McGlynn (HEASARC) circulated at document describing the suggested structure for projects in the SVN repository. This was discussed and generally approved.

3.2 Facility operations
Small modifications were made to the standard NVO web application templates, which were adopted at many of the distributed NVO sites. Some sites remain to be updated. The main NVO site at Caltech (us-vo.org) continues to operate with no significant disruptions other than occasional maintenance downtimes. The NVO TWiki site at http://chart.stsci.edu/twiki/bin/view/Main/WebHome continues to be maintained by STScI as a primary venue for communications of events among the NVO team. All major NVO facilities continued standard operations during the quarter with no major interruptions of services were reported though a number of facilities experienced scheduled and unscheduled downtimes for periods of several hours. New operational Web pages included sites for the 2008 Summer School and sites associated with the NVO Portal.

The SVN configuration management system remains fully operational at Caltech. This system is now being actively used by some of the development projects as their active repository. Other efforts continue to use local CMS systems and periodically copy updates to the central repository. The ticketing and tracking system (Trac) is now being extensively used in the Portal development and to a lesser extent in other NVO development projects. The Zoho management system developed by R. Hanisch (STScI) continues to be used to describe the progress of NVO tasks.

M. Preciado (HEASARC) has continued to operate and augment a suite of tests using the NAGIOS software system which checks every site providing VO services. Typically tests are run every hour. For sites which provide many services, only a representative subset of the services are checked to avoid undue use of the services. Currently over 100 services are tested. Preciado reviews the status of all sites periodically (~daily) and sends notifications to sites where requests are failing. Critical sites are informed immediately while less critical and non-NVO sites are normally queried if they have been down for

several hours. This continual monitoring of the health of the VO along with follow-up has helped to significantly decrease the number of sites that failed testing during the past quarter. A record of all site status queries and the responses received is maintained and available for review. The procedure for informing sites of problems is itself being reviewed and standardized. Approximately 2-3 issues are brought up each week.

3.3 User support

All user queries were directed to appropriate NVO team members and rapidly addressed.

M. Preciado (HEASARC) continues to develop Web pages designed to make it easier for end-user developers in the NVO. Many more examples of VOTable and FITS pages have been developed and an index page with clear links to the various test files is included. Links to external test files are also included. Facilities for easily downloading all or some subset of the test files have been provided.

S. Emory Bunn (Caltech) wrote and distributed the first NVO Newsletter. In preparation a mailing list was prepared using signup sheets at the January AAS meeting and convolving that with other mailings for persons involved with the NVO in the past several years. The newsletter is available on the Web at http://www.us-vo.org/newsletter/1/. A procedure for editing the newsletter was established where D. De Young (NOAO) serves as editor and each issue will be reviewed by the NVO executive committee prior to release.

4 Registries

In this quarter, we have completed migrating registries—both our own in the NVO project and those of our international partners—to the latest registry standards. Support to our partners has been mainly through the Registry of Registries, a service for both discovering and testing registries in the VO run by NVO on behalf of the IVOA. As the current chair of the IVOA Registry Working Group, R. Plante (NCSA) has been coordinating the resolution of remaining interoperability issues during and after the upgrade.

4.1 Resource metadata

A separately funded project, Virtual Astronomy Multimedia Project (VAMP) has been developing a standard for metadata tagging for graphics images created for outreach purposes called Astronomical Visualization Metadata (AVM; http://www.virtualastronomy.org/avm_metadata.php). In this quarter, Plante met with the authors to discuss if and how the IVOA standardization process could helpful to the promotion and uptake of the AVM standard. One interesting area where AVM overlaps with the IVOA standards development is in the area of controlled vocabularies for labeling the subjects of images. The IVOA Semantics working group has been promoting a standard for how to define a vocabulary and has been using the AVM subject terms as an example. Over the next six months the VAMP team and the larger outreach community will be developing the first applications that make use of AVM. Meanwhile, Plante will be working with the authors to enter the AVM document into the IVOA standardization process and promote awareness within the IVOA community. This discussion so far has raised questions about the IVOA interacts with development efforts that have not traditionally been "inside" the IVOA community.

4.2 Resource metadata schema

With most of our attention focused on the registry upgrade this quarter, there has not been much development in the area of metadata schemas; however, now that the upgrade is behind us, we will turn more attention to the roadmap for the standardizing VOResource XML schema extensions, particularly as we prepare for the spring IVOA interoperability meeting. In the queue is the general extension for describing data collections and services and a second standard covering the current set of "simple" standard data access protocols. D. Tody (NRAO) and the DAL working group has put together a draft extension for describing Table Access Protocol (TAP) services which has been important for understanding how to describe catalog services in general. The VOEvent working group is also working on an extension for VOEvent related services.

4.3 Publishing and harvesting protocols

The Registry of Registries (RofR; http://rofr.ivoa.net/), which was developed and is supported by Plante at NCSA, played an important role in this quarter's registry upgrade effort. This registry provides a listing of other publishing registries that are ready for operation under the latest registry standards. To be included in this list, a registry must be tested against the RofR's harvesting service validater to ensure its compliance with the standards. In this quarter, eight new registries were added to the list of available publishing registries for a total of 13 worldwide. Subsequently, a variety of miscellaneous interoperability issues were caught and handled. One registry in particular, the CDS Vizier catalog registry with its 6,000 records, uncovered some technical problems scaling up to a large number of records to inspect.

Through the IVOA Registry Working Group, we set a date of April 21 for having the majority of registries ready for production harvesting. This would allow a few weeks of operation to uncover any remaining interoperability issues before the IVOA Interoperability meeting in May; if at that time any unresolved issues remain, we will discuss their resolution there. By the end of that week in April, all the targeted registries in the IVOA were ready, and production harvesting commenced. Within the NVO, T. Dower (STScI) led the upgrade effort for the NVO Registry at STScI, implementing the recommended harvesting method that starts by finding registries in the Registry of Registries. G. Greene (STScI) assisted with the migration of code from our previous registry implementation, and Plante assisted with migration of the locally published resource records. With the commencement of harvesting across the IVOA, the new NVO registry was fully populated and made available to support the NVO portal.

In this quarter, we also made significant progress on the Universal Publishing Service, which will replace our current services that allow data providers to publish descriptions of their resources into NVO registries. Plante developed an initial design document based on our earlier studies of the current publishing interfaces. Then in February, the NVO registry team met at Johns Hopkins to discuss and refine the design. Dower has begun the work implementing this design and an initial prototype nearly complete. As part of our development plan, we put up an initial version of the service for internal review. Afterwards, we will have a second phase of development to bring the service to production readiness.

As that the new publishing interface is completed this year, we are relying on the publishing interface at the NCSA to support publishers that previously registered their resources

at the STScI registry. Plante modified this interface to operate on the new format records. While no effort was put into improving the actual interface (as this effort is going into the newly designed version), this did provide an opportunity to fix problems with individual records (see section 4.6).

4.4 Search protocols

Registry searching is a central capability of the NVO Portal, so we continue to make incremental improvements in the registry's search capabilities. The work this quarter has been in mainly two areas: an improved search and browse interface and in the inter-service portal communication. In particular, portal services transparently share data that the user has collected. One such set of data is a list of resources of interest to the user. This list is usually generated from an initial search of the registry. This list is then shared with other portal services in VOTable format. Greene and Dower have been responsible for ensuring interoperability of the Registry with the other portal components.

4.5 Replication, synchronization, maintenance, revision control, and curation

As part of the registry upgrade effort, Plante was responsible for the migration of resource records into the latest VOResource format. This provided an opportunity to correct a number of errors (many of them required to pass the RofR's compliance tests) in our existing records. There were a number of improvements made that went beyond simple compliance however, including improving subject keyword choices, testing service URLs, and reclassifying incorrectly classified resources. We expect this experience to be useful input in the development of a plan for production-phase curation practices that we will draw up this summer.

As part of our emerging software process, we continue to migrate, document, and organize our source code into our new software repository.

5 Data Models

5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality)

J. McDowell (SAO) visited IPAC to meet with O. Pevunova, R. Ebert and J. Mazzarella to discuss the status of the VO Data Model (DM) for application to SEDs produced by NED. An SED in the VO will essentially be a concatenation of "segments," with each photometric measurement (or group) represented by a 1-point (or few-point) spectrum characterized with the VO Spectrum DM. McDowell reviewed decisions remaining to complete the SED DM specification, most of which pertain to details of how to organize and serialize the metadata for SED segments and their combinations to form an SED. The discussion favored a hybrid approach analogous to the "Greenbank Convention" in FITS files. There was also discussion of the role of UTYPEs (for structure and semantics) and UCDs (for description of the physics) in the DM. The bulk of the NED work in this task area is dependent on work at SAO to (a) complete the specification of the SED data model; (b) program a Java implementation of the SED DM based on the existing SAO spectrum library; and (c) with the IVOA, define an IVOA passband registry service. While waiting for the above, Pevunova can only perform initial steps of mapping individual photometric data points within NED's SEDs into the VO Spectrum DM.

McDowell has been working on maintaining and updating the existing Spectrum java code originally developed by K. McCusker.

5.2 Low-level (measurement, quantity, uncertainty, relationship)
No activity during this quarter.

5.3 Descriptors and ontologies (UCDs)
M. Smith (STScI/MAST) is leading the development of a new stellar spectral classification system. He is preparing an IVOA Note for discussion at the May IVOA Interop meeting in Trieste.

5.4 Space-Time and regions
A. Rots and McDowell reviewed a proposal from F. Ochsenbein for STC serialization in VOTABLE and are preparing a response.

5.5 Standard schema
No activity during this quarter.

6 Data Access Layer

6.1 Data access services (catalog, image, spectrum, time series, visibilities, ...)
DAL development this past quarter focused on development of the new Table Access Protocol (TAP), and a major update to the Simple Image Access Protocol (SIAV2), both of which will be second generation DAL interfaces, similar to the SSA interface completed in late 2007. Work has also begun to define a standard footprint service protocol, to provide precise information on the spatial footprint of data collections and surveys.

Catalog access. Despite a successful international table access protocol (TAP) tiger team meeting held at JHU late last year, at which a number of promising agreements were reached, the IVOA TAP effort may be split into two parallel efforts. One effort, led primarily by ESAC and AstroGrid, will focus mainly on ADQL-based queries executing in a Grid context. The second effort, led primarily by NVO and CDS, will attempt to define a more general TAP interface, including support for both simplified non-ADQL queries as well as Grid-enabled ADQL queries, integrated into a common interface and providing uniform access to both table data and metadata. The longer term goal for IVOA and NVO is to ultimately replace these two interim TAP interfaces with a single unified standard TAP service interface.

NVO planning for the TAP interface continues building upon the agreements reached in the JHU TAP tiger team meeting in November. At this point, K. Noddle (AstroGrid), D. Tody, J. Salgado (ESAC), F. Ochsenbein (CDS), J. Good, and T. McGlynn, as well as A. Szalay and R. Hanisch are involved in development of the generalized TAP interface.

The scope of the interface will include both ADQL (SQL)-based queries as well as simpler parameter-based queries. Simple synchronous GET-based queries will be supported as well as Grid-enabled queries for cases where asynchronous execution is required, using VOSpace to stage and transport large tables. Uniform access to both table data and metadata will be provided. The intention is that TAP will provide a basic table access interface as well as eventually replace the very successful cone search interface, while at the same time a fully featured TAP will provide advanced capabilities for large or complex queries as well as multi-position and distributed queries. While TAP will not directly implement large distributed queries, TAP services will provide the primary interface to individual sites needed to implement such queries.

A draft TAP interface design was circulated to the design team in late April for discussion. An updated version of this will be discussed further at the IVOA Interop in May. Prototyping of a functional interim TAP interface is planned for this summer.

Spatial Footprints. NVO has recently initiated an effort to coordinate the draft of a standard Footprint Service specification. We are now entering a new phase of the virtual observatory where spatial features will be emerging in all of our services ranging from the registry to the query language and various complex applications. This effort is an attempt to work within the IVOA DAL working group to establish a consistent set of specifications for the various approaches currently implemented. The NVO organizations involved in this effort include JHU, CfA, and STScI. The NVO Trac repository contains a set of documents for further details (see http://trac.us-vo.org/nvo/wiki/Footprints).

The consistency NVO is working toward is at the keyword level mapping from ADQL to function calls in representative services and applications as well as in the way we specify spatial description. While there is an agreed upon IVOA standard spatial data model, STC, there is no such standard for the services that manage regions. There is an emerging ADQL specification that will accommodate much of the query handling for regions in which keywords are based on OpenGIS, an internationally agreed upon standard.

An initial telecon to discuss the footprint service was held in late April (G. Greene, A. Szalay, T. Budavari, A. Rots, F. Bonnarel, D. Tody). Further discussions will take place in a footprint session at the May IVOA interoperability meeting in Trieste.

Spectral and time series data. A capability schema for SSA services was completed (R. Plante, D. Tody, M. Dolensky) so that both SSA V1.0 services and older pre-SSA spectral services can be registered with the new registry, which was deployed in late April. Further extensions to this schema are planned to support SSA V1.1, which will add support for a getCapabilities service operation to allow client applications to directly retrieve the service capabilities from a service. Further development and testing of SSA services and client applications will resume later this year once the existing services are discoverable via the new registry.

Image data. A team within the IVOA DAL working group was formed last fall to begin work on the SIAV2 specification, a major upgrade to the image access protocol to bring it up to current second general DAL standards, and add new capabilities for cube access and Grid-enabled operations. Members of this team met several times this quarter to work on the SIAV2 design (D. Tody, F. Bonnarel, D. Durand, M. Dolensky). A session on SIAV2 is also scheduled for the IVOA Interop meeting in Trieste in May.

At this point the scope and concept for data access, e.g., to cube data, have been established. The current effort is focused aspects of the service interface. While the overall interface for SIAV2 will be largely based upon SSA, with much of the query interface and query response metadata in common, image access requires some new interface elements to deal with image-specific issues such as access to cube data. In particular, cube access will require specification of an image world coordinate system with at least spatial and spectral axes, and probably polarization and temporal axes as well. Neither FITS WCS nor STC fully addresses this problem, and some degree of compatibility and interoperability with both is required.

Complex data. Interest continues to be strong in adding a capability to the DAL interfaces for generic data discovery and complex data associations (generic data discovery refers to a data service which treats all kinds of data uniformly, e.g., catalogs, images, spectra, and so forth). Planning for how to deal with complex data associations is in the concept phase currently, and is being addressed in connection with the SIAV2 design. We don't expect to begin work on specifying and prototyping of the generic dataset service until SIAV2, SSA, and TAP are all available as second generation services which can be linked to generic dataset discovery, and referenced in descriptions of complex data associations in a generic dataset query response.

6.2 Data representation (VOTable, etc.)

A longstanding issue within VO has concerned whether data models should be specified as abstractions that can be represented in various formats, languages, and technologies, versus specification via a more specific physical representation such as XML. Separation of model from representation increases flexibility and enhances extensibility, but can result in a less controlled software integration. The use of generic abstractions such as the table and image can promote use of astronomy standard libraries and tools, and is consistent with common practice within astronomy today, whereas the use of non-astronomy specific technology such as XML can result in better integration with generic non-astronomical technologies such as for automatic data binding, XSLT transformation, schema verification, and so forth.

How to resolve these two points of view is still very much under discussion within the VO. Most recently it has come up in connection with the space-time coordinates standard (STC). While STC provides advanced capabilities for describing and representing astronomical coordinates and regions, it relies heavily upon a specific technology (XML) and can be hard to use within astronomical applications which take other forms or which require only limited coordinate representations, such as legacy or end-user oriented astronomical data analysis software. On the other hand software such as STC can work well in advanced modern software that does complex region manipulation or uses technology such as automated data binding to generate language bindings for schema defined data structures.

How to resolve this issue is not yet clear. VO needs to serve both the needs of end users who require use of a diverse variety of software for personal data analysis, as well as large projects and data centers who require advanced capabilities for automated data processing, where the end user may never see the software but only interact with a Web portal or application.

6.3 Framework (mediators, components)

A face-to-face meeting to discuss the applications framework phase II planning and interface design took place in Garching in late January (P. Grosbol, D. Tody, et al.). Phase II of this project is expected to go forward in 2009 (subject to the anticipated formation of the VAO), and will be a combined effort of the US and Europe, with OPTICON responsible for the European side of the effort. The purpose of the project will be to develop standards and reference implementations for the major elements of an astronomical data processing and analysis system framework. Related use-cases are also planned, which would use this technology in facility software, e.g., for client side data analysis with VO

(using VOClient as well as the applications framework), and for computation of virtual data products for VO data access services.

The initial phase of the applications framework project ends in 2008. To conclude the phase I effort, work is underway to prepare a high level specification of the major system interfaces, as well as update the system architecture specification. As part of this effort discussions of the interface design are planned with the major AUI and AURA observatories associated with NVO in the US, in addition to ongoing discussions with OPTICON in Europe. In particular, representatives of the applications framework effort within NVO (D. Tody, M. Fitzpatrick) will participate in an AURA-sponsored software planning workshop to be held in Hilo in June. Less formal discussions are also ongoing to coordinate with data processing software development within AUI.

A first working draft of the SAMP interface for applications messaging was completed in late April (M. Fitzpatrick, M. Taylor, T. Boch). This first version of SAMP addresses primarily high level inter-tool messaging, providing functionality similar to the PLASTIC prototype developed earlier by Euro-VO. A more general applications messaging facility is planned to be developed as part of the applications framework effort.

6.4 Data provider/consumer implementations and end-to-end testing

Work has begun on a load testing application and web portal, to be used to simulate multi-user loading of VO services (J. Crossley, D. Tody). This will eventually be part of the operations facilities planned for service verification and testing.

On 20 March, R. Ebert, O. Pevunova and J. Mazzarella (NED) discussed with D. Tody (NRAO) a plan to implement a Simple Spectrum Access (SSA) interface into the NED spectral database using the DALServer Toolkit. To this end, Pevunova and Ebert are studying the Simple Spectral Access protocol and the VO Spectrum Data Model. Pevunova is performing initial steps in mapping from the internal NED data model to the VO Spectrum data model by adding observation time information to the NED spectra metadata. R. Ebert downloaded the DALServer toolkit and set up a computer running Apache and Tomcat as a development/test platform for a NED SSA prototype service. Since the SED DM and DAL Java library is not ready to begin an implementation, during May - July 2008 we will be focusing primarily on prototyping the SSA interface into NED spectra, postponing implementation of SED data access until late summer.

7 Query Language

7.1 Low-level: Astronomical Data Query Language

The VOQ Technical Experts Group finalized its work on the ADQL standard. The draft has been circulated within the VOQL Working Group and will be become a proposed recommendation soon. Issues have been raised about how the STC standard is represented (A. Rots, R. Hanisch), however, there seems to be general agreement that the WD is ready to be discussed in the next IVOA meeting in Trieste.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

JHU has keeps working on implementing a cross-matching engine that addresses very large datasets. A new generation Open SkyQuery is being designed that uses as new infrastructure known a Graywulf. The Graywulf (In honor to Jim Gray) aims to combine SQL Server and High Performance computer clusters to provide seamless parallel data

access and computation. As part of the basic Graywulf infrastructure a cluster and workflow management systems are being implemented. Great progress has been done recently to define queries and data models. Both of these technologies will be incorporated into Open SkyQuery to enable it to handle arbitrary cross-matches of large datasets. Additionally, progress has been made related to VO space, thanks to B. Bauer, a graduate student from the Fakultät für Informatik of the Technische Universität Munich in collaboration with Budavari (JHU) has worked on a C# 2.0 implementation based on new Window Communication Foundation (WCF). The reference implementation is a self-contained SQL Server 2005 backend, follows a VOPipe Architecture that provides higher level services for data/work flows and establishes the basis for next generation data access VO services such as Open SkyQuery.

7.3 High-level: Complex queries
Nothing to report this quarter.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)
VOSpace is the IVOA interface to distributed storage. Work this quarter has focused on developing the VOSpace 1.1 specification that extends the existing 1.0 recommendation to support containers, links between individual VOSpace instances, third party APIs and a "find" mechanism. An IVOA Working Draft of VOSpace 1.1 with accompanying WSDL and schema was released and reference implementations have been developed at Caltech and JHU. The latter will form part of a data pipe system that will employed to remove the 5000 row limit in crossmatch operations. There has also been some effort in producing a version of VOSpace 1.1 based on the iRODS technology from SDSC.

Discussions have begun about VOSpace 2.0, a RESTful version of the interface, driven in part by CADC's intent to use VOSpace as part of the ALMA archive system infrastructure.

8.2 Grid Services (OGSA)
No activities to report this quarter.

8.3 Computational resource management
No activities to report this quarter.

8.4 Virtual data
No activities to report this quarter.

8.5 Application and service integration with Grid
In this quarter, we completed our plan for providing high-availability for single sign-on services with automated fail-over. We completed the deployment of the login server at NCSA (accessible via sso.us-vo.org) with a replica deployed at NOAO. Replication was eased by our packaging of the service as a virtual machine; thus, both sites run the same system under a VMWare server. As part of the dual deployment, the underlying user databases are continually replicated. In addition we set up a monitoring system (using the Nagios product) that lets us know when either system goes down. During this quarter, NOAO experienced an unusual series of power outages, which allowed us to demonstrate that our service remained available when one site is down.

With the deployment of these new services, some changes are required by participating portals. To ease the transition, we developed a transition plan and documentation to minimize the disruption for portal providers and users.

Our Year 7 plan includes investigating a replacement for our use of the Pubcookie product for doing web-based single-sign-on authentication. Our motivation for this stems from our assessment that Pubcookie is not adequate for developing a VO standard web sign-on. Fortunately, the open-source technology OpenID has gained considerable traction in the web community and satisfies our main requirements: it is defined by a standard document and it is widely supported by libraries in a broad set of languages. We completed some initial research into this technology and will begin a pilot project to provide OpenID authentication with the user logins we already support. We are also interested in investigating if OpenID can augment our use of X.509 certificates to support service-to-service authentication, message-level security, and delegation.

9 Applications

9.1 Data location services

The development of the NVO data discovery portal continued to be a central focus of the NVO software effort. The portal team achieved a number of milestones. The new registry interface was released at STScI. The NVO inventory service at IRSA and the integrating VIM services at Caltech were greatly enhanced and made more robust. The SimpleQuery (HEASARC) is now a fully supported operational service. A new table manipulation wizard was developed at Caltech and the table converter service at IRSA was much improved.

The key change in these services, however, was the increasing interoperability. Services now call each other appropriately (e.g., registry calling SimpleQuery to query a table, or that service calling the inventory and VIM). A format for transmitting a list of services was agreed upon and has been adopted by all services.

The portal was demonstrated at the spring team meeting. At that time there was considerable discussion of how users would begin portal usage and the consensus was that a simple initial page would be the most effective. More complex forms might be appropriate later.

A demonstration of the portal and initial release are scheduled for the IVOA meeting in Trieste in May 2008.

M. Preciado (HEASARC) began development of a notification service that would allow team members to post notices of operational issues (downtimes, etc.) Initial releases of the service were made and this service should go into operation in the next quarter. Preciado also continued to enhance the software tools for checking whether NVO (and VO) services are working properly. Tests of all of the portal services were added.

9.2 Cross-correlation services

M. Nieto-Santisteban (JHU) supported and helped users reporting problems with SkyNodes and Open SkyQuery. She also worked on the model for partitioning SkyNodes across a cluster of servers using a 10-TB simulated dataset built as a prototype for the Pan-STARRS science archive.

9.3 Image combination, registration
No activities to report this quarter.

9.4 Visualization tools and services

The VIM (Virtual Integration and Mining) application has made good progress in this quarter. It illustrates several new ideas that will extend and generalize, not just within the virtual astronomical observatory (VAO), but more generally to any of the other 'VxO' infrastructures that are growing up now to fully and richly virtualize all scientific data.

A key concept of the VxO is that data can be 'published, found, and bound'; meaning that there is a single, global library of data resources; anyone can build and publish data to that library; there are rich search tools for resources in that library; and the machines can automatically connect and utilize the found resources. Early in the NVO project, federation became important, leading to DataScope and other tools that can find out 'everything' about an astronomical source. After seven years, the global registry is well-established and has thousands of resources registered.

Another dimension of the VO concept is scalability, the recognition that modern astronomy is not only about studying single objects, but about populations of objects, where the number can range from a handful to a billion. Again, early in the NVO project, this was exploited with the OpenSkyQuery tool, where complex joint queries can be run against astronomical databases that may be large and geographically separated. A sophisticated architecture was built (SkyNode), and some success reported. However, it is a difficult technical matter for a user to create these queries, there have been difficulties with international acceptance of the required protocol, and awkward steps when the number of sources is scaled up.

VIM is a software architecture and toolbox, developed at Caltech, that allows users to investigate specified sources in multiple catalogs, with the idea that both the number of sources and the number of catalogs can be scaled up, and scaled not just in terms of the computer, but in terms of the human experience. The sources are specified by a set of identified sky positions, and the catalogs are actually any of the data services registered with the global registry. Here, we use the word 'catalog' generally, implying not only source catalogs, but also image and spectral surveys, observing logs, light curve archives, etc. VIM gives users persistent storage, called a 'workbench', which stores the original source table, then mashes up information about the sources from VO resources, so the source table gets rich annotation. Data is fetched from the VO in several protocols, so that all resources can be used through slow protocols (e.g., Cone Search), yet a resource with the fast protocol (e.g., SkyNode, TAP, SDSS) can be fully exploited. VIM offers a short list of 'primary' catalogs (e.g., SDSS, NED), together with access to the NVO registry for keyword and other semantic searching, as well as to the NVO Inventory service to find catalogs that are spatially correlated.

Some users want to compare a handful of sources, some want to work with thousands or millions. VIM is built to enable users to *start* by playing with a handful of sources—browsing the different catalogs and image surveys, understanding what the data means by reading rich metadata and the literature, visualizing, building strategy for bad and missing data, and so on. All this is done with a web browser that shows a view of a long, wide

table, together with a set of tools that can operate on the table and fetch data. As the user works, a script is automatically built, each command corresponding to mouse clicks in the browser; running that script against the VIM API library will then reproduce exactly what happened in the interactive session. However, the real utility of the script is so the user can edit it, removing dead-ends, extending and exploiting promising paths. Huge tables can be handled by the web interface, since it can show only a subset of the full data; huge tables can be handled by the underlying infrastructure, which is based on bulk access to data and code base (STILTS) that can handle tables with a billion rows.

VIM can run as a service, with workbenches on a remote server, or it can also run on a desktop or laptop, with data on the local disk. The server-based version allows users to assess and play with VIM before the commitment of a download and install; it means that powerful machines with big data and fast pipes can be brought to bear on big datasets; it also means that URL links to data can be sent to colleagues. The laptop version, however, can build workbenches without fear of purging by server administrators; it can exploit the scripting capability of VIM without the need to get an account at a computer center; it can run data fetching for hours or days without fear of timeouts. A local installation also means that programmer-users can build new tool components for the VIM toolbox, or modify the existing ones.

Each VIM workbench has a unique URL that contains a 32-digit random string, the 'bench ID', and so workbenches are secure, because a valid URL cannot be invented (security through obscurity). However, the workbench URL can be put in an email or blog, and thereby shared very simply; others just click on the URL and the whole visualization and toolbox appears in the browser. There is also a read-only mode for any VIM workbench that can only be unlocked with a password; this means that the creator of a workbench can, if they wish, share its content in a read-only mode, retaining the right to change the content.

VIM is built as a toolbox of tools. A tool, in this context, is a method that modifies the data in the workbench; it can be called by either a web form or a function call, and its code is separate from other tools, communicating only through the workbench. A tool consists of: a way to generate its web form control; a way to read the values that the user put in that form; a way to generate the equivalent script to the invocation; and a 'business logic' to take those values and compute on the workbench. A VIM application is a combination of these tools with the web-display technology; VIM uses the Yahoo User Interface to generate sophisticated menus, grids, and table displays. The astronomy-based toolbox described above is not the only one; the VIM download also includes a tiny reference implementation of a toolbox that can only do adding and multiplying numbers; in this way programmers could utilize the display and scripting capabilities of VIM, a clean separation from the tables of stars that is the business of the astronomy toolbox.

9.5 Theory

D. De Young (NOAO) and R. Wagner (SDSC) attended the EuroVO workshop “Theory in the Virtual Observatory” held in Munich (http://www.si.inaf.it/eurovow2008/agenda.htm).

R. Wagner is working with SimCat, a simulation catalog accessible through a RESTful web service API. It is designed to track simulations as they are being run, and to publish

data using a yet-to-be-defined IVOA-compliant manner. The catalog implements a simplified version of the simulation data model (see Section 2), and will be used to store descriptions of the data in the CADAC.

9.6 Statistical analysis
No activities to report this quarter.

9.7 Data mining, outlier identification
K. Borne (GMU) finished and submitted a chapter entitled "Scientific Data Mining in Astronomy" for a new book on Next Generation Data Mining. Borne continued his collaboration with UMBC distributed data mining experts—their focus is on peer-to-peer applications of data mining on extragalactic sources using the SDSS and 2MASS catalogs.

9.8 Interfaces to/from legacy software systems
M. Fitzpatrick (NOAO) worked on the IRAF VO package, bringing it closer to full-scale release.

10 Community Engagement

10.1 Documentation
We created the news@us-vo.org mailing list, comprising addresses from all summer school applicants, research initiative applicants, plus 80 people who signed up at AAS/Austin. An initial welcome message was sent out, and sign-up information was added to the main website. We currently have 319 addresses on the list, with more people signing up every week.

The first quarterly NVO Newsletter was distributed on March 12. Topics covered included the NVO book release, the 2008 Summer School, Google Sky, VO-CLI, and a VO calendar (http://www.us-vo.org/pipermail/news/2008-March/000001.html). The next issue will be sent out in June.

10.2 Web site
Updates were made to the NVO homepage, including the news mailing list signup, some reorganizing of the content, and the addition of footprint services to the applications listed. News items and sidebar links to the NVO book and the Summer School were also added.

10.3 Technical training initiatives
S. Emery Bunn traveled to Santa Fe, NM to evaluate facilities for the September 2008 NVO Summer School. She visited four hotels and met with sales reps at each, and met with CVB staff regarding social activities and dinner sites. We selected The Lodge at Santa Fe as the meeting venue, and Emery Bunn worked with the hotel to get a proposal and contract and to arrange an upgrade of Internet service for the meeting. A poster was designed and printed, and mailed to ~80 university astronomy departments in the U.S. Announcements were sent to the AAS, ADASS, IVOA, Women in Astronomy, and other mailing lists and web sites. An application form was set up on the NVO website and it was opened to accept applications (http://www.us-vo.org/summer-school/2008/).

Caltech hosted a Semantic Astronomy Workshop, Feb. 18-21 (http://www.cacr.caltech.edu/semast/).

10.4 Advocacy

We had an exhibit booth at the January AAS meeting in Austin. The theme was "NVO Inside", highlighting the extent to which the VO is permeating the data collections and data access services throughout the astronomical community. Many organizations displayed "NVO Inside" signs at their exhibits.

R. Hanisch (STScI) chaired a special session at the AAAS Annual Meeting in Boston, February 17, entitled "Virtual Observatories and Research Collaboratories: Internet Enabled Science." The program featured presentations about VO-like research environments in astronomy, space sciences, medicine/biology, and sociology. A. Szalay (JHU) spoke about the VO and data-intensive science, and A. Goodman (Harvard) spoke about the application of medical image processing techniques to astronomy.

11 Education and Public Outreach

11.1 Strategic partnerships

NVO EPO collaborations with external programs consists of cooperation with groups submitting new proposals to funding sources to create new educational programs or augment existing ones.

C. Christian (STScI) and J. Raddick (JHU) continue to participate in proposals with external collaborators. One preliminary proposal to NSF "CDI-Type II: Cyber-enabled Discovery of Cosmological Parameters through Distributed Community Research Environments" was submitted with measured success. It will be submitted as a preliminary proposal in the next round. Most of the focus has been on the use of the Sloan Digital Sky Survey because it is the most useable data source for a significant fraction of the sky as well as Galaxy Zoo.

K. Borne (GMU), C. Christian and J. Raddick are collaborators on a proposal to NSF submitted through George Mason University for studies of galaxy collisions and another by the LSST consortium on the use of data in archives for astronomy science education. Indications are that this effort has been funded.

C. Christian and STScI collaborators (primarily A. Conti) completed the augmentation of Sky in Google Earth (Sky) and are working on additional resources for that interface. Both individuals also are working with the Microsoft World Wide Telescope (WWT) group to bring accurate data from HST and GALEX to that interface. They also consulted with the major missions and Ground based observatories to assist in creating collections of their data for publication in the Sky and WWT interfaces.

11.2 Formal education

K. Borne (GMU) attended the American Association of Physics Teachers conference in Baltimore in January 2008. He presented one talk and two posters. The talk was "Astroinformatics: The New e-Science Paradigm for Astronomy Research and Education." The posters were "Robotic Telescopes for Engaging Students in Real Research Experiences" and another one with the same title as the talk. He was also co-author on two other astronomy education presentations at the conference.

11.3 Informal education

C. Christian, K. Borne and J. Raddick are collaborating, as described above, with colleagues on a project to use galaxy collision imagery and models into an interactive inter-

face for informal education activities. This is the CDI proposal. Other NVO EPO partners (such as the Adler Planetarium) are members of the collaboration.

11.4 Outreach and press activities

The core team, primarily J. Raddick, continues to post original contributions to the Data in Education Blog and encourage community participation.

11.5 Technical development

C. Christian regularly participates in the international group working towards implementation of metadata standards (VAMP/AVM) to insure that such standards are adapted to the visualization data for public release in 2008. We are working for endorsement of the standards and propagation of the use of the meta-tags throughout astronomy. The software and methodology have been shared with other observatories and presented at numerous conferences throughout the year. C. Christian is also working with other VO individuals to have the VAMP standards adopted by the IVOA.

Activities by Organization

Caltech–Astronomy Department and Center for Advanced Computational Research (CACR)

In February 2008 we published "The National Virtual Observatory: Tools and Techniques for Astronomical Research" as Vol. 382 in the ASP Conference Series (M. J. Graham, M. J. Fitzpatrick, and T. A. McGlynn, eds., with 56 contributing authors). This is a comprehensive review of the NVO design, technologies, tools, and programming interfaces. Ninety copies of the NVO book were sent to authors and participants in the NVO Summer Schools.

We co-organized the "NVO Inside" exhibit at AAS in Austin. Created posters, handouts, and booth coverage schedule. Distributed smaller "NVO Inside" signs to many booths throughout the exhibit hall, and gathered 150 names for those interested in NVO.

M. Graham ran the workshop "Practical Semantic Astronomy" with the assistance of S. Emery Bunn. http://www.cacr.caltech.edu/semast/

Produced VOSpace 1.1 Working Draft + WSDL and schema + reference implementation, and worked with JHU, CDS, UCSD, CADC on implementations of VOSpace.

Carnivore registry validated with Registry of Registries.

M. Graham taught at Australia VO Summer School.

Organized NVO Team meeting at Caltech.

R. Williams worked on NVO portal component Vim (Visual Integration and Mining), bringing it in this quarter from delicate prototype (almost) to alpha release.

Caltech–Infrared Processing and Analysis Center (IPAC)

R. Ebert, B. Madore, J. Mazzarella, and O. Pevunova (NED) attended the NVO Team Meeting held 19-20 March 2008 at Caltech. On 20 March, Mazzarella gave a presentation about the NED spectrum database. After highlighting the unique content and functionality of the current operational system (http://nedwww.ipac.caltech.edu/forms/SearchSpectra.html), he reviewed the complexities of capturing metadata from journal articles, ongoing work to provide VO-compatible access via the Simple Spectrum Access protocol, and the need to develop Web-based software tools to enable authors to contribute their data to the repository following VO standards for data and metadata curation. At the March team meeting, NED personnel and D. Tody (NRAO) formed a plan to implement an SSA interface into the NED spectrum database using the DAL Server Toolkit. During April Ebert set up a platform (Linux+Apache+Tomcat) to begin NED SSA development via DAL Server. On 4 April 2008, J. McDowell (SAO) visited IPAC and met with NED personnel. We reviewed the status of the VO SED data model (DM) and planned the tasks needed to construct a data access layer that translates between NED's existing internal SED DM and a VO SED DM based on "segments" within the VO spectrum data model. Pevunova continued to maintain 18 NED services in the NVO Registry, with access logs provided daily on an automated basis. Pevunova also participated in the weekly Technical Working Group telecons, and Mazzarella participated in the weekly Executive Committee telecons.

High Energy Astrophysics Science Archive Research Center (HEASARC)
A new member was added to the HEASARC's NVO team. Jake Wendt joins the HEASARC team from the INTEGRAL Science Data Center.

The HEASARC worked primarily in software development in the portal and software operations. T. McGlynn coordinated the efforts of the portal team. An initial release of the portal suite is now scheduled for the Trieste IVOA meeting May 2008. T. McGlynn and J. Wendt released the Simple Query tool, which allows users to query VO services, and also use supplied VOTables and can be used both within and outside of the portal. This software uses a common underpinning of XSLT and JavaScript with the registry services and legacy archive developments at STScI. J. Wendt has also take over responsibility for the upgrade to the DataScope tool that will be released in the next quarter. DataScope will also use the same underlying software capabilities. Wendt has also put up a home page for the portal.

M. Preciado has continued to augment software that monitors the health of all VO hosts. A simple report indicating the health of the VO is available through the NVO internal pages. More importantly, Preciado periodically reviews the status and has begun monitoring of the health of VO services as an operational responsibility. Whenever a failing service is discovered a dialog with the responsible parties is initiated until the issue is resolved.

T. McGlynn is a member of the NVO Executive committee. Wendt, Preciado, and McGlynn both attend the regular NVO technical working group telecons and support the general activities of the NVO including the NVO team meetings.

Johns Hopkins University
M. Nieto-Santisteban supported and helped users reporting problems with SkyNodes and Open SkyQuery, continued her work on the management of very large astronomical databases, and attended the NVO meeting held in Pasadena.

A. Thakar continues to maintain the NVO web logs and monitor the harvesting at JHU on a daily basis. He contributed quarterly and annual usage stats and charts for the NVO quarterly (4^{th} quarter 2007) and annual (2007) reports. Thakar is still interviewing candidates for a full-time VO position at JHU to work on the OpenSkyQuery re-engineering. Meanwhile, the model for partitioning SkyNodes across a cluster of servers has been developed and tested by M. Nieto-Santisteban (JHU) with a 10 TB simulated dataset built as a prototype for the Pan-STARRS science archive. Thakar continued to provide triage and support for technical problems with OpenSkyQuery and other VO Services at JHU, delegating to developers of the services as necessary. He also attended the VAO Proposal Team meeting (March 18) and the NVO Team Meeting (March 19-20) at Caltech.

G. Fekete worked to revise the Spherical Library. He added a way to directly compute the outline of a region consisting of more than 80 thousand convexes in a reasonable amount of time. He also developed two major improvements. First, the library now computes the outline of a region directly, so that the outline is an ordered list of arcs in which the end of one arc coincides with the beginning of the next arc. Second, for very large footprints, the convexes are organized in a hierarchical structure so that convexes in each bucket

have outlines unaffected by convexes of other buckets. This divide-and-conquer approach speeds up the process, as well as making it feasible to compute other processes in parallel.

A. Szalay worked with G. Fekete (JHU) on changes to the spherical library.

J. Raddick maintained the virtualobservatory.org public outreach website, and compiled the JHU reports for the 2007 annual report.

T. Budavari developed a new algorithm for running Principal Component Analysis on spectra, to be used with the VO Spectral Services. He rewrote VOSpace to work with SQL Server.

B. Bauer, a six-month visitor from the Technische Universität Munich (TUM), worked with T. Budavari and A. Szalay to develop and test a C# implementation of VOSpace 1.1 with a self-contained SQL Server 2005 backend. They are also working on a VOPipe Architecture that would provide higher-level services for data/workflows, which would be the basis for the next generation of VO Services, including the next generation SkyQuery.

National Optical Astronomy Observatories (NOAO)
In collaboration with B. Baker and R. Plante (NCSA), I. Barg led the installation of the mirror server for the NVO Single Sign On (SSO) service in Tucson. This mirror is now operational and has been integrated into the NVO SSO operational model as a fail-over server for the main SSO server hosted at NCSA. NOAO continues to use the SSO service as an integrated part of the NOAO VO Portal, providing access to proprietary data to principle investigators using NOAO's MOSAIC and NEWFIRM instruments.

M. Fitzpatrick continued work on the recently released VO Command Line Interface (VO-CLI), including enhancements for general use, compatibility with the new Registry interface, and SSAP support. VO-CLI was also modified to support its use in the SimpleQuery tool of the NVO portal being developed by the NVO team. VO-CLI was the featured application in the first NVO Newsletter. Additionally, a web page describing scripting usage of the VO-CLI was developed to encourage use in user-developed command-line scripting tasks. Fitzpatrick also provided support for NVO users who have downloaded and installed the VO-CLI tools.

Fitzpatrick participated in the ongoing development of the NVO Portal, both through the work on VO-CLI and other web pages that make up the Portal. This work included a demo of a Registry query form that interactively updates a display with the number of expected results (e.g. one search term shows 300 results, a second narrows this to 37, a bandpass term shows 5 results to user can easily understand).

Additional testing of the VOClient code and related IRAF VO package software with real data revealed problems with the tools. Fitzpatrick worked on fixes to the identified bugs, bringing the IRAF VO package closer to full-scale release.

Fitzpatrick invested additional effort in the proposed IVOA SAMP standard and related documentation. Various sections were rewritten as agreements on details regarding the protocols were worked out by the international group charged with finalizing the proposed standard. Fitzpatrick also contributed, as co-editor, to the finalization and publication of the NVO Book, which is now out. Fitzpatrick was asked to co-lead the 2008

NVO Summer School (NVOSS) and invested significant time in NVOSS preparations, including site selection and working with the NVOSS faculty to establish the content of this year's school.

P. Warner has continued work on the NOAO VOEvent software in collaboration with R. Seaman (NOAO), R. Williams (Caltech), and A. Drake (Caltech). This collaboration is partially funded by NVO and partially by an AISRP grant from NASA. An incremental version of the software is planned for release in May 2008, which includes closer adherence to the IVOA VOEvent specification, an initial framework for classification of events, and query and retrieval of events as KML for use in, e.g., Google Sky.

R. Massad has completed an initial prototype testing application for web-based NVO tools and services. This work was undertaken with support from C. Miller and A. Egaña. This prototype can be used to program both mouse and text input to a web application and parse results for correctness. Massad is continuing to work on the prototype to enhance robustness and usability.

National Radio Astronomy Observatory (NRAO)
NRAO participation in the NVO effort over the past quarter has emphasized participation in ongoing IVOA standards development and related implementations, participation in planning for the proposed operational phase of the US VAO, and ongoing production and publication of VO-ready data from NRAO instrumentation including the VLA.

In the area of standards development most of the effort was focused on development of the Table Access Protocol (TAP), and a second version and major upgrade to the Simple Image Access Protocol (SIAV2), both efforts being led by D. Tody. Work continues on spectral access as well, including preparation a capability schema for SSA, and registry support for legacy and standard SSA services so that these can be described by the new registry. In the area of data analysis standards, work is underway on interface specification for the applications framework project, which is a joint effort of NVO within the US, and OPTICON within Europe.

J. Crossley, a post graduate student who recently joined the NRAO VO effort, has begun work on software for load testing VO services, as part of the planned facilities for service verification and testing.

Raytheon/ADC (University of Maryland and George Mason University)
George Mason University staff K. Borne participated in the "Practical Semantic Astronomy" workshop at Caltech in February 2008. He gave a talk on "A Classification Broker for Petascale Sky Surveys." In addition, GMU staff participated in the NVO project team meeting at Caltech in March. Borne finished and submitted a chapter for a new book on Next Generation Data Mining—this chapter is titled "Scientific Data Mining in Astronomy." Borne continued his collaboration with UMBC distributed data mining experts—their focus is on peer-to-peer applications of data mining on extragalactic sources within the combination of the SDSS and 2MASS catalogs.

Smithsonian Astrophysical Observatory
J. McDowell visited IPAC for discussions with the NED team.

After a long search, SAO has hired James Cant as a VO programmer. Jim has extensive Java development experience and a background in science; he began work on Apr 28.

A. Rots conferred with G. Greene (STScI) and T. Budavari (JHU) on footprint standards and participated in the footprint telecon on April 24. A. Rots commented on the treatment of Regions in a draft of the ADQL proposal, deeming it inadequate. A. Rots attended the AAS meeting in Austin, TX, January 7-11 and provided support for the NVO exhibit. Also attended the NVO team meeting March 19-20 in Pasadena, CA.

G. Fabbiano also attended the team meeting.

Space Telescope Science Institute
Based on feedback from the IVOA Executive and IVOA Working Group and Interest Group chairs, the week of October 27-31 has been selected for the fall IVOA Interop workshop. STScI will co-host the meeting with JHU.

G. Greene and T. Dower attended the AAS meeting in Austin to support the NVO Exhibit.

Dower worked on improving the STScI registry search and associated web pages, bringing the OAI interface into compliance with the VOResource V1.0 schema requirements, and beginning the design and implementation of a new publishing interface for the STScI registry. For the keyword search page, the filtering mechanisms for search results and the templates for their display were improved. Basic examples and a usage FAQ for the search page were written. The OAI web service interface to the STScI registry was updated so that it can be validated and included in the Registry of Registries. Work also began on the publishing interface, which will make it easier for data-providing organizations to register their holdings and services and help to assure higher quality metadata. The new user interface borrows deeply from Carnivore (Caltech's XML-based registry), and is similarly based on the XForms technology with dynamic form content dependent on data relevance for the various resource types. As part of the publishing interface, a user registration system was implementing utilizing our Microsoft SQL Server registry database. Once the basic architectural design and the login system were completed, work began on the forms and data processing to include new resources in the registry.

Greene upgraded the registry harvesting process to include the HEASARC and NCSA OAI publishing registries. Once CDS completes the Vizier OAI registry we will be able to resume full population and provide nearly equivalent content search availability. The goal is to prepare for April/May final testing of the STScI registry in compliance with the new metadata and interface standards. We will decide at that time to accept the new resource publication interface or to continue with the public release using the NCSA publication portal.

Greene began work on a draft footprint service specification that will be presented at the IVOA Interop workshop in May. This is a collaborative work with T. Budavari and A. Szalay (JHU), A. Rots (SAO), M. Dolenksy (ESO/EuroVO), and F. Bonnarel (CDS/EuroVO). Greene also worked with staff at HEASARC to resolve some problems in harvesting resource metadata from their publishing registry.

C. Christian and J. Raddick (JHU) continue to discuss with a variety of groups ideas for education with large data archives. The LSST-related EPO activity is being formulated for input to NSF.

Hanisch chaired a special session at the AAAS Annual Meeting in Boston, February 17, entitled "Virtual Observatories and Research Collaboratories: Internet Enabled Science." The program featured presentations about VO-like research environments in astronomy, space sciences, medicine/biology, and sociology.

Hanisch completed revisions on the IVOA Document Standards process document, which clarify the review and approval process, particularly with regard to the Technical Coordination Group (the chairs and deputy chairs of all IVOA working groups and interest groups).

K. Gillies joined the staff of the Archive Sciences Branch on Feb 19. He was formerly at Gemini South. Gillies will be spending some of his time on VO-related tasks such as footprint services and distributed data storage implementations.

Hanisch, Greene, Dower, Gillies, and Christian attended the spring NVO team meeting at Caltech, March 19-20. Hanisch, Greene, Dower, and Gillies participated in the weekly NVO Technical Working Group telecons.

University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA)

R. Plante and B. Baker lead the development and support for secure VO logins (see section 8). Plante also leads the Registry WBS, working with team members at STScI (G. Greene and T. Dower) and Caltech (M. Graham). Plante continues to contribute to the development of our Year 7 software process.

Publications and Presentations

"The National Virtual Observatory: Tools and Techniques for Astronomical Research," M. Graham, M. Fitzpatrick, & T. McGlynn, eds., ASP Conf. Ser. 382.

"The Virtual Observatory Vision." A. Szalay, AAAS Annual Meeting, Boston, Feb. 2008.

"Data-Intensive Science in Astronomy and Medicine." A. Goodman, AAAS Annual Meeting, Boston, Feb. 2008.

"VOClient: Application Integration in the Virtual Observatory." M. Fitzpatrick & D. Tody, AAS Meeting, Austin, Jan. 2008. (47.14)

"VOEventNet: An Open Source of Transient Alerts for Astronomers." A. Drake, R. Williams, M. Graham, A. Mahabal, S. Djorgovski, R. White, W. Vestrand, & J. Bloom, AAS Meeting, Austin, Jan. 2008. (47.13)

"Everybody Has a Sky." C. Christian & A. Conti, AAS Meeting, Austin, Jan. 2008. (70.04)

"NRAO VLA Archive Survey (NVAS)," J. Crossley, L. Sjouweman, E. Fomalont, & N. Radziwill, AAS Meeting, Austin, Jan. 2008. (132.03)

"Transients from the Palomar Quest Sky Survey: Recent Results and Status," A. Mahabal, A. Drake, S. Djorgovski, C. Donalek, E. Glikman, M. Graham, R. Williams, C. Baltay, & D. Rabinowitz, AAS Meeting, Austin, Jan. 2008. (108.04)

"Footprint Tool for the Hubble Legacy Archive," S. Lubow & G. Greene, AAS Meeting, Austin, Jan. 2008. (47.12)

"NVO Newsletter No. 1", http://www.us-vo.org/newsletter/1/

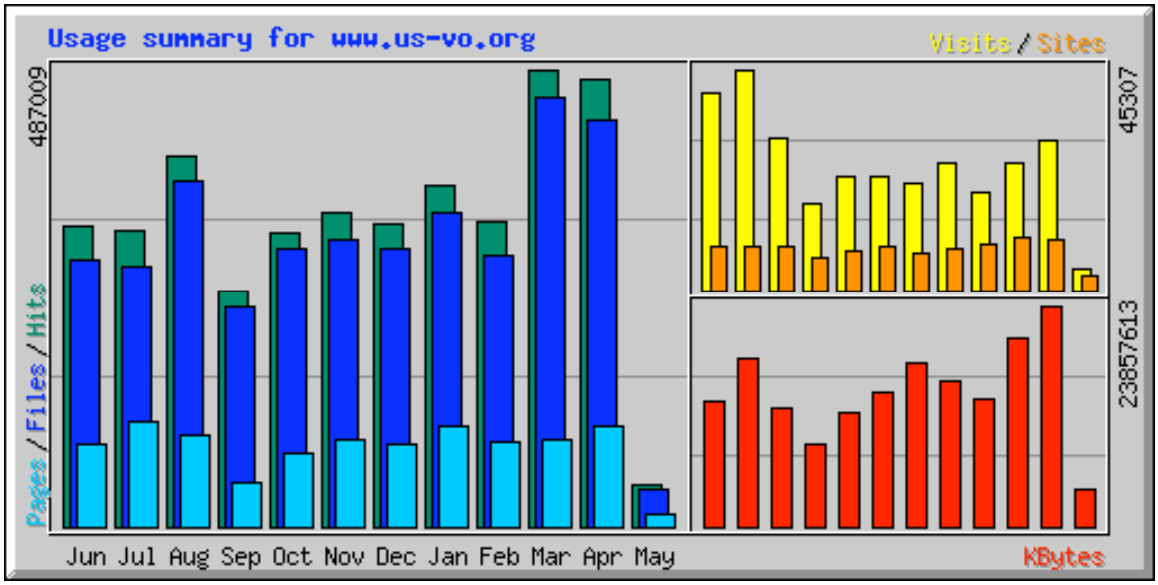
Virtual Observatory Articles in the Popular and Technical Press

None this quarter.

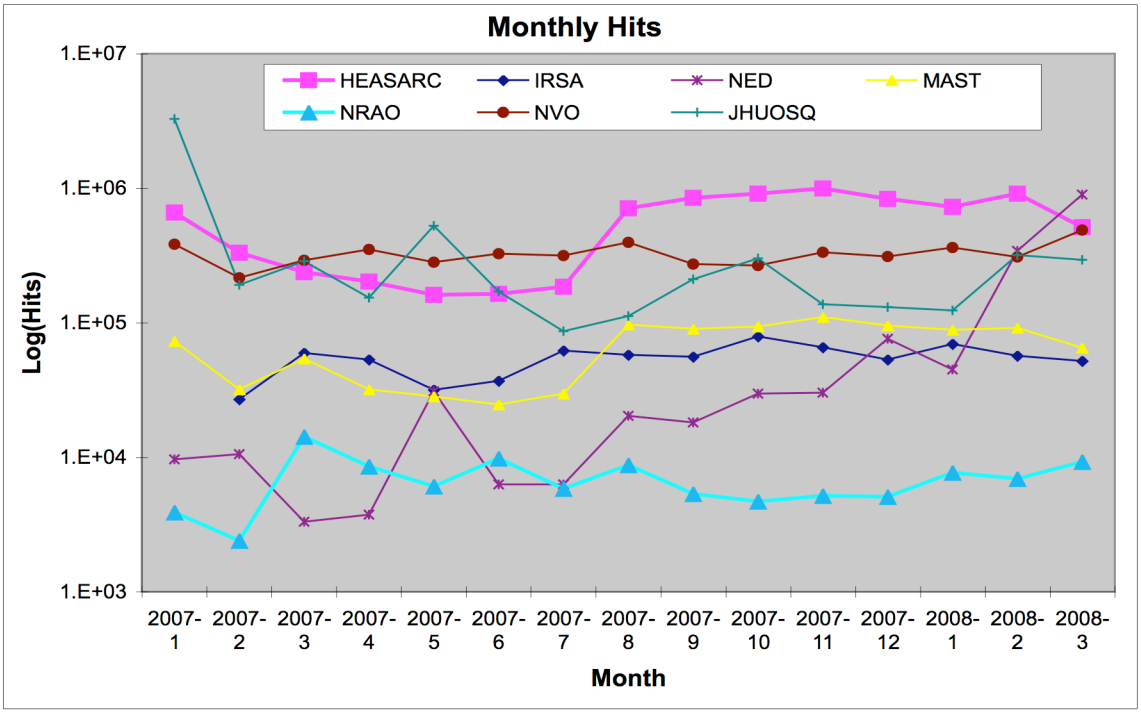
Usage Logs

A number of NVO participating organizations have implemented standard interfaces to their web and service logs, and we have begun collected these logs in order to track VO-enabled use of data and services. The main NVO web site continues to be used frequently, with an all-time peak in February of nearly 500,000 hits.

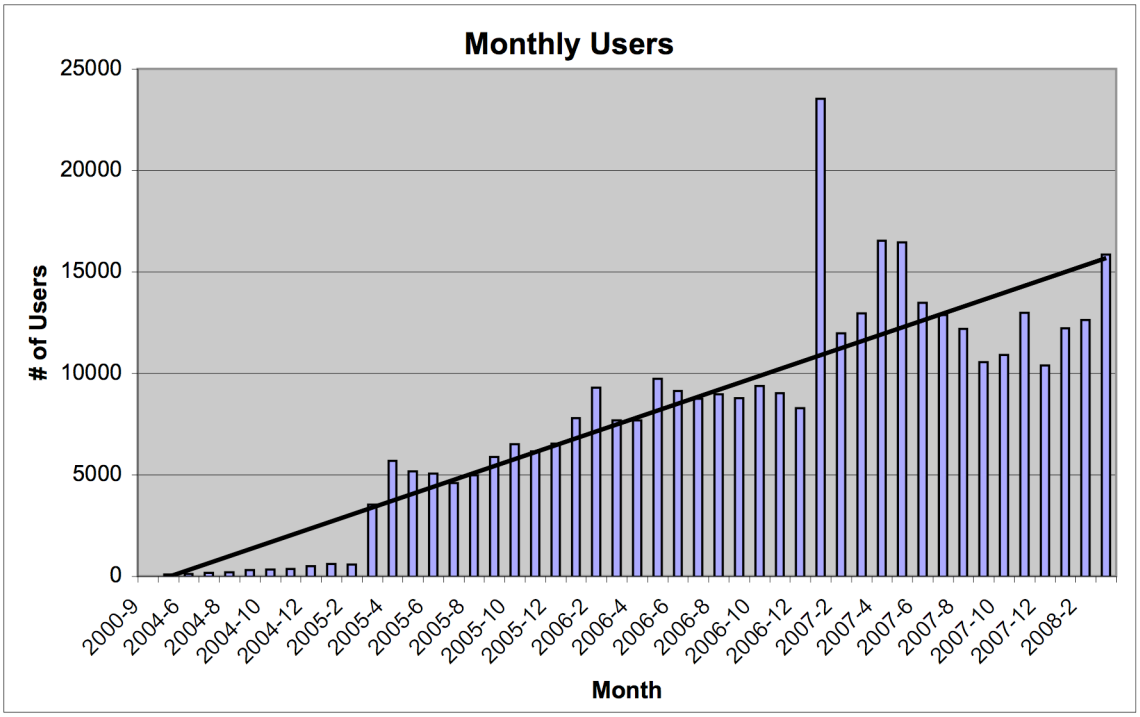
Monthly Usage of NVO Web Site (http://www.us-vo.org):



Monthly Hits at NVO Organizations Originating from NVO Applications:



Monthly Users of Any NVO Application or Website:



Acronyms

AAS	American Astronomical Society
ADC	Astronomical Data Center
ADEC	Astrophysics Data Centers Executive Committee (NASA)
ADQL	Astronomical Data Query Language
AIPS++	Astronomical Image Processing System++ (NRAO)
API	Applications Programming Interface
AVO	Astrophysical Virtual Observatory
CACR	Center for Advanced Computational Research (Caltech)
CADC	Canadian Astronomy Data Centre
CDS	Centre de Données astronomiques de Strasbourg
CMU	Carnegie Mellon University
CXC	Chandra X-Ray Center
CY	calendar year
DAG	Directed Acyclic Graph
DAGMan	Directed Acyclic Graph Manager (Condor)
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DIS	Data Inventory Service
DM	Data Model
DOE	Department of Energy
DPOSS	Digitized Palomar Observatory Sky Survey
DTD	Document Type Description
EPO	Education and Public Outreach
ESTO	Earth Science Technology Office (NASA)
ESTO-CT	ESTO Computational Technologies (NASA)
FIRST	Faint Images of the Radio Sky at Twenty Centimeters
FITS	Flexible Image Transport System
FNAL	Fermi National Accelerator Laboratory
FTP	File Transport Protocol
FY	fiscal year
GB	gigabyte
GLU	Générateur de Liens Uniformes (uniform link generator)
GRB	Gamma Ray Burst
GriPhyN	Grid Physics Network
HEASARC	High Energy Astrophysics Science Archive Center
HTTP	HyperText Transport Protocol
IPAC	Infrared Processing and Analysis Center (Caltech)
IRAF	Image Reduction and Analysis Facility (NOAO)
IRSA	Infrared Science Archive (IPAC)
ISI	Information Sciences Institute (USC)
ITWG	Information Technology Working Group (NASA data centers)
iVDGL	International Virtual Data Grid Laboratory
IVOA	International Virtual Observatory Alliance
JDBC	Java Data Base Connectivity (Sun, Inc., trademark)

JHU	The Johns Hopkins University
MAST	Multimission Archive at Space Telescope (STScI)
MB	megabyte
MOU	Memorandum of Understanding
MWG	Metadata Working Group
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputer Applications
NED	NASA/IPAC Extragalactic Database
NESSSI	NVO Extensible Secure Scalable Service Infrastructure
NOAO	National Optical Astronomy Observatories
NPACI	National Partnership for Advanced Computational Infrastructure
NRAO	National Radio Astronomy Observatory
NSF	National Science Foundation
NVO	National Virtual Observatory
OAI	Open Archives Initiative
OASIS	On-line Archive Science Information Services (IRSA)
OGSA	Open Grid Services Architecture
OIL	Ontology Inference Layer
OWL	Web Ontology Language
PB	petabyte
PMH	Protocol for Metadata Harvesting (of OAI)
Q	quarter
QSO	Quasi-Stellar Object
RC	Replica Catalog
RDF	Resource Description Framework
REST	Representational State Transfer
RLS	Replica Location Service
ROME	Request Object Management Environment
SAO	Smithsonian Astrophysical Observatory
SAWG	Science Archives Working Group (NASA)
SAWG	System Architecture Working Group (this project)
SciDAC	Scientific Discovery through Advanced Computing (DOE)
SDSC	San Diego Supercomputer Center
SDSS	Sloan Digital Sky Survey
SDT	Science Definition Team
SIAP	Simple Image Access Protocol
SOAP	Simple Object Access Protocol
SRB	Storage Resource Broker
SSAP	Simple Spectral Access Protocol
STScI	Space Telescope Science Institute
SWG	Science Working Group
TB	terabyte
UCD	Unified Content Descriptor
USC	University of Southern California
UDDI	Universal Description, Discovery, and Integration
UIUC	University of Illinois Champaign-Urbana

USNO	United States Naval Observatory
USRA	Universities Space Research Association
VDL	Virtual Data Language
VDS	Virtual Data System
VO	Virtual Observatory
VO	Virtual Organization
VOQL	Virtual Observatory Query Language
WBS	Work Breakdown Structure
WebDAV	Web-based Distributed Authoring and Versioning
WSDL	Web Services Description Language
XML	Extensible Mark-up Language
2MASS	Two-Micron All Sky Survey