

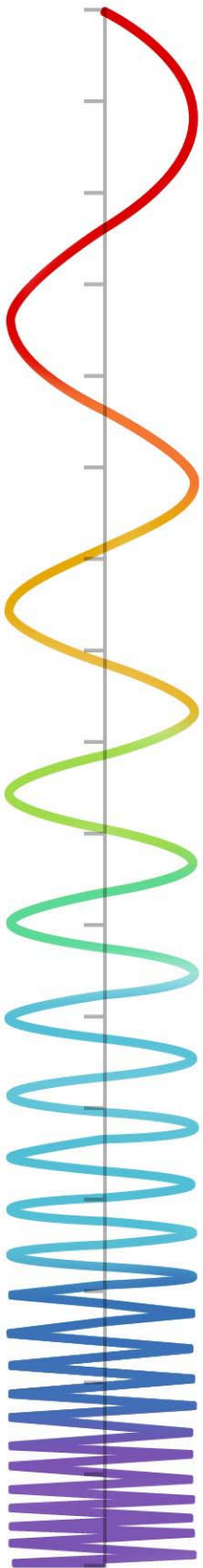
Quarterly Report
April-June 2005

Building the Framework for the
National Virtual Observatory

NSF Cooperative Agreement
AST0122449



INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE



Executive Summary	1
Activities by WBS	2
1 Management	2
1.1 General (planning, reporting, communications, team meetings, etc.)	2
1.2 Science	2
1.3 Technical (including standards, configuration management)	2
1.4 Financial.....	3
1.5 International coordination/collaboration.....	3
2 Science Requirements	3
2.1 Usage scenarios for all areas of astronomy research, including theoretical simulations	3
2.2 Requirements analysis	3
2.3 Demonstration definition and review.....	4
3 System Architecture	4
3.1 System design, components; relationships to Grid components.....	4
3.2 Computational facilities (processing, bulk data storage, network access, security, authentication).....	5
3.3 Digital library integration	5
4 Registries	6
4.1 Resource metadata	6
4.2 Resource metadata schema	6
4.3 Publishing and harvesting protocols	7
4.4 Search protocols.....	7
4.5 Replication, synchronization, maintenance, revision control, and curation	7
5 Data Models	8
5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality).....	8
5.2 Low-level (measurement, quantity, uncertainty, relationship)	9
5.3 Descriptors and ontologies (UCDs)	9
5.4 Space-Time and regions.....	9
5.5 Standard schema	9
6 Data Access Layer	9
6.1 Data access services (catalog, image, spectrum, time series, visibilities, ...)	9
6.2 Data representation (VOTable, etc.)	11
6.3 Framework (mediators, components)	12
6.4 Data provider/consumer implementations and end-to-end testing	12
7 Query Language	13
7.1 Low-level: Astronomical Data Query Language	13
7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode	13
7.3 High-level: Complex queries	13
8 Web and Grid Services	13
8.1 Web Services (SOAP, WSDL, etc.).....	13
8.2 Grid Services (OGSA)	13
8.3 Computational resource management.....	14
8.4 Virtual data.....	14

8.5	Application and service integration with Grid.....	14
9	Applications.....	14
9.1	Data location services	15
9.2	Cross-correlation services.....	15
9.3	Image combination, registration	15
9.4	Visualization tools and services.....	15
9.5	Theory	15
9.6	Statistical analysis.....	15
9.7	Datamining, outlier identification.....	15
9.8	Interfaces to/from legacy software systems.....	15
10	Community Engagement.....	15
10.1	Documentation.....	15
10.2	Web site	16
10.3	Technical training initiatives.....	16
10.4	Advocacy	16
11	Education and Public Outreach	17
11.1	Strategic partnerships.....	17
11.2	Formal education	17
11.3	Informal education.....	17
11.4	Outreach and press activities	18
11.5	Technical development	18
	Activities by Organization.....	19
	Caltech–Astronomy Department and Center for Advanced Computational Research (CACR).....	19
	Caltech–Infrared Processing and Analysis Center (IPAC).....	19
	Canadian Astronomy Data Centre/Canadian Virtual Observatory.....	19
	Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt).....	20
	Fermi National Accelerator Laboratory (FNAL).....	20
	High Energy Astrophysics Science Archive Research Center (HEASARC).....	20
	Johns Hopkins University.....	21
	Microsoft Research	22
	National Optical Astronomy Observatories (NOAO).....	22
	National Radio Astronomy Observatory (NRAO)	23
	Raytheon/ADC (University of Maryland and George Mason University).....	24
	San Diego Supercomputer Center.....	24
	Smithsonian Astrophysical Observatory.....	25
	Space Telescope Science Institute	26
	United States Naval Observatory.....	26
	University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA).....	27
	University of Southern California (USC/ISI)	27
	University of Wisconsin	28
	Publications	30
	Virtual Observatory Articles in the Popular and Technical Press.....	31
	Acronyms.....	32

**Building the Framework for the National Virtual Observatory
NSF Cooperative Agreement AST0122449
Quarterly Report**

Period covered by this report:

1 April—30 June 2005

Submitted by:

Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

The primary focus of NVO activities this quarter was the major IVOA Interoperability Workshop, which was held in Kyoto, Japan in May. NVO team members contributed major updates to the registry Resource Metadata definitions and schema. The Simple Spectral Access Protocol, and VOStore specification, the ADQL specification, the new VOEvent specification and protocol, and methods for authentication and authorization were presented and discussed, with much positive feedback, in Kyoto.

Within the NVO project substantial progress was made in curation of the resource registry records, and subtle problems in harvesting from publishing registries associated with time zones were resolved.

Major effort has gone into preparations for the second NVO Summer School. The longer duration of this year's Summer School allows for a two-day "pre-school" to familiarize participants with core VO technologies. A software development kit is being prepared and will be made available to participants for installation well in advance of the Summer School.

The NVO Project submitted a successful supplementary proposal to NSF to support Summer School expenses, but more importantly to create a program of small research initiatives to jump-start community-based, VO-enabled science projects. We will especially encourage NVO Summer School participants to apply for these stipends in order to follow through on projects started at the Summer School.

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, and the Project Manager and WBS leaders have a biweekly telecon to discuss programmatic issues.

The spring team meeting was held in Tucson, 25-26 April, and was co-hosted by LSTT and NOAO. The focus of the meeting was on preparations for the May IVOA Interop workshop in Kyoto. We also discussed requirements for handling time-domain datasets with our LSST co-hosts and a member of the PanSTARRS team.

1.2 Science

Plans were made for a summer meeting of the Science Steering Committee, to be held in conjunction with the summer Team Meeting. These will be hosted by the University of Pittsburgh in late July.

Senior NVO team members are participating in the science organizing committee for an IAU Special Session entitled “The Virtual Observatory in Action: New Science, New Technology, and Next Generation Facilities.”

1.3 Technical (including standards, configuration management)

The NVO Technical Working Group, chaired by J. Good (IPAC) has focused this quarter has been on the following issues:

- Cleaning up the NVO Registry content (validating services and service metadata) and documenting Registry use.
- Defining, setting up, and populating an official NVO code repository (CVS). This has been done at CACR. The code repository was also deemed critical path for properly supporting the NVO Summer School.
- Defining VOspace infrastructure. This is critical to supporting long-lived and/or complex VO interactions as it provides the “working space” for the distributed NVO processes. Tied to this is the issue of security, since interactions between users, processes, and VOStores cannot be done without some sort of user validation and authorization.
- Formal testing, tracking, etc. This is a wide-ranging topic, including software acceptance, service validation, and so on. Much of this is beyond the scope of the current NVO research project but it was felt it was important to at least discuss it in the context of future planning.
- There have also been preliminary discussions (though no development as yet) relating to asynchronous services, usage logging, and workflow.
- Tests of the SOAP services (and particularly the SkyNode interface) using the gsoap package. So far this has mixed success: the simpler interfaces work well (e.g. getting

a table list or a column list for a table). However, the actual catalog query interface is heavily layered and polymorphic, and it has proven difficult to make that work. Gsoap is used here for two reasons: a C/C++ implementation will be faster and lighter-weight than Java and will overcome Java's memory limitations.

1.4 Financial

Overall project spending continues as planned. Our current carry-forward is \$860k, down \$90k from last quarter. This is as planned, and is sufficient to keep the project team in place well into Year 5. Indeed, projecting from the current slope of the decreasing underrun, we anticipate closing out the project very close to the overall target budget. Planning is getting underway in the Executive Committee to consider budget allocations for the final year of the project.

1.5 International coordination/collaboration

A number of NVO team members participated in the IVOA Interoperability Workshop in Kyoto, Japan, 16-20 May. Over eighty VO personnel from projects around the world participated in the week-long workshop. JHU database expert M. Nieto-Santisteban has been asked to co-chair the IVOA VO Query Language Working Group. R. Williams has assumed the chair of the IVOA Event Working Group, which is defining the international standard for exchanging information about transient events. R. Hanisch attended the EuroVO workshop in Garching, Germany, 27 June-1 July, to support EuroVO's efforts in technology take-up and training. Seeing what worked well, and not so well, provided extremely useful information in planning for the NVO Summer School II.

2 Science Requirements

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

An immediate priority for the US NVO is the creation and dissemination of science applications that will be readily accepted by the US astronomy community. After the successful demonstration and deployment of five science applications at the January 2005 AAS meeting, the next tasks in the science requirements area will be to develop a list of candidate "next phase" applications. This process will be initiated, as before, through the NVO Science Steering Committee (SSC). This committee has already provided a tentative list of such applications from its extended discussion during the 2004 SSC meeting in Aspen. Further development of this list will take place during the July 2005 SSC meeting in Pittsburgh. That meeting will also address other key science requirements issues such as community engagement and the evaluation of proposals for funding of small grants under the NVO Research Applications program.

2.2 Requirements analysis

The development of the next generation of science applications will be reviewed for technical feasibility by the NVO Team. The first such review will occur immediately after the SSC meeting in Pittsburgh during the NVO Team meeting at that same location. In addition, the SSC will, throughout the year, be consulted about the continuing scientific viability of these applications projects as they are developed.

2.3 Demonstration definition and review

No activities in this area during this quarter.

3 System Architecture

The system architecture is now relatively well fixed with the following components:

- Portals such as sky query
- Web services (currently WSDL based)
- Data grid (an implementation based on VOSpace and VOStore is being pursued)
- Collections (both original collections from sky surveys, and replications of collections onto more powerful computing resources)
- Grid (based on Globus Tool Kit)
- Infrastructure (resources include the NSF TeraGrid, project specific systems)

The remaining technologies that are being evaluated for inclusion in the system architecture are:

- Digital libraries (two examples are DSpace from MIT and Fedora from Cornell University and University of Virginia)
- Semantic grid technology (support for reasoning across inferred attributes associated with items in collections through use of ontologies)

3.1 System design, components; relationships to Grid components

The VOStore and VOSpace system components were discussed at the IVOA meeting held in Kyoto on May 16-20, 2005. Given the multiple implementations that are being evaluated for VOStore, this topic is now relegated to the Web Services section of the report.

At Kyoto, the discussion on VOSpace included comparisons between the VOSpace design and the existing Storage Resource Broker data grid technology (SRB). It was recognized that the SRB provides the capabilities that are envisioned for VOSpace.

At the Global Grid Forum meeting, held in Chicago on June 27-29, a new naming infrastructure was proposed by the Grid File System working group. A central issue in VOSpace is an understanding of the appropriate levels of naming indirection that are needed to allow images in existing sky surveys to be included in the VOSpace file name space. The Global Grid Forum now uses three levels of naming for files that are being federated across sites:

- Resource Naming Service. This is a user-friendly logical file name that can be organized in a directory hierarchy.
- WS-names. These are unique abstract names that can be expressed as Globally Unique Identifiers, or Object Identifiers, or handles. This provides an indirection layer to allow multiple types of name resolvers to be used in the Grid.
- WS-addressing. These are the end-point references that point to the actual location of the file.

It is possible to build data grids that have only two levels of naming indirection (logical

file names and end-point references). Such a system assumes a single type of name resolver is sufficient to map from the logical file name to the actual location.

One question for the DSpace implementation will be whether two or three levels of naming indirection are needed for federating existing collections.

3.2 Computational facilities (processing, bulk data storage, network access, security, authentication)

Storage resources and collections. SDSC has made a major commitment to the NVO project to provide 50 Terabytes of disk space to house the major NVO sky image collections on the TeraGrid. While the disk has been installed at SDSC, testing and final configuration of the file systems is still being done. The first use of the disk cache will be to hold an image of the final release of the 2MASS sky survey. The TeraGrid disk and compute resources will be used to create mosaics for deposition into the Hyperatlas that is being created by Roy Williams.

A Java implementation and demonstration of the VOStore interface has been demonstrated, which includes the full security infrastructure previously agreed at the IVOA and which is based on the WS-Security model of the W3C. This implementation has been shown to interoperate successfully with the JHU .NET implementation.

TeraGrid data intensive applications. The project to build mosaics of 2MASS is being supported by two teams (J. Good, L. Brieger, R. Williams) and (J. Good, E. Deelman, J. Jacobs, R. Williams). The two separate approaches include:

- Optimal parallelization of the mosaic code for execution on a single TeraGrid node (L. Brieger)
- Distributed processing of the mosaic code for execution across multiple TeraGrid nodes (E. Deelman)

The parallelization of the mosaic code for a single TeraGrid node combines the following strategies:

- Assignment of the construction of individual mosaics to separate processors in the cluster.
- Parallelization of mosaic modules such as the co-add function through use of MPI calls.

The two levels of parallelization enable the generation of the mosaics to make optimal use of the computing resources. An initial test run has been completed for the Galactic Plane (450 plates) and for the Virgo Cluster (18 plates). Each band (150 plates) took between 9 and 10 hours to generate using 150 processors on the TeraGrid, implying each 6-degree by 6-degree mosaic took less than 4 minutes. The initial tests revealed problems with system I/O and mosaic processing that were fixed by modifying the MPI-based code and installing the latest version of J. Good's code.

3.3 Digital library integration

SDSC is pursuing two digital library efforts:

- Provision of a DSpace instance for the creation of personal collections. A total of 5

Terabytes of disk space will be allocated to a DSpace instance that has been installed at SDSC. This is a collaboration with MIT to explore use of the integrated DSpace/SRB data grid system.

- Porting of the Fedora digital library on top of the SRB data grid. The port has been started, with an instance of the Fedora digital library installed at SDSC.

4 Registries

In the last quarter, focus has returned to advancing registry-related standards through the IVOA registration process, although general registry clean up and improvements continue. To help keep the standardization process on track, the IVOA Registry Working Group defined a revised roadmap that ends with global support for Registry Interface standard, v1.0 in early 2006. It is worth noting that there are a number of standards-related changes that must happen along the way (e.g. metadata schemas, OAI profile, interface). Instead of upgrading our implementations many times over the next six months to be compliant with each change, we have decided to postpone full adoption until all the changes to the standards are complete.

4.1 Resource metadata

After presentation to the IVOA Registry Working Group at the IVOA Interoperability Workshop in May, R. Hanisch released version 1.1 of the IVOA Specification, *Resource Metadata for the Virtual Observatory* (RM) as a Working Draft for more general review (see <http://www.ivoa.net/Documents/latest/RM.html>). One of the most significant of the changes is the introduction of the *ResourceValidationLevel* metadatum, which allows registry curators to rate resource records (and the associated service interfaces) by their conformance to standards as well as the quality of the descriptions that make them useful to VO applications. The working group recommended that an resource metadata encoding (e.g., VOResource for XML) allow this metadatum to be tagged with a reference to the registry or organization that assigned the value; this would allow resource descriptions to include multiple *ResourceValidationLevel* values to be included in a description.

4.2 Resource metadata schema

R. Plante generated the first working draft of VOResource: *an XML Encoding Schema for Resource Metadata*, the document that defines and describes the VOResource XML Schema. This document reflects a stable format for the core resource metadata defined in the RM that has matured through several years of actual use by VO applications. It was presented to IVOA Registry Working Group and is on track to be moved through the IVOA standardization process by the end of the year. With this core standard moving to a version 1.0 Recommendation, further evolution of our VO metadata will occur in specialized extension schemas.

One such extension is the VODataService schema, which defines metadata for resources specifically related to data collections and services that access them. Plante is leading a small IVOA tiger team to do a revision of this schema to incorporate some of our lessons learned registering services. In particular, we are addressing the CDS problem of many catalogs being served by a single or few services. With more effective separation of the description of services and data collections (as separate resources) it will be possible to

reduce the number of services that a large repository like CDS would have to register. We are also looking at adopting the AstroGrid approach to registering “application services”—services that are especially well suited to incorporation into automated workflows. There is also a need to describe more traditional software applications and packages that run on a user’s local platform but can access VO services.

4.3 Publishing and harvesting protocols

In the last quarter, we developed the idea of a Registry of Registries that provided a central doorway for new registries to enter the VO web. This registry would be run by the IVOA and would include browser-based interface that allows a new registry curator to register his/her registry. Other harvesting registries would harvest from the central registry to determine who to harvest from. Since then Plante has refined the definition and put up a prototype implementation (<http://nvo.ncsa.uiuc.edu/VORoR/check.html>). This prototype was presented at the May IVOA Interoperability Workshop and was accepted by the Registry Working Group. Technical arrangements have been made with the IVOA webmaster for migrating Registry of Registries to the *ivoa.net* domain when we transition to the Registry Interface v1.0 in early 2006.

4.4 Search protocols

The Registry Working Group (RWG) noticed that there were a few types of registry queries that could not be easily expressed using ADQL. This motivated the decision to add optional support for XQuery to the Registry Interface specification.

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

In the last six months, we have seen a blossoming of new registries around the world, including in the US. Publishing registries are being installed at the Harvard-Smithsonian Center for Astrophysics and the National Optical Astronomy Observatory (NOAO). In addition to installing their publishing registry, M. Fitzpatrick has created some interesting end-user services that leverage the programmatic interface to the full registry at Space Telescope. His interface includes several helpful features for drill-down searching. For example, keyword-based search results are updated in real-time as the user is typing. These results are categorized by resource type, allowing the user, after seeing an initial set of results, to quickly zero in on the resource types of interest. In addition to a browser-based interface, he has packaged the service into a MacOS “dashboard widget”, providing perhaps a premonition of what the window into a successful VO might look like.

This quarter, G. Greene made a number of important upgrades to the STScI registry. In addition to changes to track our evolving draft interface standard, a number of improvements reflected feedback from our curation studies. These include improved documentation, correction of errors in metadata, and correction miscellaneous system errors. The STScI registry currently harvests from six other registries distributed worldwide (Japan VO, NCSA, Caltech, UK’s AstroGrid, CDS’s Vizier, and HEASARC). The contents are updated every hour with a total of some 12,000 resources now registered. In related work, M. Graham (Caltech) released version 1.3 of the Carnivore XML-based registry,

which features an improved interface and an improvement in search performance by a factor of 5.

IPAC/NED has now completed VO-compliant version of most commonly used functionality. This includes

- Search by object name
- Search near object by name
- Search near object by IAU name
- Search by object reference code
- Advanced all-sky searches
- Search references
- Search NED notes
- Search position information
- Search redshift information
- Search photometry data
- Search NED "diameter" data

These searches are in addition to the NVO-standard NED Cone Search (which encompasses the standard NED Search Near Position) and SIA services.

We are seeing greater use of our publishing interfaces both inside and outside of the NVO project. This reflects not only a greater willingness of the community to register resources, but more importantly effort building VO-compliant services. IPAC has completed and registered VO-compliant versions of all its commonly used NED services such as “search near object name” and “search redshift information”.

5 Data Models

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

Modifications were made to the Spectral (formerly Spectral Energy Distribution) data model to keep it consistent with the new spectral Data Access protocol. The Spectral data model was discussed at the Kyoto meeting and further minor changes proposed. One significant change is to provide metadata to indicate the level of calibration of both the axes and the fluxes. We distinguish between uncalibrated data (e.g. raw counts), relatively calibrated data (correct shape but unknown normalization, common for narrow slit optical spectra), and absolutely calibrated data (e.g. fluxed spectra). The approach taken generalizes to other kinds of observation, and will later be folded in to the more general Observation data model effort. The main work remaining on the Spectral model is to complete test implementations, and to review suggestions for changes to the FITS and VOTABLE serializations. The model will then be put through the IVOA standards process by the end of 2005.

Work has continued on the VOEvent protocol to support the burgeoning field of event-based astronomy. A small group led by R. Seaman (NOAO) and R. Williams (Caltech)

has been converging on a working draft of the IVOA standard, due for release in July 2005. The VOEvent schema has been designed for upward sophistication, meaning that it provides both a simple, pragmatic data representation, but also a richer, more sophisticated representation based on extension schemas.

A proposal has been submitted to NSF by a team at Caltech, Los Alamos, and Berkeley for the creation of VOEventNet, a rapid-response telescope grid that can integrate and reprioritize a collection of independent observations of a given astrophysical event.

5.2 Low-level (measurement, quantity, uncertainty, relationship)

Work on the Characterization schema continues, with the current emphasis on reconciling the Space-Time Coordinates schema of A. Rots with the schemata developed by the Strasbourg team. Face-to-face discussions at Kyoto resulted in some progress on this front.

5.3 Descriptors and ontologies (UCDs)

The IVOA has established the UCD science board and UCD tech board to manage the list of UCDs and the definition of UCD-related software. NVO members are serving on the boards and contributing to the revision of the UCD list. UCDs proposed by J. McDowell to support the spectral data model have been added to the list. The UCD-1+ definition document is expected to be approved by the IVOA during the next quarter.

The discussion of semantic dictionaries for astronomical object types (arising from the VOEvent discussion reported last quarter) remains controversial and has been deferred from the initial IVOA UCD document.

5.4 Space-Time and regions

The IVOA Proposed Recommendation for the Space-Time Coordinate metadata proposal drew significant comments from the community. The comments have been carefully considered and responded to. As a result, the PR and STC schemata are being revised.

5.5 Standard schema

Nothing to report this quarter.

6 Data Access Layer

6.1 Data access services (catalog, image, spectrum, time series, visibilities, ...)

SSA. A first draft of the Simple Spectral Access (SSA) query interface (D. Tody, M. Dolensky, et. al.) was released in early May, and was discussed in detail at the IVOA interoperability workshop in Kyoto. SSA provides a uniform interface for data discovery and retrieval of spectrophotometric tabular data including 1D spectra, time series, and SEDs, complementing the SIA interface already available for image data.

SSA is the first major upgrade to the VO data access services since SIA was released over two years ago. Much work has gone into development of the technology used in SSA, particularly of data models for generic dataset identification and characterization,

and representation of these data models in VOTable, FITS, native XML, and other data formats. An expanded query interface supports queries by spectral and time bandpass in addition to position on the sky, as well as by other data attributes such as resolution in the spatial, spectral, or time axes, the calibration status of the data, and so forth. The query response is based on the new standard dataset data models mapped into grouped fields in VOTable.

SSA is still undergoing revision in response to the broader discussions begun in Kyoto. The goal is to advance SSA to the Proposed Recommendation stage by the time of the next IVOA interoperability meeting this fall in Madrid.

Work continues on the **Simple Line Access Protocol**. A fairly detailed physical model for spectral lines was presented at Kyoto by M.L. Dubernet and P. Osuna. A working group with broad representation from within astronomy has been formed to help define a standard interface. This is expected to result in an initial draft of the interface sometime later this year.

SIA. A major upgrade of simple image access (SIA) is planned for early 2006. The most significant changes will be to the query interface. The revised query interface and query response metadata will be based largely on the work that has been done for SSA, with most of the generic dataset query parameters and metadata in common. Initial planning meetings were held in Kyoto and in connection with the EuroVO workshop in Garching at the end of June (D. Tody, F. Bonnarel, D. Durand). The goal is to have a first draft of the revised SIA interface ready for discussion in time for the Madrid meeting in October. When the new version of SIA is released an important issue that will need to be addressed is versioning. Many existing SIA services will continue to implement the old V1.0 interface even while new implementations come online. Client applications will need to be able to determine which version of the interface a given service supports, and may want to use only a certain version of a service. It will be necessary to identify in the registry which interfaces a given service supports. In many cases a given service implementation may simultaneously expose multiple interfaces, such as old and new versions of a particular interface. This approach will make it possible to evolve interfaces over time while allowing some interval of time (one or two years) for service providers and client applications to update to a new version.

Data Staging. We have begun to look again at the issue of asynchronous staging of data, coordinating with members of the Grid and Web Services working group (G. Rixon, R. Williams). Currently all data access is synchronous; data is generated on the fly at access time, and returned directly in response to a data access request. Only a single data file can be returned per request, and the request will time out if it takes too long to generate the data.

Asynchronous data staging will add a new *accessData* request that tells the service to start up a job to generate one or more datasets for later retrieval. Multiple datasets can be staged in one request, and messaging or polling can be used to monitor the status of the data staging job. Once generated, datasets can be stored by the service to later be picked up by the client with a standard synchronous *getData* request, or the service may be

directed to deliver the generated data directly to a specified location, e.g., a *VOStore* managed by the client or as part of a workflow. A mechanism at the same level will be used to support single sign-on authentication. Support for asynchronous data staging and authentication will be optional and will not affect the basic query and synchronous data access interface.

The current plan is to integrate these capabilities directly into the DAL services, basing the functionality on a standard job management mechanism such as the recently released Web Services Resource Framework (WSRF). The next steps will be to become more familiar with WSRF (a beta release of the V1.0 implementation from Apache was recently released) and then determine how to use this within SIA to asynchronously generate and stage images.

Query Language Integration. A joint session of VOQL and DAL in Kyoto (Y. Shirasaki, D. Tody) considered the issue of how to go about integrating query language capabilities into the data access services. The strategy adopted relies upon the observation that the data access services already separate the query from dataset retrieval. A query against a service is first issued to determine what available datasets satisfy certain search parameters, after which some number of individual (often virtual) datasets may be physically retrieved. Data access is based upon a family of data models describing the different types of astronomical data objects that can be accessed (image, spectrum, time series, source catalog, generic dataset, and so forth). Queries are posed in terms of the object attributes defined by the data model for the particular type of data object to be accessed.

To add query language (ADQL) support to the DAL services, all we need to do is add an alternative query interface. What we have currently is a parameter-based query interface, using a set of parameter=value constraints to refine the query. The ADQL-based query interface will instead use an SQL-like query language to define a query in terms of the data model attributes of the object to be accessed. Once the query has been posed to the service the remainder of the data access interface is the same. The query response is the same for both parameter-based queries and ADQL-based queries, consisting of a *VO-Table* containing standard dataset metadata for the class of object in question. Data staging (if used) and data access are identical operations regardless of the type of query used.

In many cases, data access will involve complex operations to compute a virtual data product (image cutout or projection, spectrum, SED, etc., conformant to a VO-defined object data model). Ultimately, the more sophisticated data access services will support both parameter and query language-based data discovery interfaces, and will employ a scalable computational framework on the back-end to generate virtual data products on the fly at access time or in response to a staging request.

6.2 Data representation (VO-Table, etc.)

SSA represents an effort at data representation based on a formally defined science data model (SDM). A data object, defined in terms of a formal SDM, possibly with data

provider-defined extensions, may be represented equivalently in any of a number of data formats, e.g., VOTable, FITS, a straight XML serialization, tabular text, XML with binary data attachments, and so forth. A first cut at this has been made in the current SSA specification, but further experience with actual data will be required to develop these techniques further.

Now that we are close to having a standard data model for spectral data, data providers are naturally interested in storing spectral data in archives in a form that is consistent enough with SSA to make runtime conversion at access time straightforward. R. Thompson, A. Micol, and D. Durand have begun an effort to do this for HST spectral data. A splinter meeting was held in connection with the Euro-VO workshop in late June to advance this effort. The goal is to define a FITS format (since this is archival data), based upon the FITS binary table format defined by SSA, to provide a more uniform way to store HST spectral data. This is a good test case as others will want to do the same later. An issue is how to provide a metadata extension mechanism for FITS, similar to what has already been defined for VOTable by F. Bonnarel and others.

6.3 Framework (mediators, components)

A second face-to-face meeting of the Option/VO working group on future data analysis environments was held at ESO in late June. Major topics for discussion were requirements, a reiteration of the system architecture, and the design of the component-container interface. An updated version of the generalized parameter model was prepared for the meeting and discussed.

The AstroGrid CEA implementation (as well as the proposed Universal Worker Service) was examined carefully for this meeting to see how much overlap there is with the computational framework we propose. There are some similarities as both can carry out complex computations, and both support a task/parameter mechanism to interface computational components. The chief difference is that CEA deals primarily with workflows linking together applications on a wide-area Grid, whereas the framework we propose is used to link together more finely-grained components in a single computer (workstation or cluster), and could be used to build the sort of applications which might participate in a larger-scale Grid workflow. In particular, a “local area” computational framework such as we are planning is needed to build services which generate complex virtual data products (images, spectra, etc.) for data access services.

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

CADC (D. Durand) and F. Bonnarel (CDS) implemented a number of SIA services to interface HST WFPC2 associations, CFHT mosaic data (both with full cutout support), and SCUBA data. This was done using SIA V1.0, but was in part a test case for SIA V1.1 as it used the new metadata extension mechanism.

L. Dobos (JHU) has implemented an early version of the SSA query interface for the JHU (Sloan) spectral archive. We plan to update this over the summer to support the final version of the SSA interface.

7 Query Language

7.1 Low-level: Astronomical Data Query Language

The main work this quarter was in updating ADQL 0.91 to prepare it for promotion to a Proposed Recommendation after the IVOA meeting in Kyoto. Issues raised at the previous Interop workshop were addressed. The document remains in review within the working group but should be ready very soon for general review in the IVOA community. A good deal of effort went into making ADQL 0.9 parsers in C# and Java. The new ADQL is available on dev.openskyquery.net.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

OpenSkyQuery is now in operational mode. Several bugs in the OpenSkyQuery portal were rectified while the nodes have not needed any changes.

U. Pittsburgh published OpenSkyNodes for the Phoenix Deep Survey (1.4 GHz) and the POSSUM KPNO UBVRI survey. A Java FullSkyNode implementation was checked in to the NVO CVS and has been documented for use in the Summer School. We have some XML related problems with this code when we use the latest versions of AXIS and a newer version of Tomcat. ESAC (Madrid) personnel have also produced a full OpenSkyNode for XMM in Java, but they have implemented a different cross-match algorithm in their node, leading to possible inconsistencies in results. Discussions are underway to either restrict OpenSkyNodes to one particular algorithm, or provide for more options (at the cost of greater complexity).

A proposal is being prepared for the NASA AISRP to extend OpenSkyQuery to allow full sky matches; this will be submitted in the next quarter.

7.3 High-level: Complex queries

Nothing to report this quarter. This activity remains a low priority.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)

A major new sky survey (Palomar-Quest) is being integrated with the NVO. A. Mahabal has made an image-access service (SIAP) for this 13 TByte synoptic survey, although it is only available to members of the survey team, since the data is still sequestered. PQ data is being processed on the NSF TeraGrid node at Caltech.

8.2 Grid Services (OGSA)

Progress is being made on the VOspace/VOSTore services that provide a secure, distributed data storage system for the NVO. The protocol has been underdevelopment by M. Graham, W. O'Mullane, and others from the UK AstroGrid project, and a formal description language (WSDL) developed. The Support Interfaces for web services have advanced significantly after discussions at the IVOA meeting in May, with contributions from M. Nieto, W. O'Mullane, and A. Thakar. A SkyNode interface has been created for the latest data release of the Sloan survey (SDSS-DR3). Spectrum services have been

mirrored at STScI by T. Budavari. M. Nieto and others have made a preliminary implementation of the Spectrum Access protocol (SSA).

Security infrastructure has been advanced by W. O'Mullane (JHU) working with M. Humphries and D. DelVeccio (U of Virginia). This is interoperating well between Java and .NET in generic form. This has also successfully interoperated with the CasJobs system, and continued with Caltech where interoperability of a VOStore prototype was shown for IVOA in Kyoto. A specification for VOStore was iterated between JHU, CACR, and AstroGrid, resulting in v0.18 for Kyoto Interop meeting. A prototype for Logging Support Interface was produced and the Support Interface Specification was revised for the Kyoto meeting. Work remains to be done producing the full WSDL for that specification.

R. Plante and R. Williamson, in collaboration with the NCSA Grid Security team and the GRIDS Center project, have been researching an architecture for managing authentication and authorization in the VO, and Plante has written two white papers on the topic describing requirements and use cases. Plante has also participated in IVOA GWS effort to develop the necessary standards.

At the core of the NCSA effort is a certificate-based architecture that can leverage existing technologies from the Grid community that interfaces well into a service-based architecture. With help from the GRIDS Center, they have put together a demonstration of a VO-friendly certificate management system that allows users to create and use certificates. This demonstration illustrates two important features need by the VO: first, it allows portals to manage certificates in a secure way on behalf of users, so that the users need not know that certificates are being used. Thus, usability does not sacrifice security. Second, it provides a common infrastructure (and just enough infrastructure) to provide interoperability across portals, and more importantly, across services administered by different organizations. To enable this interoperability, we recommend that the NVO maintain a small set of services for creating and managing identities for use in the VO.

8.3 Computational resource management

Nothing to report this quarter.

8.4 Virtual data

Nothing to report this quarter.

8.5 Application and service integration with Grid

Nothing to report this quarter.

9 Applications

Relatively little work was done on the development of new end-user applications during this time period. Software development efforts have focused on preparing user libraries and software for the second NVO summer school (see WBS 10.3). A set of all software libraries and client tools to be used at the summer school was collected and is being

placed under CVS control. During the third quarter of 2005 there will be extensive testing of these VO services.

9.1 Data location services

A few small bug fixes were made in the DataScope service and issues regarding the delivery of datasets into the JHU registry and subsequent appearance in DataScope were diagnosed.

9.2 Cross-correlation services

Efforts were made at the STScI, NCSA, and HEASARC to provide a Java-based (and hence portable) Full SkyNode implementation. The Java implementation remains due not so much to the VO elements, but to the very exacting compatibility requirements between the Tomcat Web Server and the AXIS Web-services library. Given this delicacy required here testing in alternative Java environment (e.g., BEA) will likely be needed to ensure a robust Java implementation.

9.3 Image combination, registration

A Java implementation of the SkyView image service is now available for download and distribution. This uses the SIA protocol to access many remote services and allows users to easily add new SIA servers by modifying only a few lines in a simple XML file.

9.4 Visualization tools and services

Nothing to report this quarter.

9.5 Theory

Nothing to report this quarter.

9.6 Statistical analysis

Nothing to report this quarter.

9.7 Datamining, outlier identification

Nothing to report this quarter.

9.8 Interfaces to/from legacy software systems

The NOAO has continues to integrate access to data resources using the VO data access layer. As part of the Summer School preparations the NOAO has looked at building VO based data analysis pipelines.

10 Community Engagement

10.1 Documentation

Documentation efforts in this quarter focused on the installation procedures and tutorials for the NVO Summer School II (to be held in September). Installation scripts were being prepared for testing in July.

10.2 *Web site*

In this quarter, the front page has been kept current with news announcements and upcoming events, such as AAS, ADASS, NVO team meetings, and IVOA meetings. A special section of the web site was created for the NVO Summer School, including an on-line application form.

10.3 *Technical training initiatives*

Preparations for the NVO summer school were a major focus of many members of the NVO team and some faculty recruited from outside. Major milestones in this quarter included:

- The finalization of the faculty,
- Development and distribution of the NVO Summer School Poster,
- Advertising of the Summer School in a variety of astronomy related publications and mailing lists and posting at the summer AAS meeting,
- Development and release of a summer school application web site,
- Receipt of 60 applications,
- Review of all applications by the a sub-committee and full faculty,
- Acceptance of 43 applicants (including 2 who have decided against coming),
- Development of the summer school curriculum and course descriptions and course assignments,
- Development of a science theme for the summer school,
- Establishment of software requirements for the summer school courses,
- Establishment of configuration control procedures and CVS repository for summer school software,
- Development of a summer school TWiki and permanent web sites,
- Development of a test plan for summer school software,
- Collection of summer school software into the repository for the initial pre-testing release.

During July and August extensive testing of the Summer School software packages and development of the summer school presentations is planned.

10.4 *Advocacy*

The NVO Project requested and was granted supplementary funding from NSF to create a program of small research initiatives for the astronomy community. These funds will enable ~15 research projects to get underway, or be continued from projects initiated at the Summer School, resulting in a first generation of VO-enabled science results.

Senior members of the NVO project team continue to give presentations in a variety of venues, informing the community about the Virtual Observatory.

11 Education and Public Outreach

11.1 Strategic partnerships

The NVO EPO program, directed by C. Christian, continues to foster collaborations with a variety of EPO professionals with established experience and programs reaching a diverse audience. The program spans news and public information, formal and informal science education and undergraduate education. Outreach to the science and technical communities is covered in WBS 10. The EPO personnel work with well-established projects to plan the types of technical support required to integrate NVO data into educational activities. The core NVO EPO projects as a collaborative are also strategizing on methods to broaden interactions with other existing educational programs especially focused on astronomical data sharing and large data archives.

11.2 Formal education

Development of pre-college education projects continues at JHU to be incorporated into the NVO EPO Portal for use by educators. The portal itself is being redesigned to better address target audiences and identify materials of interest to educators, as well as other specific audiences.

The module *Adopt an Object* as well as the NVO EPO site were tested by 20 Illinois elementary and middle school teachers at Yerkes Observatory. These tests are used to modify and improve the accessibility of resources for appropriate application in formal education.

Work on modules within Project LITE (PI K. Brecher, Boston U.) and CLEA (PI L. Marshall, Gettysburg College) continues. With these collaborators and a number of others from a variety of institutions, a proposal was crafted to create MyVOICE, an NVO EPO School, based on the NVO Summer School. In the school, educators would learn about NVO services and data sets and collaborate with NVO developers to augment existing materials and craft new materials for educators. MyVOICE was submitted to NSF but rejected without review owing to a minor budget discrepancy with the published guideline. The concept will be submitted to other opportunities as they arise.

11.3 Informal education

The Adler Planetarium in Chicago continues to design interactive exhibits within the museum hall, based on NVO data. Dr. M. SubbaRao of Adler participated in designing the MyVOICE proposal to bring informal science educators into the workshop environment to expose them to NVO services and data relevant to their displays.

The Caltech team is almost finished generating the large image for the new exhibition space of the Griffith Observatory in Los Angeles. This image will be 158 feet wide and 20 feet high, rendered at 80 pixels per inch. The data comes from the Palomar-Quest survey, and gains great depth by digitally combining up to 20 passes over the same sky.

11.4 Outreach and press activities

STScI and JHU staff hosted a booth at a local JHU Physics Fair in May. Outreach educators and young students were given demonstrations of NVO tools and the Education site.

An invited talk at the Rochester Institute of Technology was given at the Center for Imaging by C. Christian describing the power of education and outreach based on imagery as an engagement tool. NVO data access was highlighted. A presentation on the development of the Hubble Space Telescope (HST) press release image collection development was given by L. Frattare in Munich at a meeting for public access to astronomical data. This presentation mainly stressed the need for and premise of how to find the World Coordinate System (WCS) in press release astronomical images that do not have one associated.

For our NASA Applied Information Systems Research Program (AISRP) grant, STScI Office of Public Outreach is developing an image collection to be incorporated into the NVO. The image collection is a subset of the 16 years of press release images from the Hubble Space Telescope. A student intern is building the World Coordinate System database for selected images based on the completed code that allows WCS to be associated with each press image, based on knowledge of the original data acquisitions. The web service is in development by F. Summers built on an SIAP service. This work will be used to offer the image collection through the NVO and also through a public interface provided through our collaboration with the University of California Berkeley.

11.5 Technical development

NVO EPO personnel are collaborating on the development of metatags for public information through an international working group.

Activities by Organization

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

M. Graham has been instrumental in the development of an authenticated, secure VOStore, both at the international (standards) level, and also in implementing a Java version that has fully interoperated with other versions. The Carnivore Registry has advanced to a new release, and is faster due to better database indexing. Also on the Registry, Caltech has been working on curation and maintenance issues.

Progress on VOEvent has been strong following the international workshop at Caltech April 13/14. The specification is converging well for a release of a Working Draft in July, and prototype implementations are appearing that involve Caltech, Los Alamos, Berkeley, and the UK eStar project.

Caltech has been running the NVO web presence and mailing lists, as well as registry and other NVO services. All of this has been ported to new hardware, and the mailing lists to new software (mailman). Much of the Summer School administration has been at Caltech, including the application process and review.

Work at the Caltech Astronomy department has concentrated on building the Palomar-Quest data service in the NVO context, including an image access service, a SkyNode implementation for the source catalog, and VOStore for managing the data products. A major thrust has been the creation of the "Big Picture" for the Griffith Observatory in Los Angeles (see section 11.3).

Caltech–Infrared Processing and Analysis Center (IPAC)

J. Good has continued to lead the NVO Technical Working Group discussions. The focus this quarter has been on: Registry curation and logging; Defining, setting up, and populating an official NVO code repository (CVS); Defining VOspace infrastructure; Formal software testing and tracking; and preliminary discussions of asynchronous services. IPAC has also been testing the SkyNode interface using the gsoap package.

IPAC/NED has registered VO-compliant versions of 11 new NED functions. Also IPAC has focused much of its technical work this quarter on design and implementation of custom mosaicing services using NVO-compliant image sets and TeraGrid processing resources.

Canadian Astronomy Data Centre/Canadian Virtual Observatory

CADC has prepared a report on our experience and implementation strategy for a common SIA 1.0 framework in the context of a data center with multiple data collections. This report is available on the CVO web site.

CADC has released a prototype system for time-based querying of image data. The goal of this system is to allow users to find sets of overlapping observations in the same band with user-selected time sampling. Such observation sets could then be used for time-dependent analysis, such as searching for variable or moving objects.

CADC is currently developing a prototype wavelength-based querying system. This system allows users to specify two or more bands and find sets of overlapping observations with members in all bands. Public release of this system is expected in September 2005.

Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt)

WESIX (S. Krughoff, J. Brewer, A. Connolly, and C. Miller (NOAO)) extended WESIX to incorporate URLs as a mechanism for uploading data in order to simplify the overall interface of the service. The WCS component of WESIX has been updated to enable a general transpose of input images to account for non-standard images such as those from the SDSS. WESIX was deployed at the NOAO Science Archive (NSA) providing a quick look image analysis facility for images returned by queries on the NSA. The web service functions of WESIX are being incorporated into a pipeline for automated calibration (photometric and spectroscopic) of imaging data. They are currently being tested on SDSS coadded images.

Connolly, Krughoff, and Brewer have started developing web services for anomaly detection utilizing data returned by queries on openskyquery.net. Scranton has developed an all-sky pixelization scheme (SDSSPix) optimized for the SDSS survey strategy. The utility of this pixelization approach and its ability to define arbitrary survey footprints on the sky (and for querying these footprints) is under investigation.

Gardner, Connolly, and McBride have implemented a parallel *npt* code for use on the TeraGrid. The underpinnings of the code are a general tree code that can be deployed on Grids in excess of 128 nodes. The parallel *npt* point implementation is being used as a prototype for a general grid-based data mining facility.

Fermi National Accelerator Laboratory (FNAL)

Fermilab held an NVO training session for its experimental astrophysics group that was led by W. O'Mullane of JHU. About 15 people attended. Several successful exercises were carried out using to construct clients and servers using Java and the ANT and AXIS toolkits. Several bugs in the NVO supplied code were found and fixed. A full set of Java programs to interface with the OpenSkyQuery web service was written. A HEALPIX web service (useful for cross-matching areas of the sky) is now included in the Starlink STARJAVA package.

High Energy Astrophysics Science Archive Research Center (HEASARC)

During the last quarter the HEASARC personnel were deeply involved in the development of the NVO summer school, leading the summer school faculty in developing the curriculum and software requirements for the school. The HEASARC is also responsible for the Perl summer school software libraries. The HEASARC worked on the full-

SkyNode implementation, extending the Java code and making it easier to incorporate site-specific Java modules.

In related efforts, a standalone version of the *SkyView* service has been developed. This application fully incorporates the VO SIAP protocol for access to remote resources. Adding appropriate SIAP services are SkyView surveys can now be done in a few minutes essentially by specifying the SIAP URL and a few other lines in a simple XML file.

HEASARC staff attended NVO weekly technical working group telecons and led bi-weekly telecons for the NVO summer school.

Johns Hopkins University

T. Budavari worked on creating a mirror for Spectrum Services at STScI, so that MAST can extend its services to spectrum data. He also met with ESAC visitors (I. Barbarisi and J. Salgado).

W. O'Mullane delivered a successful short court course at Fermilab in April, using activities and experience from the NVO summer school. He worked with U. Virginia on web service security. He worked with M. Graham (Caltech) on VOStore prototypes, and with Graham and AstroGrid staff on specifications. He gave an invited talk at the University of Newcastle computer science department on NVO and SDSS data services (<http://www.cs.ncl.ac.uk/research/events/colloquia/abstract.php?id=137>). He attended NVO team meetings and IVOA Interop meetings, where he pushed ADQL and OpenSkyNode specifications up to version 0.91. He helped set up and staff the NVO booth for the JHU physics fair. He supported Summer School planning efforts, particularly in regard to notes and code from last year.

N. Li has continued development on OpenSkyQuery, upgrading it to ADQL 1.0 and fixing bugs.

M. Nieto-Santisteban worked on an AISR proposal to extend OpenSkyQuery to do all-sky cross matches. She also worked on ADQL 1.0 and SkyNodeInterface 1.0 documents, to be released as proposed recommendations to the IVOA community. With J. Gray (Microsoft Research), she met with the LSST database working group at SLAC, where she demonstrated SDSS and NVO services and applications. Formal collaboration with SLAC has been established as a result of this meeting.

T. Malik and Nieto-Santisteban continued their work to create a parallel query framework that uses partitioning and scheduling technologies on a cluster of nodes holding SQL Server.

A. Thakar worked on VO support interfaces, prototyping the logging interface and updating the specifications. With Nieto-Santisteban, he attended the AISRP workshop meeting at NASA AMES.

A. Szalay continued his work on the footprint and shape services.

J. Raddick worked with C. Christian (STScI) and Krista Wildt, a graphic designer at STScI, to produce a new design for the virtualobservatory.org web site. The site will become a portal to all projects that use NVO services to obtain data for use in education and outreach activities.

S. Carliles has produced a new FITS library written in C#, based on nom.tam.fits.

J. Vandenberg performed system administration and database maintenance tasks for servers used by the NVO, and answered a few OpenSkyQuery helpdesk items.

L. Dobos came as a visitor to update the spectral web services. The extended services will provide flux-preserving rebinning and dereddening of spectra using the WMAP dust map, and mapping VOTable spectra to C# objects. He went through many iterations of the services to keep them up to date with the emerging spectral data model and the SSAP protocols.

Microsoft Research

No activity this quarter.

National Optical Astronomy Observatories (NOAO)

NVO Project Scientist D. De Young continues to provide guidance to the NVO project in terms of setting project priorities from the perspective of scientific relevance to the astronomy and astrophysics community. During the period covered, De Young provided scientific perspective during the development of the curriculum for the second NVO Summer School to be held in September 2005. This included suggestions for possible science themes for the school as well as consultation during the development of the science theme finally selected. De Young is also developing a proposal solicitation process for NVO sponsored community research support grants to be made available to the community in the fall of 2005. De Young continues to work with the NVO Science Steering Committee in his role as chairperson; this committee will provide critical input to the project at the July 2005 Team Meeting and will in addition provide assistance in implementing the NVO research grant program.

De Young continues to serve on the Executive Committee of the International VO Alliance and as a member of the Theoretical Astrophysics Interest Group of the IVOA. In that context he attended the recent IVOA interoperability meeting in Kyoto, which included participation in both the Executive Committee discussions and meeting as well as participation in a special Theory Interest Group meeting. This latter meeting established some well defined next steps in developing both databases and metadata standards for theory datasets. Initiatives in these areas are to be undertaken by members from several different countries and cover datasets of differing sizes and types. The US effort in this area will concentrate on moderate to large-scale datasets in the areas of N-body simulations and multidimensional fluid dynamics and MHD calculations.

M. Fitzpatrick and O. Malik completed work on a reference implementation for the IRAF web-services framework. They are in the process of defining the first set of public services and documenting the system for release. The new system is now able to automatically generate test client applications from the same service descriptor file, allowing users to test the new service without needing to write custom clients. The framework will be featured at the NVO Summer School in 2005 Sept. Work was begun on enhancing the web-service adapter used in the system to allow greater flexibility in calling non-IRAF legacy applications. Fitzpatrick has been invited to present a talk on this work at STScI in late July.

M. Fitzpatrick has been preparing course materials for the NVO Summer School and configuring the nvo.noao.edu server to host user accounts for the participants requiring access to a Unix system. A network-based install/setup script for the needed IRAF tools was developed as part of the Summer School software distribution.

NOAO and LSST Corp. hosted the NVO team meeting in Tucson in late April, which was very well attended. Also in April, R. Seaman attended the first VOEvent Working Group meeting at Caltech. R. Seaman is the first author of the VOEvent Specification (V1.0), which was published at the end of June. M. Fitzpatrick also participated in the IVOA Interoperability meeting in Kyoto in May.

National Radio Astronomy Observatory (NRAO)

The NRAO VO plan prepared last quarter was circulated within NRAO and AUI by the Director's office in late April. This was generally well received; however there is no concrete plan yet for how to proceed or for what new resources might be available. There is considerable overlap between NRAO E2E and VO, hence a broader effort to advance VO at NRAO is closely linked to planning for E2E, particularly in connection with archive development.

Meanwhile work on integrating image data into the NRAO archive has begun. A simple image access (SIA) service for publishing atlas images from NVSS and FIRST to the VO has been implemented and was undergoing initial testing at the end of the quarter.

Work continues on the VLA archive imaging pilot project. Software is under development for staging data from the archive for pipeline runs. Production data processing has not yet begun.

The yearly week-long spring IVOA interoperability workshop was held this year in Kyoto, Japan. D. Tody attended, chairing the data access layer (DAL) working group. The main focus for DAL was the new simple spectral access (SSA) interface. This included definition of a second generation query interface for data access services, definition of query response metadata including data models for generic dataset identification and characterization, specification of the architecture required for future integration of astronomical data query language (ADQL) capabilities into the data access services, and planning for the next version of simple image access.

The second face-to-face meeting of the OPTICON/VO working group on future astronomical data analysis environments was held at ESO in late June (adjacent to the EuroVO workshop). D. Tody attended representing the US NVO. A design for a standard parameter mechanism, required to define a standard interoperable interface to astronomical data processing tasks, was presented and discussed.

Raytheon/ADC (University of Maryland and George Mason University)

K. Borne (GMU) continues to collaborate with the LSST (Large Synoptic Survey Telescope) Project Data Management team. Borne gave a presentation on EPO-related NVO activities to the LSST Data Management Team, including a review of the results of the July 2002 NVO EPO community meeting and a review of the SEGway team's "NVO-EPO Needs Assessment Study." Borne contributed inputs to the LSST Data Management Team in areas related to the use of NVO registries, the VOEvent protocol development activities, NVO science use cases, NVO EPO use cases, and NVO Summer School plans. Borne participated as co-PI on an NVO/EPO-related NSF proposal that was submitted by STScI/NVO staff to the NSF CI-TEAM research opportunity; the proposal was entitled "My Virtual Observatory Infrastructure for Cyber Education (MyVOICE)." Borne also gave a presentation entitled "Grid-Enabled Science with the NVO" to a Grid Computing Workshop in Landover, Maryland, sponsored by the NASA Goddard Space Flight Center. This presentation surveyed NVO science results utilizing grid computing, plus NVO activities in the areas of distributed computing, VOspace, VOStore, workflows, service composition, single sign-on, authentication, registries, and data grids (distributed data discovery, access, mining, and analysis). The talk slides are available at <http://romulus.gsfc.nasa.gov/msst/gridws/>. Borne also contributed to the planning of a paper to be presented at the Fall 2005 AGU Meeting, entitled "Developing Architectural Alternatives and Best Practices in Cooperating Registry/Repositories for Application to Space Science." Finally, Borne is second author on a refereed paper that has been accepted for publication in the Data Science Journal (www.datasciencejournal.org). The paper is entitled "eScience and Archiving for Space Science," and it describes data-intensive e-Science activities that are empowered by the NVO, including data mining, distributed data access, and the integration of data, modeling, and simulations. The paper is loosely based on a talk given by T. Eastman (GSFC) at the 2004 CODATA Conference (<http://www.codata.org/04conf/papers/>).

E. Shaya (U. Maryland) is experimenting with the Catalog data model using real examples in the ADC catalog repository. Shaya is also developing in-depth semantics for future versions of UCD, SED, VOEVENT, and SIAP.

San Diego Supercomputer Center

The main activities at SDSC continue to be integration of NVO processing with the NSF TeraGrid. Towards this effort, SDSC is again replicating major sky surveys onto TeraGrid disk resources. The current status is:

- DPOSS, housed on tape, but will be replicated onto the TeraGrid disk cache
- USNO-B, still waiting shipment of the collections
- SDSS, starting download of the SDSS DR4 release
- 2MASS, reloading the final release in collaboration with R. Williams and J. Good

SDSC is collaborating with Lawrence Livermore National Laboratory and Mike Norman of UCSD. The project will generate synthetic views of the early universe to test the whether the proposed processing pipeline for the LSST will be able to detect the predicted features. One objective is to build upon the NVO services to create the processing pipeline and data management environment. The processing of the 2MASS mosaics is being used as an early evaluation of the required processing power. To understand the ability to the current TeraGrid to process 10 TBs of data in 3 days, the following evaluation of the mosaic service is planned:

- Complete the replication of the 2MASS data collection onto the TeraGrid GPFS disk system. This will be started July 19, 2005 after a planned reconfiguration of the GPFS file system is completed.
- Schedule 6 dedicated runs on the TeraGrid node at SDSC, of 12,000 CPU hours each. SDSC currently schedules such dedicated processing twice a week, implying that a total of 3 weeks will be needed to process the entire collection
- Review the results of each dedicated run before starting the next computation. We will rely upon the expertise of J. Good to evaluate the scientific quality of the results.
- Publish the mosaics into the Hyperatlas in collaboration with R. Williams.

The goal is to show that given stable algorithms, the entire process could be completed in three days. Meeting the requirements of the LSST project will require roughly three to five times the computing capacity of the current TeraGrid node at SDSC.

Smithsonian Astrophysical Observatory

G. Fabbiano, besides coordinating the CfA NVO efforts, continues to be engaged in scientific outreach activities, including contributing to the discussion towards the planning of the VO IAU symposium in Prague. She is also the principal organizer of a workshop at the COSPAR symposium to take place in Beijing in July 2006, dedicated to the challenges of high-resolution space astrophysics. This workshop will consist of three sessions, one of which will be entirely dedicated to data challenges with strong VO flavor. G. Fabbiano attended the Kyoto Interop meeting.

J. McDowell continues leading the Data Models group, and has been working on the Spectral and Characterization data models, as well as contributing to the UCD effort and to the DAL simple spectral access discussions. He attended the Tucson team meeting and the Kyoto interoperability meeting, where he presented work on the Spectral data model, UCDs for the spectral model, and reported on an experiment in porting theory simulations to FITS with added VO-related metadata. He has also begun discussions with M-L Dubernet (Paris) on common formats and a data model for spectral line atomic data.

K. McCusker has been developing an improved interface to the Chandra SIAP service. She has installed a Carnivore registry at the CfA and provided beta testing and feedback to M. Graham on this software. She is continuing work on implementation of Java library software for the Spectral data model.

A. Rots has responded to the comments resulting from the RFC on the PR for the Space-Time Coordinates; as a result he has been revising the Space-Time Coordinates document and schema in the light of feedback received from the community. He has also been involved in the core group developing the PR for VOEvent. Rots attended the VOEvent workshop in Pasadena, the NVO team meeting in Tucson, and the VO Interoperability meeting in Kyoto. At the Kyoto meeting he presented STC issues in the context of Data Model, VOEvent, and VOTable.

Space Telescope Science Institute

R. Hanisch worked on updates to the Resource Metadata definitions document, taking into account experiences with the resource registry curation from the past several months. The definitions of a number of metadata elements have been clarified. These revisions were brought forward for discussion at the International VO Alliance interoperability workshop in Kyoto, Japan in May, and were received positively.

Hanisch wrote an overview of NVO activities for the newsletter of the European Astronomical Society and gave an invited talk on VO standards at the Joint Conference on Digital Libraries in Denver. Hanisch also attended the EuroVO Workshop in Garching, Germany. This workshop was similar to the NVO Summer School, though focused more on data providers than on end-users.

G. Greene updated Registry auto-harvesting services to correct errors in the harvesting times. The standardization of UTC must be incorporated in all publishing registries to accommodate time differences between the various sites. The complete Vizier VO publishing registry was harvested and descriptions were corrected.

A redesign of the NVO OPO web site was undertaken in collaboration with JHU. The goal is to provide a more efficient and attractive web portal for NVO OPO to better meet the functional needs of the Outreach community.

C. Christian led the development of a proposal to an NSF Cyber-Infrastructure program that would provide funding for training educators in the use of VO tools and technology. The proposal was submitted on 5/27/05. Unfortunately, a misunderstanding over budget limits led the NSF program executive to reject the proposal without review. This situation has been appealed.

United States Naval Observatory

Work has continued on replicating the USNO Precision Measuring Machine (PMM) project archive of digitized Schmidt plates. The 12,000 plates that were scanned comprise almost 10 TByte of image data which are being transferred from tape onto disk. During this quarter, all of the remaining Schmidt plate data was copied onto disk. A modest amount of processing is being done on the raw images, to ensure basic astrometric and photometric calibration of the images (S. Levine). They will then be integrated into the TeraGrid storage at SDSC (R. Moore).

Catalogue search services at USNO have been upgraded to improve batch processing of data requests. Work by S. Levine is ongoing to upgrade the current web based image cutout service to be compliant with the VO SIA protocol. Part of these upgrades includes copying the archive image data from tape onto hard disk, and working with SDSC to replicate the data on the TeraGrid.

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

R. Plante and R. Williamson continue to maintain the NCSA publishing registry. Williamson added preliminary support for Simple Spectral Access (SSA) services; full support will be completed as the SSA standard becomes complete.

Plante remains active in the IVOA Registry Working Group (see section 4). He is lead author on the VOResource specification document and a major contributor to the Registry Interface specification document. He is leading a modeling tiger team to look at VOResource extensions for data collections and services.

Plante and Williamson have been putting in a major effort to develop a security framework for the VO (see section 5). Plante has engaged a collaboration with the NMI Grids Center project and the NCSA Security team to bring existing grid technologies to address security needs in a user-friendly way. Williamson has been working with the GRIDS Center team to assemble a login prototype for the NVO. Plante is also engaged in the larger Security standards development under the IVOA Grid and Web Services Working Group.

University of Southern California (USC/ISI)

The work at USC/ISI has focused on providing infrastructural support for making Montage an operational service and on workflow optimizations. ISI participated in discussions with other NVO members from Caltech, JPL, and SDSC on listing the requirements for making the Montage mosaic service operational on the TeraGrid. Among the main requirements were

- i) A need for a uniform GridFTP access to the image datasets irrespective of the underlying data storage (SRB etc).
- ii) A need for a TeraGrid allocation for the required storage and the CPU cycles.
- iii) A need for naming and storage strategy for the generated image mosaics.

Towards this goal, ISI installed the Montage software on the TeraGrid community software area on the Caltech, NCSA, and the SDSC sites. Deployment of the other software pieces for making the service operational is expected to begin next quarter.

ISI also worked on developing runtime optimizations that can drastically reduce the workflow completion time for the Montage workflow. The optimizations can be applied in general to fine granularity workflows that are based on the bulk synchronous parallel model of computation with a large number of similar tasks between the synchronization

points. A report on the work was presented at the NVO Team meeting at Tucson in April 2005.

University of Wisconsin

No activity reported this quarter.

Publications

R. Plante & T. Linde, *VOResource: an XML Encoding Schema for Resource Metadata*,
<http://www.ivoa.net/internal/IVOA/VOResourceV10/WD-VOResource-v1.html>

R. Plante & B. Loftis, *A VO-friendly, Community-based Authorization Framework Part 1: Use Cases, Requirements, and Approach*,
<http://us-vo.org/pubs/files/CommunityAuthorizationP1.pdf>

R. Plante & B. Loftis, *A VO-friendly, Community-based Authorization Framework Part 2: Use Cases, Requirements, and Approach*,
<http://us-vo.org/pubs/files/CommunityAuthorizationP1.pdf>

R. Plante, *Interoperable Authentication and Authorization in the VO*,
<http://us-vo.org/pubs/files/SecurityTMApr2005.ppt>

Virtual Observatory Articles in the Popular and Technical Press

Nothing this quarter.

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
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