

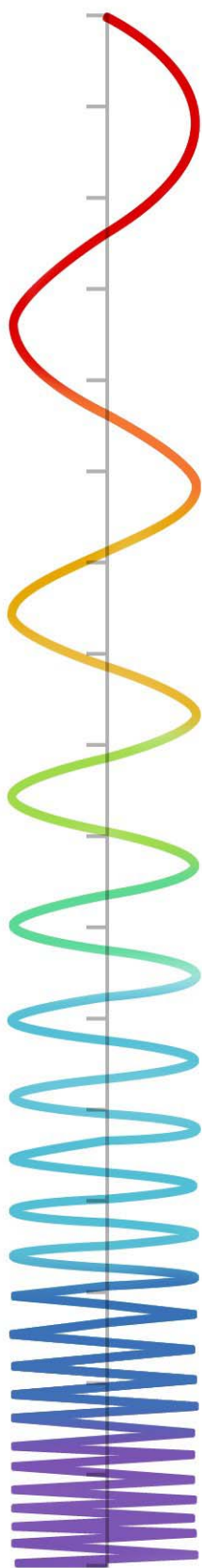
Quarterly Report  
January-March 2005

Building the Framework for the  
National Virtual Observatory

NSF Cooperative Agreement  
AST0122449



INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE



<b>Executive Summary</b> .....	<b>1</b>
<b>1 Management</b> .....	<b>3</b>
1.1 General (planning, reporting, communications, team meetings, etc.) .....	3
1.2 Science .....	3
1.3 Technical (including standards, configuration management) .....	3
1.4 Financial.....	3
1.5 International coordination/collaboration.....	3
<b>2 Science Requirements</b> .....	<b>4</b>
2.1 Usage scenarios for all areas of astronomy research, including theoretical simulations .....	4
2.2 Requirements analysis .....	4
2.3 Demonstration definition and review.....	4
<b>3 System Architecture</b> .....	<b>5</b>
3.1 System design, components; relationships to Grid components.....	5
3.2 Computational facilities (processing, bulk data storage, network access, security, authentication).....	6
3.3 Digital library integration .....	7
<b>4 Registries</b> .....	<b>7</b>
4.1 Resource metadata .....	7
4.2 Resource metadata schema .....	7
4.3 Publishing and harvesting protocols .....	8
4.4 Search protocols.....	8
4.5 Replication, synchronization, maintenance, revision control, and curation .....	9
<b>5 Data Models</b> .....	<b>10</b>
5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality) .....	10
5.2 Low-level (measurement, quantity, uncertainty, relationship) .....	11
5.3 Descriptors and ontologies (UCDs) .....	11
5.4 Space-Time and regions.....	11
5.5 Standard schema .....	11
<b>6 Data Access Layer</b> .....	<b>12</b>
6.1 Data access services (catalog, image, spectrum, time series, visibilities, ...) ..	12
6.2 Data representation (VOTable, etc.) .....	12
6.3 Framework (mediators, components) .....	13
6.4 Data provider/consumer implementations and end-to-end testing .....	13
<b>7 Query Language</b> .....	<b>13</b>
7.1 Low-level: Astronomical Data Query Language.....	13
7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode .....	14
7.3 High-level: Complex queries .....	14
<b>8 Web and Grid Services</b> .....	<b>14</b>
8.1 Web Services (SOAP, WSDL, etc.).....	14
8.2 Grid Services (OGSA) .....	14
8.3 Computational resource management.....	15
8.4 Virtual data.....	15
8.5 Application and service integration with Grid.....	15

<b>9</b>	<b>Applications .....</b>	<b>15</b>
9.1	Data location services .....	15
9.2	Cross-correlation services .....	15
9.3	Image combination, registration .....	16
9.4	Visualization tools and services .....	16
9.5	Theory .....	16
9.6	Statistical analysis .....	16
9.7	Datamining, outlier identification .....	16
9.8	Interfaces to/from legacy software systems .....	16
<b>10</b>	<b>Community Engagement.....</b>	<b>16</b>
10.1	Documentation .....	16
10.2	Web site .....	17
10.3	Technical training initiatives.....	17
10.4	Advocacy .....	17
<b>11</b>	<b>Education and Public Outreach .....</b>	<b>17</b>
11.1	Strategic partnerships.....	17
11.2	Formal education .....	18
11.3	Informal education .....	18
11.4	Outreach and press activities .....	18
11.5	Technical development .....	19
	<b>Activities by Organization.....</b>	<b>20</b>
	Caltech–Astronomy Department and Center for Advanced Computational Research (CACR) .....	20
	Canadian Astronomy Data Centre/Canadian Virtual Observatory .....	21
	Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt) .....	21
	Fermi National Accelerator Laboratory (FNAL).....	22
	High Energy Astrophysics Science Archive Research Center (HEASARC) .....	22
	Johns Hopkins University .....	22
	Microsoft Research .....	24
	National Optical Astronomy Observatories (NOAO).....	24
	National Radio Astronomy Observatory (NRAO) .....	24
	Raytheon/ADC (University of Maryland and George Mason University).....	25
	San Diego Supercomputer Center.....	25
	Smithsonian Astrophysical Observatory.....	27
	Space Telescope Science Institute .....	27
	United States Naval Observatory .....	28
	University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA).....	29
	University of Pennsylvania .....	29
	University of Southern California (USC/ISI) .....	29
	University of Wisconsin .....	30
	<b>Publications .....</b>	<b>32</b>
	<b>Virtual Observatory Articles in the Popular and Technical Press.....</b>	<b>33</b>
	<b>Acronyms .....</b>	<b>34</b>

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

*Period covered by this report:* 1 January—31 March 2005  
*Submitted by:* Dr. Robert Hanisch (STScI), Project Manager

**Executive Summary**

Significant progress has been made on the VOStore concept, by which users of the virtual observatory can store data files and tables in a distributed and secure manner. Results from remote services will have their results stored in VOStore, with location transparency, and the results can be shared with collaborators. VOStore motivates the implementation of authentication and authorization capabilities, and various approaches are being evaluated.

Substantial effort was spent this quarter on evaluating the quality and completeness of registry entries and providing feedback to resource and service publishers about how to improve their resource descriptions. Common errors and omissions in resource metadata are being addressed through modifications to the Resource Metadata definition document. Issues concerning compliance with the Open Archives Initiative (OAI) metadata harvesting protocols have been worked out, and collaboration has continued on defining the standard interface specification for registry queries.

Work has also progressed on the spectral data model and the associated Simple Spectrum Access Protocol (SSAP), with both to be presented and discussed at the spring IVOA Interoperability workshop in Kyoto. A number of prototype implementations of spectral data access services and spectral data display and analysis tools have been deployed, and these have helped to inform and refine the data model and SSAP work.

An IVOA VOEvent workshop was held at Caltech and concluded with an agreement by an international team on an information infrastructure to support the burgeoning field of event-based astronomy. The objective of the VOEvent working group is to build an open standard for exchanging messages about immediate astronomical events, including publication, archiving, query, subscription, and aggregation. The VOEvent standard has been agreed in rough form at the workshop, and has buy-in from projects including GCN, LSST, Pan-STARRS, Palomar-Quest, LIGO, eStars, Raptor, Pairitel, ATEL, and Hands-On Universe.

The Space-Time Coordinates (STC) metadata specification and associated XML schema were promoted to Proposed Recommendation status. VOEvent discussions have helped to focus the development team on the importance of the STC and have led to comments during the open Request-for-Comments period.

Planning for the second NVO Summer School has gotten underway. A site has been secured (Aspen Meadows Resort, Aspen, Colorado), dates have been set (6-15 September 2005), the faculty has been recruited, and announcements have been made. A detailed curriculum is under development. Full details are available on the NVO Summer School website at <http://www.us-vo.org/summer-school/2005/>.

## 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.

Following the January 2005 meeting of the NVO Advisory Committee, the Executive Committee updated the project plan and issued a revision that defines work areas and sets priorities for the remainder of the NVO development project. The plan is available at <http://us-vo.org/pubs/files/Project%20Plan%20Update%20Feb%202005.pdf>.

#### *1.2 Science*

Science priorities for 2005 were documented and circulated to the Science Steering Committee for comment. E-mail responses have been limited, however, so a telecon meeting is planned for spring. We will most likely have a face-to-face meeting with the SSC in conjunction with the summer NVO team meeting, as we did last year.

#### *1.3 Technical (including standards, configuration management)*

The CVS software repository has been established at Caltech, with 14 authorized contributors and numerous downloads. Authorization is obtained through the existing account creation channels provided by Caltech CACR: by filling in a web form and signing a responsibility agreement. Software may be either “contributed” or “managed”; in the latter case it must be structured and documented in a uniform way. There is complete documentation on using CVS, including the ssh key infrastructure.

#### *1.4 Financial*

Overall project spending continues as planned. Our current carry-forward is less than \$950k. 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. Some reallocations of funds among subawards were made this year, accounting for an increase in the scope of work at SAO as well as the relocation of U Penn collaborators to SAO.

#### *1.5 International coordination/collaboration*

R. Hanisch represented the NVO Project at the final meeting of the Astrophysical Virtual Observatory (European VO) Science Working Group in Villafranca, Spain, January 25-26. The AVO project is now concluded as the Euro-VO begins. AVO showed two science cases, one aimed at detecting stars as they evolve from the asymptotic giant branch to planetary nebulae and another focused on studying galaxy evolution through comparison with models of spectral energy distributions.

NVO Executive Committee members continued their involvement in the IVOA Executive Committee meetings and telecons. An IVOA Executive meeting was held at San Diego Supercomputer Center in January. R. Hanisch has been working with N. Walton (AstroGrid) and F. Genova (EuroVO, CDS) in planning the May 2005 IVOA Interoperability Workshop in Kyoto, Japan.

## 2 Science Requirements

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

The US NVO is continuing its efforts to bring to the US astronomical community a series of useful tools and applications that will facilitate astronomical research on a daily basis. The first release of a suite of tools and applications was made at the January 2005 AAS Meeting; the development of these applications was described in the last quarterly report. This activity continues as ideas for more complex and more sophisticated tools and applications are considered for development. A new list of possible projects has been produced by the NVO team in collaboration with the Science Steering Committee (SSC). This list includes:

- Large Scale Cross-correlations
- Large Scale Image Mosaicing and Stacking
- Footprint/sky Coverage Intersection Services
- Variability/Change Detection
- Spectral Index Mapping
- Incorporating VO Capabilities into “Legacy” Datasets

This list is currently being reviewed by the SSC, and the result of this review will be a prioritized list of projects. These priorities will be assigned exclusively on the basis of scientific merit. After the SSC review, the Team will evaluate the prioritized list for technical feasibility and level of effort/time to completion analysis. Based upon all of these reviews, the project will then decide which of the suggested new applications will be developed and in what order they will be completed.

### *2.2 Requirements analysis*

Analysis and review of the proposed new applications and tools is described in Section 2.1 above. In addition, the development of these new tasks will be continually reviewed as work upon them progresses.

### *2.3 Demonstration definition and review*

Once the new applications and tools to be developed have been put into final form, their efficacy and ease of use will be tested through the construction of demonstrations that display the capability of each new tool. These demonstrations will then be made available to the US astronomy community.

### 3 System Architecture

The services that implement the NVO architecture are being extended to support image retrieval and image survey archive federation. The image retrieval services will then be used to support large-scale data manipulation on the NSF TeraGrid under an NVO time allocation. The use of TeraGrid resources by the NVO project has been formalized, with disk storage resources allocated to support NVO image collections.

#### *3.1 System design, components; relationships to Grid components*

A major effort has started on the design of a standard data access mechanism, called VOSTore, and a standard image archive federation mechanism, called VOSpace. The idea is that users of a VO will have the ability to store and share files and tables that are to be processed or delivered by services. Data will be stored in any convenient place and protocol and users will not need to know these details. VOSTore will offer a sophisticated security model that allows each data object to have individual and group permissions. The design effort has involved multiple members of the NVO team, and is being submitted to the IVOA as a new standard service. W. O'Mullane (JHU) and R. Williams (Caltech) created an initial document defining the VOSTore interface. A. Szalay (JHU) developed a specification of the features that would be supported within the data access service versus the data federation service. R. Williams (Caltech) then led a discussion of the proposed interface definitions and the technologies that could be used to implement the new services.

As has been done for other NVO services, multiple implementations are being sought to demonstrate the generality of the VOSTore data access service. A first implementation, based on WebDAV, was created by M. Graham (Caltech). The test VOSTore interface can be accessed at <http://nvo.caltech.edu:8080/vostore>. The VOSTore web service interface is at <http://nvo.caltech.edu:8080/vostore/services/VOSToreSoap>. The WebDAV implementation of the VOSTore interface resides at <http://nvo.caltech.edu:8080/vostore/webdav>.

The VOSTore web service currently supports: put, get, list, listAll, rename, delete. “put” and “get” will return a VOSToreResponse containing the URL to the WebDAV where you upload/download your file. Upload/download with DIME attachments is being implemented.

The web service is secure and requires the use of digitally signed SOAP messages. These need to be signed with a certificate that has been signed by the VOSTore CA (certificate authority). The WebDAV interface is currently insecure but when it becomes secure, the same certificate may be used in calls to it (see `managed/vostore/vostorejava/samples/net/ivoa/vospace/webdav/WebDAVClient.java` for some idea of what this will look like). When the WebDAV interface is secure, you will still be able to use your favorite web browser to explore it by loading the pkcs12 certificate.

The next steps in the demonstration of the VOSTore interface will be a WSDL implementation based on the Storage Resource Broker and an interface to the SDSS image archive.

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

**Storage resources and collections.** SDSC is making a major commitment to the NVO project through the provision of disk space to house the major NVO sky image collections on the TeraGrid. A total of 44 Terabytes of disk space is being allocated for use by the NVO project. The disk space is on a GPFS file system that will be accessible from all nodes of the TeraGrid. The disk is being installed and tested for NVO use in May 2005. Towards this goal, the final version of the 2MASS image collection has been replicated onto storage systems at SDSC. This required the loading of the 2MASS images onto a 1-Terabyte disk (R. Williams), transfer of the data onto CACR disk at Caltech, and then transfer of the 10-Terabytes to SDSC over the TeraGrid network (G. Kremenek, SDSC).

Work has begun on replicating the USNO Precision Measuring Machine (PMM) project archive of digitized Schmidt plates. The 12,000 plates that were scanned comprise almost 10 TB of image data that are being transferred from tape onto disk. These are the images from which the USNO-A and USNO-B catalogues were constructed. A modest amount of processing is being done on the raw images, to ensure basic astrometric and photometric calibration of the images (S. Levine). A 1-TB disk has been sent to USNO to support the replication of the PMM/USNO-B image collection onto the TeraGrid. A little over a terabyte has already been read off the source data tapes onto spinning disk. After processing the WCS information and creation of image headers, the data will be moved to SDSC by physical transport of the 1-TB disk.

The NVO image collections will be managed on the TeraGrid with the Storage Resource Broker (SRB) data grid. Two new releases of the SRB were provided in March 2005 to provide enhanced support for bulk operations. The NVO image collections are being migrated onto the new SRB version (G. Kremenek).

**TeraGrid data intensive applications.** A second major effort has been the porting of the Montage image mosaicing software onto the TeraGrid and the integration with SRB managed image collections. J. Good provided upgrades to the Montage service, which was installed at SDSC by L. Brieger. The mosaics that are being generated will conform to Hyperatlas pages in the project led by R. Williams. The initial test of the system produced a 6-degree by 6-degree mosaic that can be seen at <http://nvo.sdsc.edu/NVO/2mass.html>.

The Montage software is set up to run on the SDSC and NCSA IA-64 TeraGrid nodes and the SDSC IBM DataStar system. The Mosaic service will be able to execute simultaneously across multiple TeraGrid nodes as one of the few TeraGrid distributed applications. Two resource allocations were requested and received to enable the analysis: a TeraGrid roaming allocation for use of multiple TeraGrid nodes, and an SDSC Data Allocation Committee grant to enable J. Good to test his code on the DataStar system.

Several observations about the Hyperatlas project are worth making. First, this data intensive effort requires the coordinated efforts of multiple persons (J. Good – algorithm

developer, R. Williams – Hyperatlas design, G. Kremenek – collection management, L. Brieger – code porting). Second, the project will be a major demonstration of simultaneous use of TeraGrid nodes, while accessing 10-Terabyte collections over the TeraGrid network. Third, the application will be extensible to other image surveys, including DPOSS, SDSS, and USNO-B. Fourth, the results will form a Hyperatlas that can then be used to support multi-wavelength image composition, and will serve as a major resource developed by the NVO project.

### *3.3 Digital library integration*

NVO team members A. Szalay, R. Hanisch, and M. Kurtz partnered with S. Choudhury et al. (JHU Library) on a proposal to NSF's Digital Archiving (DIGARCH) program aimed at establishing mechanisms for the long-term preservation and curation of the large datasets associated with peer-reviewed publications. The proposal was not successful, but explorations continue with potential partners (American Astronomical Society, University of Chicago Press, UCSD-Chronopolis) to see what pilot projects might be carried out. A component of this work aims to construct a database of astronomical instrumentation characteristics. Such a database could support throughput calculations and the computation of virtual observations. Initial work in this area is also underway at EuroVO/ESO, and we will explore possible collaboration.

## **4 Registries**

In the last quarter, the effort in this area has been focused on two areas: curation of registry records (4.5) and revising the IVOA draft standard Registry Interface document.

### *4.1 Resource metadata*

The study of metadata quality (4.5) has generated new insight on how publishers are interpreting metadata differently and has pointed to ways that we can improve our documentation. Based on these results, the registry curation team generated some specific guidelines for interpreting particular metadata. For example, many publishers interpreted the Creator tag as the person who built and deployed the service that accesses a particular collection, whereas the proper meaning of Creator is the person or group that created the scientific content (i.e., the data collection) that the service provides access to. The team, then, developed specific guidelines for choosing a Creator value depending on the type of resource (e.g. Service, Data Collection, Organization, etc.).

R. Hanisch has taken some of these results and folded them into a revision of the IVOA Specification, *Resource Metadata for the Virtual Observatory*, tightening up some definitions, improving examples, and removing ill-used or rarely used concepts. This document will be presented at the IVOA Interoperability meeting in May.

### *4.2 Resource metadata schema*

In this last quarter, R. Plante (NCSA) has been working on developing the necessary documentation needed to move the core metadata schema, VOResource, to a version 1.0 IVOA Proposed Recommendation. This includes a prose document describing the VOResource schema, its use and mechanisms for extension. Documentation also includes a separate developer's guide, to be published as an IVOA Note.

Through the IVOA Registry Working Group, Plante is also looking at small revisions in the schema that will allow it to operate flexibly with the Registry Interface specification as well as aid with registry curation. These include proposals to add attributes called “harvestedFrom” (4.3) and “verificationLevel” (4.5). These raise an important issue with regard to registry curation: these attributes are intended to be updated by any registry that holds the record, unlike the rest of the metadata, which should only be changed by the original publisher. This has led to a proposal to pull registry-edited information out of the VOResource schema make it specifically part of the schema supporting the Registry Interface; the purpose is to make clear what information is homogeneous across the VO and which might vary depending on where the record is retrieved from.

Finally, as part of the work to complete the IVOA Registry Interface specification, Plante has been working on a revision of the Registry resource metadata, an extension of the VOResource core metadata.

#### *4.3 Publishing and harvesting protocols*

We have made effective use of the Open Archive Initiative (OAI) Protocol for Metadata Harvesting for about two years now. An important issue that we need to address to complete our framework for harvesting is how new publishing registries enter into the IVOA community. Another way to phrase this problem is how registries that wish to build a full repository of everything known to the IVOA know whom to harvest from. Initially, the IVOA Registry Working Group envisioned a de-centralized mechanism in which registries would discover new publishers from any of their peers; though, less had been worked out how a publisher first taps into the network. In the interest of simplifying this “joining” mechanism, R. Hanisch recommended that IVOA sponsor a small, simple “registry of registries” that features a web form for registering a new registry. This simple solution can be put into place within the next quarter.

The other important harvesting issue comes from the larger issue of curation. We have been looking at whether our current metadata is sufficient to effectively trace provenance of registry resource records, particularly tracing errors and quality issues (4.5). To that end, we are looking at a proposal to add a “harvestedFrom” attribute to the VOResource schema. Our current registries in the NVO support this attribute internally; however, by making this part of the standard metadata, this information would be exposed to other registries and applications.

#### *4.4 Search protocols*

In this last quarter, Plante has been working on revisions to the Registry Interface specification in preparation for the May Interoperability meeting. A major part of this effort has been driving discussions within the IVOA Registry Working Group to resolve remaining technical issues. One important issue we are looking at is our decision to base advanced registry queries on a combination of ADQL (an emerging IVOA standard for database queries) and XPath (a W3C XML standard). We have had some experience now implementing support for ADQL & XPath, so we have a better practical sense of the advantages and disadvantages of using an XML encoded language. In addition, we are trying

to understand the ramifications of having one IVOA standard depend on another, which obviously has implications beyond just registries.

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

In this quarter, we have stepped up our efforts to address registry curation issues. We assembled a tiger team including S. Emery Bunn (Caltech), R. Hanisch (STScI), and R. Plante (NCSA) to systematically review all resource descriptions for consistency and effectiveness for practical scientific use. As seen in every other community that has attempted to collect metadata about resources, we see great variation in the amount of information provided as well as the quality and accuracy of that information. Emery Bunn, as the person with the least connection to the resources being described and the standard used to describe them, led the systematic review. Working first with resources associated with major partners in the NVO project (namely, NCSA, IPAC, and STScI), she examined the values of all core resource metadata as well as extensions specifically for Conesearch and Simple Image Access services, looked for problems, and rated the overall quality. The team then followed up on problem areas and attempted to derive by induction some general guidelines for providing high quality values for specific metadata (see also 4.1). We are using these results to provide feedback to publishers for improving the values for current records. We expect this to become general practice for registry administrators.

Once we have a good set of metadata guidelines in hand, we plan to establish a practice of rating the quality of the resource descriptions and encode it into the records via a proposed “verificationLevel” attribute. Partly based on software verification and partly on human inspection, this system would allow each registry to “stamp” a record according to their own standards of quality. Applications, then, can use this attribute to select resources they will use based on the level of quality they need. Note that this rating does not cover the scientific content of the resource, but only its description and compliance with standard interfaces. An important part of making the verificationLevel effective is the practice of providing feedback from the registry to the publishers to improve the metadata quality.

The discussion of how new registries join the IVOA web (4.3) is important to the discussion registry replication and synchronization. Also related to this is the way we harvest to ensure synchronized resource records. To avoid receiving duplicate copies of resource records from different registries, our publishing registries have only been emitting records that originate from that registry (rather than those that were acquired by harvesting another registry). As part of the Registry Interface specification, we now allow one to choose whether to harvest all records or just those that originate from the registry. This allows a harvester, particularly those outside of the VO community, to grab all records from a single, full registry.

A significant amount of (unglamorous) effort has gone into improving support for the registry standards across all our NVO registries. G. Greene has been improving the STScI registry’s interface and support for harvesting. M. Graham has released an exportable version of Carnivore, allowing easy installation of the registry system by other

organizations. Plante and R. Williamson (NCSA) are using results from the curation study to improve their publishing registry through better documentation as well as software improvements.

## 5 Data Models

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

Work on test implementations of the SED (Spectral Energy Distribution) model continued during this quarter. T. Budavari (JHU) and I. Busko (STScI) have partial implementations and SAO is continuing work on a full implementation. (European collaborators are also working on implementations.) This work has led to discussions clarifying some aspects of the serializations of the model such as the need to support both fixed and variable length arrays in the FITS version.

The IVOA VOEvent workshop at Caltech has concluded with an agreement by an international team on an information infrastructure to support the burgeoning field of event-based astronomy. The detection and understanding of these transient and immediate phenomena requires new kinds of surveys, fast data reduction, and fast response by the astronomical community. Providers of the event stream include the Swift and Integral satellites for gamma-ray bursts, supernova "factories", and synoptic surveys coming online now and planned for the future. There are robotic telescope projects that will respond in seconds to these discovery events, giving a comprehensive, panchromatic view. Until now, events have been distributed in various formats and protocols, so that aggregation and federation have been difficult. Users of such events have been forced to create specialized software to receive such streams. The objective of the VOEvent working group is to build an open standard for exchanging messages about these immediate astronomical events, including publication, archiving, query, subscription, and aggregation. The VOEvent standard has been agreed in rough form at the workshop, and includes the following features:

- Buy-in for the new standard from projects that include: GCN, LSST, Pan-STARRS, Palomar-Quest, LIGO, eStars, Raptor, Pairitel, ATEL, and Hands-On Universe.
- The possibility of simplicity and also the possibility of positional and semantic accuracy,
- Division into sections: Who, Where, When, What, How, Hypothesis, and Citations.
- Contact and publisher information so that consumers of events can restrict event subscription to trusted sources,
- Rich semantic content through the IVOA UCD vocabulary to express the meaning of project-specific data in an interoperable way,
- Integration with the IVOA Space-Time Coordinate schema, so that locations and timing can be defined either very simply or with sophistication and precision,
- Integration with the Remote Telescope Markup Language (RTML) so that VOEvents can be rapidly converted to scheduling instructions for a telescope,

- A vocabulary of event types that can be used to express a hypothesis about the astronomical meaning of the event, a way to express association with, and positional offset from a given object,
- A global identifier structure for events so they can be cited into the future,
- Message types such as Discovery, Followup, Retraction, and Supersede to provide a coherent picture of distributed knowledge about a discovery,
- Modular XML syntax, allowing the use of pre-built tools for parsing, storage, and filtering, as well as a clear path to extensibility.

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

B. Thomas (U. Maryland) has presented a schema for Characterization, describing the context of an observation or dataset. The CDS group has presented a similar schema and we will attempt to reconcile them at the Kyoto IVOA Interop meeting. Both approaches develop the “standard schema” approach proposed by J. McDowell at the Pune IVOA Interop meeting last fall in which elements from the Space-Time Coordinates and Quantity models are reused.

### *5.3 Descriptors and ontologies (UCDs)*

The DM group will review the existing UCDs needed for the SED and characterization models and present them to the UCD working group at the May Interop meeting.

The VOEvent activity (WBS 5.1) has generated fresh requirements for the UCD working group in IVOA. The VOEvent model includes description of relevant observational parameters with semantic information encoded as UCD. Further, the VOEvent includes a “hypothesis” section where the creator of the event can express what the nature of the event might be, for example “supernova”, or “cataclysmic variable report”, or “gamma-ray burst”. In the next quarter, these words will be refined and classified into a new branch of the UCD tree that describes immediate astronomical events.

### *5.4 Space-Time and regions*

A. Rots further revised the Space-Time Coordinate metadata specification and XML schemata. The final result was version 1.20 of the document and version 1.21 of the schemata. The latter validated under Xerces 2.3. The Working Draft was promoted to a Proposed Recommendation. In addition, two Notes were issued on the two STC implementations, STC-X and STC-S. Rots joined the VOEvent working group and is investigating the use of STC in that context. This application requires the use of Xincludes, which are now being implemented.

### *5.5 Standard schema*

Work on standard schemas in this quarter was carried out as part of the Characterization work.

## 6 Data Access Layer

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

Most of the effort on DAL services this past quarter focused on the Simple Spectral Access (SSA) protocol specification, and related changes to the SSA data model. The early work on this was done by M. Dolensky (ESO) and D. Tody. Toward the end of the quarter the draft interface was circulated to a small working group for discussion, with the goal of releasing the draft SSA interface publicly in early May in advance of the Kyoto interoperability meeting.

The scope of the SSA interface is to provide access to all tabular spectrophotometric data, including 1D spectra, time series, and SEDs. However, much of the work in developing SSA has been of a more general nature, dealing with how to model datasets in the VO and characterize them uniformly in the DAL query. Dataset metadata can be broken down into two parts, object metadata, which is specific to the type of data object represented (image, spectrum, etc.), and generic dataset metadata, which applies to all datasets. The generic dataset metadata describes datasets in terms of identification, curation, provenance, and physical characterization of the data. Physical characterization includes specifying the coverage, sampling, resolution, and accuracy of the spatial, spectral, and time axes of the data.

Once the new SSA interface is stable we will produce a new version of the Simple Image Access (SIA) interface that uses much the same generic dataset metadata, differing primarily in the object model. Other new interface elements, such as the mechanism used to query for service metadata, and the approach used for compression, should also carry over to SIA.

An effort has begun to develop a new DAL interface for spectral line lists, led by M. Dubernet (VO France). A sub-group of a dozen or so people has been formed, and an initial survey performed to assess the line list standards used in the various branches of astronomy. This will probably require a separate small interface dedicated to spectral line lists. This would consist of a query interface that returns spectral lines as rows in a VOTable.

A separate effort (I. Chilingarian, VO Russia; P. Prugniel, VO France; F. Bonnarel, CDS) has begun to begin to look at what is required to be able to support 3D data. This includes support for 3D images such as spectral data cubes, and MOS and IFU data. Probably there is no one solution for 3D data, and the approach required will involve extensions of the SIA and SSA interfaces to support data cubes and MOS data, plus more explicit support for complex 3D datasets along the lines of what is being done by the Euro3D effort.

### 6.2 Data representation (*VOTable, etc.*)

F. Bonnarel has updated the metadata extension mechanism proposal in preparation for the Kyoto discussions. This allows auxiliary metadata to be included in a VOTable (such

as the DAL query response table) to provide a way to describe the details of specific datasets.

### *6.3 Framework (mediators, components)*

The proposed component-framework architecture for data analysis, developed in collaboration with OPTICON in a workshop held at ESO last December, was discussed within the NVO and IVOA at the AAS and at an IVOA executive committee meeting in San Diego in January. The architecture was adopted as the basis for further design within the OPTICON group at the end of January. Related prototypes are underway within projects at NRAO, NOAO, and ESO. Technical discussions this past quarter have focused mainly on the component container architecture and the communications and messaging infrastructure.

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

The DPOSS (Digital Palomar Observatory Sky Survey) image data (3 TB) is having additional access services for its TeraGrid presence. In addition to a service that returns the complete 1.1 GB image file, there are also services to build cutouts of DPOSS images, and a service to provide a “replacement WCS” for the image. The purpose of this is to resample DPOSS into wide-area Hyperatlas pages, and the existing polynomial plate solution makes this a very computationally intensive process. However, when the plate is split into many small pieces, each can have a very simple (TAN) astrometry projection, and the reprojection is faster by a factor of 20 or so.

The DPOSS catalog is now being served as a SkyNode from Caltech, so that dynamic cross-matches can be done between SkyNodes in Baltimore (Sloan, 2MASS, etc) and Pasadena. The publishing of this SkyNode is being held up for data quality issues.

I. Busko, P. Osuna, and others have implemented data analysis clients that can consume SSA data. P. Protopapas has begun to look at the use of SSA for time series data from MACHO, OGLE, and other time series datasets. R. Thompson has begun looking at reworking the STScI spectral archives to be more consistent with SSA, and subsequently implementing spectral services for STScI spectral data.

## **7 Query Language**

### *7.1 Low-level: Astronomical Data Query Language*

A focused working meeting on ADQL was organized in Tokyo by M. Ohishi (Chair of IVOA VOQL Working Group). W. O'Mullane and N. Li (JHU) attended this meeting in February. Unfortunately F. Ochsenbein (CDS) could not attend. The current state of JVOQL was demonstrated, which includes advanced region functionality. A one-hour presentation on NVO was delivered. Minor modifications were made to the ADQL specification, and draft version 0.91 was forwarded to M. Ohishi.

### *7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode*

The OpenSkyQuery interface and portal was presented at the San Diego AAS meeting. A good to deal of effort of was put into manning the booth and demonstrating this and other applications. This was a very successful endeavor.

At JHU a preliminary parser for version 0.91 has been constructed and has been incorporated into some nodes and the portal on dev.openskyquery.net. A number of groups are working using the AXIS toolkit to produce a SkyNode implementation in Java. HEASARC has made the most progress with this project. VO India, ESAC (Spain), and JVO remain active in this area as well.

Technical problems experienced last quarter have now been surmounted. An attempt is being made to extract the query builder software from the portal and make an independent version that may be of use to others. As a first project this will be integrated with the sky server.

New SkyNodes have been built by CACR and U. Pittsburgh.

### *7.3 High-level: Complex queries*

Nothing to report this quarter

## **8 Web and Grid Services**

In this quarter, the NVO has been concentrating on building a security structure for NVO services that is simple to use – or invisible – for simple applications, yet sophisticated and powerful when necessary (see WBS 6.5). Another thrust has been the scaling of services: so that multiple users of a service can be farmed out to a cluster, and so that large requests can be handled asynchronously, meaning submit the job, monitor/notify until done, collect results. Services are being deployed on the NSF TeraGrid at Caltech and SDSC for exposure of major image surveys such as DPOSS, 2MASS, Palomar-Quest – see WBS 6.4 and 9.3.

### *8.1 Web Services (SOAP, WSDL, etc.)*

WESIX (Web Enabled Source Extraction and Cross-match) has been released to a broader astronomical community and has been extended to enable images to be specified as URLs. A client access to the WESIX web services is being implemented to enable automated source detection and calibration of user defined images (returning fully calibrated catalogs not just cross-matched source lists). WESIX is being implemented under the NOAO archive (NSA) to provide a tool for source detection on NSA image data products.

### *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. For full details, see WBS 3.1.

### 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

Caltech is collaborating with the High Energy Physics group (LHC-CMS) in creating service-based science gateways, using the Clarens secure web server. Such gateways will be, for example, for image mosaicing, source extraction and matching, etc. Users will be able to use the gateway in a graduated way through either (a) anonymous jobs that have little privilege, (b) temporary or local certificates with somewhat greater privilege, or (c) a well-authenticated certificate from a user with an account on the system, for which very great privilege will be granted. The secondary purpose here is to acclimate the astronomy community to the concepts of certificate use.

The connection between NVO and NSF TeraGrid has been renewed. In January, a team led by Connolly (U Pittsburgh) and Williams (Caltech) summarized how the first-year TeraGrid allocation was used for astronomical data processing and service exposure, and made a case for continued collaboration between the two projects. This application for compute resources was accepted and a further grant of 407,000 CPU-hours made to the NVO.

## 9 Applications

At the beginning of this period, substantial support was given to the presentation of the initial release of VO services at the Summer School. All services in this release remain functional. Work is underway to provide for building a framework for collecting VO-wide statistics on the use of released VO services.

Preparations for the second NVO Summer School were a main focus in this quarter. The schedule and testing plan for VO applications and libraries to be used in the summer school was decided.

### 9.1 Data location services

Responding to user requests, some new capabilities were provided in the DataScope tool. All location services also benefited from a general ST/JHU-led effort to improve the quality of metadata stored in VO registries.

### 9.2 Cross-correlation services

Substantial development efforts in making a Java implementation of a full SkyNode implementation were underway at the HEASARC in collaboration with JHU and the NCSA. A simple test correlator services were made available. This preliminary services illustrates how SkyNodes can be used in very simple applications as well as providing complete ADQL access to underlying databases. These services are not yet ready for public use.

### *9.3 Image combination, registration*

The Montage image reprojection/mosaicing service prototype has been extended at IRSA to handle all large-scale NVO image sets (2MASS, SDSS, DPOSS and DSS). The Request Object Management Environment (ROME) system for handling asynchronous user requests has been incorporated into the framework. Efforts are underway to identify adequate resources to allow fielding this as an operational SIA-compliant service.

A new implementation of SkyView services at the HEASARC using the VO SIA protocol to access image sets is being prepared for release in mid-April. This new Java-based service uses the SIA protocol to enable it to mosaic images from any service that provides lists of images in a given region. SkyView access to the SDSS DR3 will be enabled using this new capability.

### *9.4 Visualization tools and services*

Nothing to report this quarter.

### *9.5 Theory*

Nothing to report this quarter.

### *9.6 Statistical analysis*

A data mining tool for the statistical analysis of the distribution of sources on the sky), measuring the angular correlation function of galaxies, has been integrated with the SDSS SQLServer databases and will be released this month for testing.

A web service interface to a principal component analysis of spectra is under development. This tool will return spectral eigenvectors for uploaded spectra together with classification coefficients of these spectra and enable the projection of external spectra onto SDSS defined eigenspectra.

### *9.7 Datamining, outlier identification*

Nothing to report this quarter.

### *9.8 Interfaces to/from legacy software systems*

The NOAO has been successful in provide VO integration capabilities within the IRAF environment. This allows users of the most-widely used astronomy package to access VO resources transparently.

## **10 Community Engagement**

### *10.1 Documentation*

A major documentation effort was undertaken in support of the software release in January at the San Diego AAS meeting (described in the previous Quarterly Report). Efforts continue to improve this documentation in response to user feedback.

### *10.2 Web site*

The NVO website (<http://us-vo.org>) has been running with dynamic content management throughout 2004. 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.

The NVO web server at us-vo.org has been replaced with a modern machine with 80 GB disk, running Fedora Linux. Much of the software infrastructure has been ported (Apache, MySQL, ColdFusion, Carnivore, Hyperatlas), and the mail lists are in process of being ported and updated from the majordomo software to mailman. The old web server machine is still in place as a hot-spare for the new one: it gets regular updates of the web content so that it is reasonably up-to-date at all times.

### *10.3 Technical training initiatives*

Planning for a second NVO Summer School got underway in the current quarter. A site was selected (the Aspen Meadows Resort in Aspen, Colorado) and a contract was put in place for the facilities and rooms. The duration of the Summer School is being extended in response to participant feedback from the first Summer School. This year the NVOSS will run from September 6-15. Advertisements were developed and announcements were sent to a number of e-mail distribution lists. Our focus is on enabling astronomical research with the NVO, and to this end we are putting a larger emphasis on participation from young astronomers, particularly graduate students and post-docs. A request for supplemental funding was submitted to NSF, and a request to NASA is in preparation.

The website for the Summer School is <http://www.us-vo.org/summer-school/2005/>.

### *10.4 Advocacy*

R. Hanisch assisted project director A. Szalay in preparing a briefing on NVO to the Astronomy and Astrophysics Advisory Committee (AAAC), which advises NSF, NASA, and other federal agencies. The presentation went very well. Jeffrey Hayes (NASA HQ) and Eileen Friel (NSF) told the AAAC that a jointly funded NVO Facility AO would likely be available late this summer.

The NVO Project submitted a White Paper in response to NASA's long-term planning initiative. The White Paper describes the role of NVO for NASA's Universe missions in the coming decades. See <http://us-vo.org/pubs/files/NVO%20Universe%20Roadmap.pdf>.

## **11 Education and Public Outreach**

### *11.1 Strategic partnerships*

The NVO EPO program interacts 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 program is coordinated by C. Christian.

The EPO personnel are working with the various 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*

Pre-college education projects developed at JHU, for example, *Adopt a Galaxy*, are incorporated into the NVO EPO Portal for use by educators. J. Raddick is responsible for working with teachers to create and maintain these modules. The NVO EPO team is strategizing methods for enabling testing of modules by teachers.

Spectroscopy exercises for astronomy and chemistry are being developed within Project LITE, under PI K. Brecher (Boston U.). This program is positioned well to incorporate spectra from the NVO, as the spectroscopic data holdings and access protocols develop. Students are introduced to temperature and composition concepts through comparison with templates. Actual data will be accessed to incorporate stellar spectra representative of common stellar types.

Project CLEA (PI L. Marshall, Gettysburg College) is working towards incorporating Sloan Digital Sky Survey data into their virtual observatory exercises. These activities enable students to “observe” the sky and obtain data for subsequent analysis. The CLEA structure can ingest NVO data for student exercises. CLEA personnel have plans to conduct feasibility studies, in collaboration with the NVO, to understand the best method for accessing NVO data within student exercises.

### *11.3 Informal education*

The Adler Planetarium in Chicago uses NVO data for interactive exhibits within the museum hall, demonstrating galaxy classification. Adler plans more workshops and classes based on the NVO summer school, conducted by M. SubbaRao.

The NVO group at Caltech has been working at a furious pace with the Griffith Observatory – a planetarium and education center that is a Los Angeles landmark. The Observatory is building a new facility that will include a sky mural, “The Big Picture”, that will be 150 feet wide by 20 feet high. The mural will be a picture of about 15 degrees of sky, including the Virgo cluster, and is being built with NVO-supported software and Tera-Grid hardware.

### *11.4 Outreach and press activities*

A NASA Applied Information Systems Research Program (AISRP) grant is being used as the vehicle to create a prototype process for creating a press release image collection to be registered in the NVO. This work is done in collaboration with the University of California Berkeley. The schema for describing the images is being formulated for a few test images. Berkeley is responsible for the public interface for accessing the press images.

NVO EPO personnel are coordinating the development of metatags for public information including NVO products with colleagues at the European Space Agency.

*11.5 Technical development*

The NVO Registry accommodates metadata geared to EPO consumers, and feedback from the EPO community may point towards adaptations to the NVO infrastructure as to accommodate EPO material.

## Activities by Organization

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

Good progress has been made over this quarter in the definition of the VOStore/VOSpace concept as a secure, distributed data system. M. Graham has created a prototype of the VOStore, and has been interacting with the UK AstroGrid project to ensure international interoperability.

The Carnivore registry, developed at Caltech, was released in early January as the “1.0 Final” version. Carnivore is a fully-featured open source VO registry service developed at Caltech.

The Caltech group has been involved with STScI and NCSA in building a “curation workflow” for the registry. When records are published to a VO registry, this series of steps will check and correct problems with the records, starting with automated checking tools, then human checks, and a way in which corrections can be made by interacting with the original publisher of the records.

The IVOA VOEvent workshop at Caltech has concluded with an agreement by an international team on an information infrastructure to support the burgeoning field of event-based astronomy. The objective of the VOEvent working group is to build an open standard for exchanging messages about these immediate astronomical events, including publication, archiving, query, subscription, and aggregation.

The DPOSS (Digital Palomar Observatory Sky Survey) image data (3 TB) is having additional access services for its TeraGrid presence. Also the DPOSS catalog is now being served as a SkyNode from Caltech, so that dynamic cross-matches can be done between SkyNodes in Baltimore (Sloan, 2MASS, etc) and Pasadena.

Caltech is collaborating with the High Energy Physics group (LHC-CMS) in creating service-based science gateways, using the Clarens secure web server. Such gateways will be, for example, for image mosaicing, source extraction and matching, etc. The connection between NVO and NSF TeraGrid has been renewed.

The NVO website (<http://us-vo.org>) has been running with dynamic content management throughout 2004. 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.

The CVS software repository is now established and populated at Caltech. The NVO web server at us-vo.org has been replaced with a modern machine with 80 GB disk, running Fedora Linux. Much of the software infrastructure has been ported (Apache, MySQL, ColdFusion, Carnivore, Hyperatlas), and the mail lists are in process of being

ported and updated from the majordomo software to mailman. The old web server machine is still in place as a hot spare for the new one.

There has been much progress on planning for the second NVO Summer School. Caltech has been active in many aspects, including significant faculty involvement (Graham, Williams).

The NVO group at Caltech has been working at a furious pace with the Griffith Observatory – a planetarium and education center that is a Los Angeles landmark. The Observatory is building a new facility that will include a sky mural “The Big Picture”, that will be 150 feet wide by 20 feet high.

#### *Canadian Astronomy Data Centre/Canadian Virtual Observatory*

CADC has made major improvements to the CVO web site and web-based data exploration tools and added a new forms-based tool to provide direct access to the same content. Visualization with Aladin for the images (WFPC2 Associations and ROSAT fields) and SpecView for spectra (2QZ) has been integrated, including WFPC2 source catalog overlay when viewing the WFPC2 Association images. We now support download of query results in VOTable and CSV format and loading VOTable data directly into VOPlot for analysis.

We have registered the CADC, several CADC archives (data collections), projects, and services in NVO registry, using the portal at NCSA.

The CVO project has completed implementation of SIAP 1.0 compliant services for CFHT (CFH12K and Megaprime), HST (WFPC2 B-Associations), and JCMT (SCUBA) with all optional output (WCS, time, and bandpass metadata) available. The CFHT and HST services perform image cutouts and all returned images have corrected WCS parameters in the FITS headers. Total content: ~500,000 images (mosaic cameras produce one SIAP “image” per extension).

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

S. Krughoff, J. Brewer, and A. Connolly worked on extending WESIX to provide access via URLs rather than uploading images. WESIX was publicized to a broad astronomical community through the SDSS mail exploders. With C. Miller (NOAO) WESIX is being deployed as part of the NOAO archive to provide source detection for NSA users.

J. Brewer, S. Krughoff, and A. Connolly are implementing a client application for WESIX that provides photometric calibration for input data. This application is a prototype to demonstrate the use of web services as a pipeline processing tool.

J. Brewer and S. Krughoff with help from W. O'Mullane (JHU) and A. Conti (STScI) have automated the scripts and are documenting the procedures for deploying SkyNodes. This documentation is aimed at users without intimate knowledge of the NVO. J. Brewer has deployed a SkyNode (the Phoenix SkyNode) as part of the clean up and documentation of the scripts.

C. W. Yip, S. Krughoff, and A. Connolly are implementing a web service interface for a series of spectral classification tools based on principal component analysis.

*Fermi National Accelerator Laboratory (FNAL)*

S. Kent and N. Sharma continued testing and building interfaces to the OpenSkyQuery web service portal (WBS 9). Client side tools were developed to allow automation of queries and conversion of returned results to FITS binary tables.

*High Energy Astrophysics Science Archive Research Center (HEASARC)*

During the last quarter the HEASARC was deeply involved in beginning the process of setting up the curriculum for the second NVO Summer School. Initial drafts of the summer school content and schedule were made and the summer school faculty was recruited. This will be the primary focus of activities in the next two quarters.

Some modifications of the DataScope tool were made during this period mostly to address minor bugs and to collect better statistics on DataScope usage. No major problems were found in DataScope operation. During the coming quarter access to spectral services will be added to the DataScope.

HEASARC metadata were reviewed and numerous small problems fixed. Verification of the HEASARC metadata using the OAI explorer showed some inconsistencies with the current version of XERCES parsers. While the HEASARC metadata seem to be acceptable to current full registries, these will be addressed. A problem with updates not propagating correctly to the ST/JHU registries was brought to the attention of personnel there who will address it.

*Johns Hopkins University*

T. Budavari continued his work developing Simple Spectra Access Protocol (SSAP) and the Spectrum Services web site.

W. O'Mullane spent a week in Kyoto with N. Li and M. Ohishi (JVO/NAOJ), where they developed SkyNodes and installed a version of CasJobs on NAOJ's server. O'Mullane also spent three days in Cambridge working on VOStore, VOspace, and security issues with N. Walton (AstroGrid). O'Mullane also worked on the specifications for VOStore, helped set up and take down the NVO display at the AAS meeting, and answered questions from the community at that meeting.

Budavari, O'Mullane, and Nieto-Santisteban participated in the preliminary discussions of the next NVO Summer School.

N. Li linked ADQL 0.9 to the OpenSkyQuery development site ([dev.openskyquery.net](http://dev.openskyquery.net)), worked in Japan with W. O'Mullane, and developed SkyNodes in Java. He made improvements to CasJobs. He also began work on a new query builder, based on SkyQuery, which will allow users to build queries for single databases.

M. Nieto-Santisteban extended the zone-based data partitioning algorithm to the Tera-byte-sized SDSS DR2 database (BESTDR2), and ran the test problem (MaxBCG cluster finding algorithm) on the entire BESTDR2 data (3300 square degrees, 1 TB). She has obtained a performance increase of at least a factor of 50 over the file-based TAM (Tera-byte Analysis Machine) implementation that J. Annis et al. (Fermilab) developed. This impressive speed-up is mostly due to the better native indexing system that the Database Management System (DBMS) provides in combination with the Zone algorithm. The MaxBCG computation for the whole BestDR2 database demonstrates the benefits of bringing the analysis to the data in the form of a sequence of SQL functions and stored procedures.

T. Malik and Nieto-Santisteban worked on creating a scheduler for parallel execution of database jobs using the workflow model used by the sqlLoader data loading pipeline, which was developed by A. Szalay, J. Gray (Microsoft) and A. Thakar for SDSS data loading.

Nieto-Santisteban and J. Gray have also been experimenting with parallel, purely spatial USNOB $\leftrightarrow$ 2MASS all-sky cross-matches. The next steps are to include science criteria in the cross-matching, and to execute the cross-matches within the job scheduler framework, so that the parallel workflow can be automated.

The zone-based partitioning, the job scheduler and the parallel cross-matching will eventually be applied to SkyNode cross-matches to speed them up for large areas.

T. Malik worked on caching for SkyQuery.

A. Thakar has built a prototype of a logging system that has not yet been made public. This prototype is based on version 0.22 of the IVOA Support Interfaces Working Draft written by W. O'Mullane.

A. Szalay has worked with J. Gray to complete modifications inside SQL Server to allow footprint, intersection, and algebra services. These have not yet been released to the public.

G. Fekete continues his work on HTM, particularly to get it to compile with new Linux C++ compiler.

J. Raddick continued his education and outreach work on the virtualobservatory.org website. He added a new project where students use a Registry query to build an H-R diagram. He continues working with C. Christian (STScI) to fit his projects into the larger scheme of NVO education and outreach. He presented his first education project (Adopt an Object) at ADASS, and attended the AAS meeting, where he collected a list of teachers interested in NVO education.

S. Carliles worked on a C# port of T. McGlynn (HEASARC)'s Java FITS library. He worked on porting Starmaps, a program to display constellation images and boundaries,

into C#. He will make this program available to give a larger view for programs that display data on patches of sky.

#### *Microsoft Research*

J. Gray collaborated with JHU personnel on several VO-related papers.

#### *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. In particular, De Young is working with the NVO Science Steering Committee (SSC) to review and prioritize a new set of tools and applications to be released to the US and international astronomical communities. This new set will provide extensions to the recently released first set of tools made available in January 2005, and in addition it will contain tools that provide completely new capabilities. In addition, De Young is participating in the Theory Interest Group of the International Virtual Observatory Alliance (IVOA) as it prepares for theory demonstrations and reports to be made at the May 2005 interoperability meeting in Kyoto. De Young also continues to provide input to the design of the Science Reference Mission of the European VO project.

Having demonstrated last quarter the technology to build important web services using IRAF-based applications as a back-end, M. Fitzpatrick and O. Malik (a new CS student intern) began refactoring the IRAF web-services framework to allow for easier configuration by users and to simplify long-term support. The new model defines the web-service as an XML configuration file and XSLT files are used to auto-generate the web-service Java code. Much of the original code has been moved to support servlets, making the web-service layer much lighter.

A new server designated *nvo.noao.edu* was purchased and configured to act as the main NVO machine in Tucson. This will host the IRAF web services as well as other VO-related projects within NOAO, a mirror system is planned for CTIO. The NCSA Registry-in-a-Box was installed with the intent that it serve as a Publishing Registry for NOAO resources. Discussion is underway concerning which resources other than access to the Survey Data in the NOAO Science Archive, plus IRAF web-services, should be made available.

M. Fitzpatrick and O. Malik also began work on a web application designed to make searching for VO data and related information easier for users. Fitzpatrick was also invited to give a talk at STScI on the IRAF web services work, it is hoped that other institutions with IRAF-related projects will be willing to collaborate on creating additional such services for the community.

#### *National Radio Astronomy Observatory (NRAO)*

A comprehensive NRAO VO plan has been prepared and was submitted to the Director's office at the end of March. A key element of the plan is the development of a capability to publish user-generated science data products to the VO. The NRAO archive infra-

structure would be enhanced to provide a project view including all raw and processed data products for an observing project. Standard VO services would be implemented to provide access to catalog, image, spectral, and time series data products. Included is an effort to experiment with the generation of reference images for a subset of the observational data in the VLA archive. The VLA archive imaging pilot project will produce images for continuum data for the VLA B configuration at 5 and 8.4 GHz.

The annual NVO Advisory Committee meeting was held at the San Diego Supercomputer Center in January. D.Tody reported on the data access layer (DAL), including the VO data access services and plans for integration of data analysis software with VO.

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

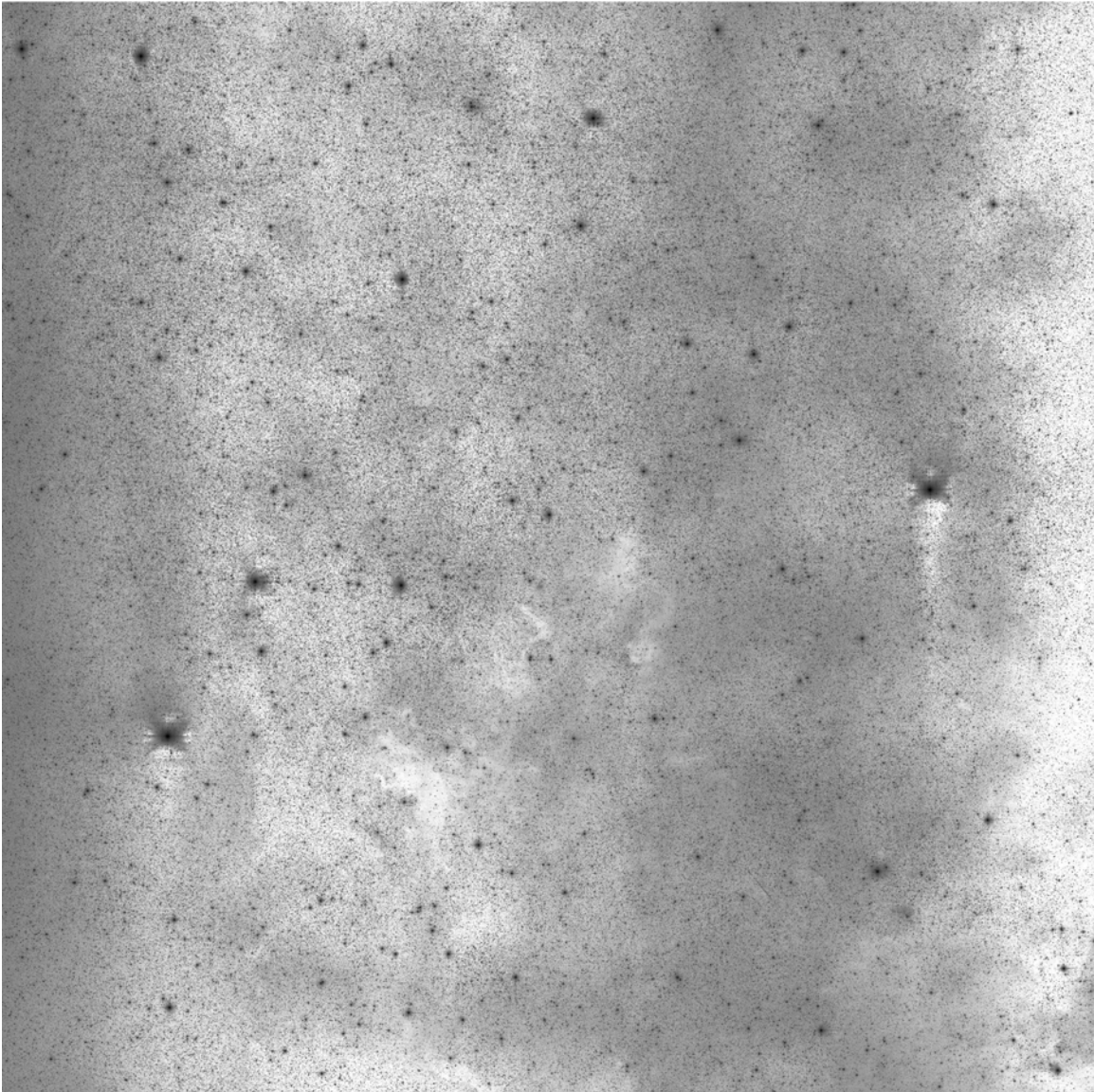
GMU staff contributed to the IVOA discussions on the new VOEvent requirements and protocols, gave a presentation on Distributed Data Mining in the NVO to the NASA Applied Information Systems Research workshop at the Ames Research Center, participated in planning activities for an Education and Public Outreach (EPO) component for NVO, and began collaborating with the LSST Data Management team. For the LSST project, GMU staff contributed to an LSST team meeting in areas related to NVO protocols for data access, data integration and interoperability, VOEvent requirements, and EPO. GMU staff also participated in the on-going development of an NVO-centric EPO proposal to the NSF. Finally, GMU staff gave public lectures on the NVO project to faculty and students at UMES (University of Maryland, Eastern Shore) and to the general public at the Astronomical Society of Harrisburg (Pennsylvania).

U. Maryland staff worked on developing the schema and examples for the Catalog data model. Staff have proposed a more proper division of concepts between STC, Characterization, and Quantity that reduces redundancy and inconsistencies. A set of XML Schema and UML diagrams define the proposed model.

*San Diego Supercomputer Center*

Version 3 of Montage, along with updated utility routines (for querying the IRSA database, etc) that were supplied by J. Good have been installed on three systems, SDSC TeraGrid node, NCSA TeraGrid node, and SDSC DataStar. R. Williams generated a list of the 1734 plates that will make up the Hyperatlas archive. L. Brieger chose the first plate to use for testing (located at the pole), and generated a 1-degree mosaic. The very high density of coverage and overlap in the raw data at/near the poles caused an incredible I/O bottleneck, requiring hours of CPU time to get through some of the computationally simplest steps. By parallelizing the differencing routine and taking advantage of the GPFS parallel file system and MPI2, both of which are available on the systems where the codes will be running, L. Brieger was able to use MPI I/O in the parallel differencing routine, speeding up computation of the "pathological" example at the pole by a factor of between 4 and 5. Examples near the equator do not show such a speedup and indeed may be penalized by the overhead of MPI I/O. The MPI I/O code had some trouble with crashing the GPFS file system on the TeraGrid systems for large numbers of CPUs. At 128 CPUs or fewer, the problems do not occur.

L. Brieger generated an example 6-degree mosaic for inspection, at which point it was decided that the header files for these mosaics should be slightly different from that generated by the routine provided with Montage. The routine to generate the new headers was created and additional 6-degree examples were generated. These examples have been placed in the SRB repository for viewing by the NVO team. J. Good has examined the first plate and provided the image shown in Figure 1.



*Figure 1.* J. Good took the 10X-shrunken version of the 6-degree square plate (#1484) and made the JPEG. The background was flattened and stretched to enhance the low-level structure. There may be some observing artifacts in the large scale bright and dark bands or it may be real. More plates will need to be created to tell.

It has also been decided that storing a 6-degree mosaic as a single 4GB file, is too unwieldy, and that the mosaics should be “tiled”, so that the final image will be composed of 144 sub-images which are part of the single 6-degree mosaic. The tiling routines, again provided by J. Good, have been tested on the three systems. These routines provide a simple way of parallelizing the final co-add step as well as of storing the final image data in a piecewise form. The parallelization of the co-addition step has been implemented and tested so far on DataStar.

#### *Smithsonian Astrophysical Observatory*

G. Fabbiano participated in CfA and collaboration meetings on the future of the NVO. She is also the organizer of a workshop at the COSPAR meeting in Beijing 2006, on the needs and challenges of high resolution (imaging and spectral) space observatories. She plans to have a VO-flavored session and is in the process of finding sponsors.

J. McDowell has been working on the SED and Characterization models and standard schema issues.

K. McCusker has been working on implementation of an SED model library in Java, with FITS, XML and VOTABLE serializations. She has also been working on installing a copy of Caltech's Carnivore registry on our system. We plan to make a CfA registry site public next quarter. McDowell and McCusker, with help from E. Galle, have begun updating the CfA's VO website.

A. Rots continued work on Space-Time Coordinates and is pushing the document through the approval process.

#### *Space Telescope Science Institute*

G. Greene upgraded the local searchable NVO Registry for the standard Open Archives Initiative (OAI) protocol publishing and harvesting. This is part of an ongoing IVOA initiative to have fully compliant Registry schema exchange between all registries. The enhancement of Dublin Core schema publication is nearly complete.

STScI and JHU VO team members met with T. McGlynn (NASA GSFC) and discussed technical interface issues between the local VO Registry and DataScope application. Inconsistencies have been resolved, and with the correction of the “Date” fields harvested from the HEASARC OAI repository, automated harvesting of registry resources is now working properly.

The STScI computer infrastructure for support of NVO activities has been stabilized on

Windows 2003 servers. The study for operating the NVO TWiki on IIS concluded with the decision to maintain this site on the existing server with Apache given the technical incompatibilities between the IIS Web Server and TWiki software. All NVO data services affected by OS/server migrations were successfully reinstalled and are operational.

R. Hanisch attended the final meeting of the Astrophysical Virtual Observatory (European VO) Science Working Group in Villafranca, Spain, January 25-26. The AVO project is now concluded as the Euro-VO begins. AVO showed two science cases, one aimed at detecting stars as they evolve from the asymptotic giant branch to planetary nebulae and another focused on studying galaxy evolution through comparison with models of spectral energy distributions.

In February we started a new effort on registry metadata curation, assisted by D. Norman (CTIO). Norman is documenting her use of the registry as a research user, and a small team (R. Hanisch, R. Plante/NCSA, and S. Emery Bunn/Caltech) undertook a systematic review of registry entries based on Norman's feedback. The review was undertaken one provider at a time, allowing data publishers to make some repairs "en masse".

The Aspen Meadows Resort in Aspen, Colorado has been confirmed as the site of this year's NVO Summer School (6-14 September). A room and facilities contract was negotiated by R. Hanisch and signed by JHU. A supplementary funding proposal was submitted to NSF; NASA will also be approached for support.

R. Hanisch prepared an updated NVO Project Plan and distributed it to the NVO team. The project plan defines work areas and sets priorities for the remainder of the NVO development project. The plan is available at <http://us-vo.org/pubs/files/Project%20Plan%20Update%20Feb%202005.pdf>.

C. Christian worked with L. Christensen (ESO) in a review of resource metadata for education and public outreach resources. She also visited Gettysburg College to meet with L. Marschall (Project CLEA) about how they can utilize NVO resources in their educational environment. In addition, work is getting underway in a collaboration with UC Berkeley, "Virtual Cosmos," in which HST press release images will be made available through standard NVO interfaces. This project is funded by NASA's Applied Information Systems Research Program (AISRP).

#### *United States Naval Observatory*

Catalog search services at USNO have been upgraded to provide VO conesearch access to the USNO-B1, USNO-A2 and NOMAD catalogues. 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, as lead for registries, continued to coordinate the upgrade of NVO registries to VOResource v0.10 and moderated discussion of future changes to the XML schemata. He continues to contribute to the refinement of the IVOA Registry Interface in collaboration with the IVOA Registry Working Group. He has also participated in the Registry Curation tiger team to study metadata quality issues. Finally, he is participating in the planning of the NVO Summer School as a faculty member.

R. Williamson and Plante have been making improvements to the VORegistry-in-a-Box package in response to the study of metadata quality (see section 4.5). We are avoiding drastic changes to the package in anticipation of merging our efforts into the Carnivore project (from Caltech) later this year.

Williamson and Plante are collaborating with T. McGlynn on the extension of the Java Sky Node package and its incorporation into the new software repository.

Williamson has joined the VOEvent collaboration with a special focus on application to the Large Synoptic Survey Telescope. In April, he attended the VOEvent workshop in Pasadena.

*University of Pennsylvania*

U. Penn participants C. Alcock and P. Protopapas have relocated to SAO, and their activities are described in the SAO report.

*University of Southern California (USC/ISI)*

This report highlights the work done at ISI during the first quarter of CY2005 within the NVO project. The work during this period has mostly focused on the changes required to the Montage DAG service in order to make it work seamlessly with the three image archives (2MASS, DPOSS, SDSS) and added flexibility in specifying the desired mosaic size. ISI was involved in testing and debugging the new Montage DAG service interface with the above mentioned features.

The Montage DAG service invocation syntax was changed to specify the survey to be used as source of the input images (2MASS, DPOSS, SDSS) and the size of the desired in terms of height and width. Earlier the DAG service only worked with 2MASS data, and the size of the mosaic was specified as a single parameter. The generated mosaic used to be in square degrees. With the current interface the user can specify any arbitrary height and width. Integration of the DAG service with the other image survey access tools necessitated further testing. Several errors were discovered due to the format of the output produced and the file naming conventions of the image access tools. Each survey uses a specific set of bands and the band specified in the user request should be in agreement with the image survey specified.

The SDSS survey returns image files in zipped format. The files could be unzipped before being send to the compute resources or the Montage modules could unzip it on the

compute site before using it. A decision to use the latter option was taken since it reduces the amount of data transfer, which as we had seen in one of our benchmark studies, constitutes a significant portion of the workflow execution time. Also it requires no changes to the existing Montage workflow.

*University of Wisconsin*

No activity reported this quarter.



**Publications**

Djorgovski, S. G.; Williams, R., "Virtual Observatory: From Concept to Implementation." Invited review, to appear in proc. "From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science", eds. N. Kassim, M. Perez, W. Junor & P. Henning, ASPCS vol. 3xx, in press (2005). astro-ph/0504006

**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