Quarterly Report
April – June 2006

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

IVOA

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

Executive Summary **1**
Activities by WBS **3**
1 Management **3**
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 4
1.5 International coordination/collaboration 4
2 Science Requirements **4**
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
3 System Architecture **4**
3.1 System design, components; relationships to Grid components 5
3.2 Computational facilities (processing, bulk data storage, network access, security, authentication) 5
3.3 Digital library integration 6
3.4 Preservation 6
4 Registries **6**
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
5 Data Models **9**
5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality) 9
5.2 Low-level (measurement, quantity, uncertainty, relationship) 10
5.3 Descriptors and ontologies (UCDs) 10
5.4 Space-Time and regions 10
5.5 Standard schema 10
6 Data Access Layer **11**
6.1 Data access services (catalog, image, spectrum, time series, visibilities, …) 11
6.2 Data representation (VOTable, etc.) 15
6.3 Framework (mediators, components) 15
6.4 Data provider/consumer implementations and end-to-end testing 16
7 Query Language **17**
7.1 Low-level: Astronomical Data Query Language 17
7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode 17
7.3 High-level: Complex queries 18
8 Web and Grid Services **18**
8.1 Web Services (SOAP, WSDL, etc.) 18
8.2 Grid Services (OGSA) 18
8.3 Computational resource management 19

8.4 Virtual data ... 19
8.5 Application and service integration with Grid ... 19
9 Applications ... 20
9.1 Data location services ... 20
9.2 Cross-correlation services ... 20
9.3 Image combination, registration ... 21
9.4 Visualization tools and services ... 21
9.5 Theory ... 21
9.6 Statistical analysis ... 21
9.7 Data mining, outlier identification ... 22
9.8 Interfaces to/from legacy software systems ... 22
10 Community Engagement ... 22
10.1 Documentation ... 22
10.2 Web site ... 22
10.3 Technical training initiatives ... 23
10.4 Advocacy ... 23
11 Education and Public Outreach ... 23
11.1 Strategic partnerships ... 23
11.2 Formal education ... 23
11.3 Informal education ... 24
11.4 Outreach and press activities ... 24
11.5 Technical development ... 24
Activities by Organization ... 25
Caltech−Astronomy Department and Center for Advanced Computational Research (CACR) ... 25
Caltech−Infrared Processing and Analysis Center (IPAC) ... 25
Canadian Astronomy Data Centre/Canadian Virtual Observatory ... 26
Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt) ... 27
Fermi National Accelerator Laboratory (FNAL) ... 27
High Energy Astrophysics Science Archive Research Center (HEASARC) ... 27
Johns Hopkins University ... 28
Microsoft Research ... 29
National Optical Astronomy Observatories (NOAO) ... 29
National Radio Astronomy Observatory (NRAO) ... 30
Raytheon/ADC (University of Maryland and George Mason University) ... 31
San Diego Supercomputer Center ... 32
Smithsonian Astrophysical Observatory ... 33
Space Telescope Science Institute ... 34
United States Naval Observatory ... 34
University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA) ... 35
University of Southern California (USC/ISI) ... 35
Publications and Presentations ... 36
Virtual Observatory Articles in the Popular and Technical Press ... 37
Acronyms ... 38

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

Period covered by this report: 1 April — 30 June 2006
Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

The two major events of this Quarter were the project team meeting at Caltech, 3-4 April, and the IVOA Interoperability workshop in Victoria, BC, 15-19 May. NVO and IVOA efforts are focused on completing Version 1.0 standards and protocols for all core VO functions and services by the end of the current fiscal year. The NVO team continues work on its contributions to these standards and protocols, with the goal of reaching closure at the fall IVOA Interop meeting in Moscow (18-22 September).

The program for the Special Session on Virtual Observatories, which will be held during the IAU General Assembly in Prague in August, has been finalized. It includes both scientific and technical presentations, including several from NVO team members and collaborators. Several NVO team members are on the organizing committee. NVO will join with our IVOA colleagues in a demonstration booth at the General Assembly.

Preparations for the third NVO Summer School continue. Supplemental funding requests to support the NVOSS were approved by NSF and NASA. Applications were received from 56 people and 41 were accepted by the NVOSS faculty.

The major technical activities of the past quarter included:

- Reaching agreement on restructuring of the registries that more easily accommodates multiple types of services for the same data collections and that allows schema extensions for detailed metadata caching. Software revisions to the NVO registries are being carried out this summer and fall.
- Establishment of a "registry of registries" (RoR). The RoR provides a single top-level directory service, analogous to a domain-name service, where IVOA registry services are themselves registered.
- Reaching agreement on the VOEvent protocol.
- Making further advances in the spectrum data model and simple spectrum access protocol (with closure planned by the Moscow Interop workshop).
- Discussion of approaches to distributed cross-matching and planning for the evolution of the SkyNode database interface.
- Simplification of the Astronomy Data Query Language (ADQL) syntax.
- Further development of VOSpace for distributed data storage, with the associated authentication and authorization tools.
- A major reworking of the DataScope user interface.

Financially the project remains in good condition, with spending proceeding at expected levels. A proposal was submitted to NSF for continuing funding in FY2007, to accommodate the delays in issuing the solicitation for the NVO Facility.

Activities by WBS

1 Management

1.1 General (planning, reporting, communications, team meetings, etc.)
Regular weekly telecons of the Technical Working Group (TWG) continue. Similarly, the Executive Committee also continues to meet weekly by telecon.

The spring team meeting was held on 3-4 April at Caltech. The agenda and presentations are available at http://chart.stsci.edu/twiki/bin/view/Main/TeamMeetingSpring2006.

Supplemental funding proposals were submitted to NSF ($93,650) and NASA ($33,000) for the NVO Summer School. Both were approved.

A supplemental funding proposal was submitted to NSF ($1,343,674) to sustain the NVO development project through FY2007. This support is needed in order to assure continuity between the development team and the eventual NVO Facility. The solicitation for the NVO Facility continues to be delayed owing to the complications of setting up a jointly funded (NSF, NASA) program.

1.2 Science
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." This will be held during the IAU General Assembly in Prague in August. Also during the IAU there will be a discussion on archiving and data management in astronomy, just following the VO Special Session. Team members also submitted abstracts for both oral and poster presentations.

1.3 Technical (including standards, configuration management)
J. Good of IRSA continues to chair the NVO Technical Working Group, which has focused its efforts on cleaning up the Registries and providing validation and curation tools and procedures, designing and setting up logging infrastructure, and generally monitoring the development of data access protocols and operational infrastructure such as distributed data stores (VOSpace). M. Graham (Caltech) manages the project-wide CVS code repository and is coordinating the assembly of the software library for the Summer School.

The VOEvent standard has advanced to a V1.1 Proposed Recommendation and is undergoing an open review.

The spectrum data model and Simple Spectrum Access Protocol are being finalized, with V1.0 Working Drafts to be completed in term for consideration at the Moscow Interop workshop in September.

1.4 Financial

The NVO project had expenses of $718k in the third quarter of FY2006. This is a major increase from the previous three quarters, and reflects a catch-up in invoicing from several organizations.

As of 31 March 2006 the project is carrying an underrun of $620k, which is $329k less than the previous quarter and less than 4% of the total project budget. A modest amount of funding is expected to remain at the nominal end of the project, and this has been taken into account in preparing a budget for continuing support in FY2007.

1.5 International coordination/collaboration

M. Graham (Caltech) spent two weeks with AstroGrid (UK) personnel, chiefly to refine the specification of VOSpace (see section 8.2), and also to understand the nature of their effort, and the possibilities for collaboration. R. Hanisch visited ESAC near Madrid, Spain in April to discuss progress on the VO registry standards.

The next IVOA Interop workshop will be held September 18-22, 2006 in Moscow, Russia, and will be hosted by the Russian Virtual Observatory project.

2 Science Requirements

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

At the IVOA Interoperability meeting in Victoria, BC in May 2006 significant progress was made toward integration of theoretical astrophysics datasets and programs into the overall VO architecture. More details are given in Section 9.5 (Theory), but in summary the Theory Interest Group has established active liaison with other VO working groups in order to define compatible theory structures in the areas of semantics, data modeling, data formats and analysis protocols. When these are in place, they will have significant impact upon the rest of the VO functionality, because they will enable for the first time general and large-scale comparisons to be made between observations and theoretical simulations and predictions. They will also enable the construction of "virtual telescopes" that can simulate observations of astrophysical phenomena. This capability will be of particular value during the design of new instrumentation and new observing facilities.

2.2 Requirements analysis

No activity in this area this quarter.

2.3 Demonstration definition and review

No activity in this area this quarter.

3 System Architecture

The system architecture is now reasonably stable. However, the definition of the VOSpace interface to manage access to a distributed collection is still under development as an IVOA activity (M. Graham, P. Harrison). Once the specification is complete, multiple implementations are expected. One implementation of the VOSpace interface will provide access to data stored in SRB collections.

3.1 System design, components; relationships to Grid components

The GGF Astronomy working group met at the 17th Global Grid Forum meeting on May 10, 2006 in Tokyo, Japan (R. Moore, N. Walton). The meeting focused on a discussion of grid services that will become available for use in NVO services. An attempt was made to identify the most relevant grid services, and the specific GGF working groups that are leading appropriate standards efforts. The conclusions may change as the Global Grid Forum is evolving into the Open Grid Forum, with inclusion of standards from commercial vendors. After the next meeting, GGF18, in Washington, DC on September 11, 2006, we will have a better understanding of the new grid standards activities. In particular, the GGF standards for data management are in danger of falling two generations behind academic research practice. Grid standards have focused on storage virtualization, while current practice focuses on data virtualization. Storage virtualization provides uniform access mechanisms to multiple storage systems. Data virtualization provides uniform management mechanisms for data stored in multiple storage systems.

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

Work continues 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. The raw image set has been reprocessed to ensure basic astrometric and photometric calibration of the images (D. Monet and S. Levine). This recalibration will improve the alignment of the images supplied by the USNO cutout service. The images will then be integrated into the TeraGrid storage at SDSC (R. Moore). Modifications by Levine to the existing USNO Image archive server are ongoing and will enable it to rapidly serve the entire collection from recently constructed Terabyte class servers.

NVO applications can still take advantage of TeraGrid compute and storage resources. The largest scale NVO application has been the creation of mosaics of the 2MASS image collection. This generated a 20 TB image collection. The output from the first round of production runs for 2MASS was converted from 64-bit to 32-bit and the size of the archive duly went from 20TB to 10T (L. Brieger, J. Good).

Zero-point corrections to track changes in photometry have been incorporated into the Montage code at SDSC, on the TG platform. Testing with the new code has shown that correction factors are correctly incorporated and indicates how increased iterations and slope fitting in the background model near the poles will be necessary to compensate for the heavy coverage in those areas. (This is a matter of supplying appropriate options to the background model routine.) (J. Good, L. Brieger)

Mosaic test results using the new version of Montage can be accessed at http://users.sdsc.edu/~leesa/cgi-bin/list-mosaics.cgi. The jpeg images are the 16x compressed browse products derived from the formal 6-degree mosaic plates. Plate 1 is almost at the south pole, plate 94 is just off the galactic plane, and the other plates are in the Virgo cluster. The jpegs are calculated using a Gaussian log stretch from -1.5σ to the plate maximum, so that low-level signal is accentuated (purposely so in order to identify potential problems).

Preparations are being made now to begin the formal round of production runs for 2MASS. This will generate high quality mosaics of the 2MASS image collection. The mosaics will be registered into the Hyperatlas maintained by R. Williams.

3.3 Digital library integration
The Fedora interface to the Storage Resource Broker data grid is still being evaluated by the Fedora community. The DSpace interface to the SRB data grid is available for public access, with a storage repository provided at SDSC for testing of the integrated system.

3.4 Preservation
The GGF Data Grid Interoperability working group also met in Tokyo. A demonstration was given of the federation of 14 international SRB data grids (R. Moore). A user could log onto one of the data grids, and then transfer to another data grid to read or write files without having to log in again. The SRB data grids were located in Australia, New Zealand, Taiwan, Japan, Chile, the US, the UK, France, Netherlands, and Italy.

A demonstration is planned for GGF18 on the use of SRB data grid federations to build preservation environments. A collection will be migrated between the 14 federated SRB data grids, preserving both the metadata and the data file records. The choice of collection for the demonstration has not been made. One possibility is to pick a set of astronomy images, and extract the metadata from the FITS headers to create as set of preservation metadata. The images and the FITS metadata can then be replicated across each of the 14 federated data grids.

Planning is proceeding, and funding has been secured, for a digital preservation prototype study in partnership with the American Astronomical Society, University of Chicago Press, and the research library community as represented by The Johns Hopkins University and Cornell University libraries. SPARC, the Scholarly Publishing Academic Research Coalition, Microsoft, and the TeraGrid have all committed financial support. A proposal to IMLS, the Institute of Museum and Library Services, has been submitted and appears likely to be approved. R. Hanisch gave presentations about this project at the Library and Information Services in Astronomy (LISA) conference at Harvard and the American Library Association meeting in New Orleans.

4 Registries

In May 2006 at the IVOA Interoperability meeting, the IVOA working group on Resource Registries wrapped up the remaining issues of the Registry Interface (RI) standard. Since then, we have been focused on sharpening the standard document and upgrading our registries to become compliant. In parallel with that, we are upgrading our registry client software so that we can teach and take advantage of the new standard interface at the 2006 NVO Summer School.

4.1 Resource metadata

As we reported last quarter, the small "tiger team" was assembled out of the IVOA Registry working group to update our model for describing services, data collections, and applications, which R. Plante (NCSA) participated in on behalf of the NVO project. The recommendations were presented at the IVOA Interoperability Meeting and subsequently incorporated into the resource metadata model and our standard documents. In particular, Plante has fully documented the new model for describing services within the document, VOResource: an XML Encoding for Resource Metadata, now an IVOA Working Draft.

One long-standing issue that resurfaced out of the tiger team's work is one of "graininess" in the registry—that is, how much detailed resource information should appear in the registry. Efforts centered within the UK AstroGrid project to support workflows have been driving registries toward containing more information. In the NVO, more emphasis has been placed on curation considerations and making sure the information we have is correct, up-to-date, and useful, which calls for less information. We led a discussion of this issue at the Interoperability meeting and made some important headway. We recognized that registries need not all be exactly the same in their features and behavior; in particular, each registry needs to be able to meet the specific needs of its local regional community as well as innovate with new ideas that could eventually propagate to the global web of registries. Thus, we concluded that different registries could maintain different levels of information as needed. To maintain interoperability, however, we concluded that every full registry must harvest every legal record it harvests, and it must export every record that it harvests in its original form (apart from validation annotations). Nevertheless, a registry is not required to support searching on all parts of the record. This lets a registry choose what metadata extensions it will support (and effectively ignore others) as well as what information it attempts to curate.

Progress was made this quarter with integrating the VOEvent infrastructure with the VO registry, and a preliminary sketch of two new extension schemas made. These are for a Publisher of VOEvents, and for a Repository of VOEvents. The former takes data from an event Author and assigns a unique VO identifier, and the latter stores events so that identifiers can be looked up to get previous events. We expect that Publisher and Repository schemas will be formally defined in draft form by the end of 2006.

4.2 Resource metadata schema

R. Plante (NCSA) led the update and standardization for the VOResource metadata schema as part of the overall roadmap for upgrading registries to RI V1.0. The core schema VOResource V1.0 represents the realization of our Resource Metadata standard in XML. The remaining issues of the core schema were presented at the May Interoperability meeting, and following that, *VOResource: an XML Encoding Schema for Resource Metadata, V1.0,* was released as an IVOA Working Draft to the IVOA Document Repository, marking the start of its standardization process.

Plante, in collaboration with the Registry Working, also released updated versions of the main extension schemas, including one focusing on data collections and services, and several for our standard services (e.g. ConeSearch and SIA). He also produced canonical examples as well as a style sheet for translating old records into the new format. The working group established a roadmap in which the core VOResource schema goes through the standardization process during the summer of 2006 and the main extensions follow in the fall.

4.3 Publishing and harvesting protocols

The IVOA registry working group's roadmap to support RI V1.0 includes an update to the harvesting protocol. Still based on the Open Archives Initiative (OAI) standard, the new RI harvesting interface a small change to the previous version as it mainly clarifies what information needs to be available. This new version should address the existing discrepancies between current implementations. G. Greene (STScI) and M. Graham (Graham) are now upgrading the RDBMS and XML-DB registry implementations accordingly. As part of this process, we plan to unify the publishing interfaces as well, drawing on the best elements of all our registries.

Plante updated the implementation of the Registry of Registries (RofR) and deploy the OAI interface for testing during the upgrade to RI V1.0. The RofR is the mechanism for discovering IVOA-compliant registries that can be harvested from. That is, when a new registry comes on-line, its administrator registers it via a web page at the RofR. Other registries that seek to assemble all known resource records will start their harvesting process by first harvesting from the RofR to discover harvestable registries. Plante is now completing the registry registration page that will include a validater that checks to make sure the new registry is compliant with the harvesting part of the RI standard.

4.4 Search protocols

The biggest benefit of the new Registry Interface V1.0 (RI) will be its standard programmatic interface for searching a registry. The biggest challenge has been developing a query mechanism that maps easily to both an RDBMS and XML-DB implementation *and* without stipulating how the data is actually stored in the database. This is important because we have found that registry search performance depends heavily on these two factors. The common search mechanism RI specifies an interface that uses an ADQL Where clause to select resource descriptions, using XPath-like labels to refer to the metadata to search against. This approach allows search against not only standard metadata but extensions as well (to the extent that the registry implementation supports it). We have found that this approach is not perfectly flexible however, and there are some queries that are not possible to support. Partly to address this problem, XQuery is an optional search protocol that can be supported to do arbitrary searches. Once we have upgraded to RI V1.0, the IVOA Registry working group (through our various implementations) plan to experiment with modifications to the query syntax to become more flexible.

An important goal as part of our upgrade is to allow us to teach and use the new interface at the NVO Summer School. To that end, Plante is developing a new client library that makes it easier to search the library from a Java application. It provides a usability layer on top of the service interface that allows users to submit query constraints as a simple string (ADQL/s) and easily pull out important information from the results.

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

Our on-going effort to build validation services has now moved to the registry interfaces themselves. As mentioned in Section 4.3, the RofR will contain a harvest interface verifier. Not only does it validate the OAI interface itself, it checks the validity of the resource records themselves. The software that does the validating, however, is detachable from the RofR. We plan to plug this validation software into the publishing process supported by our registries.

NED supported maintenance of its NVO-registered, including corrections and enhancements to existing services. In particular, the NED Image Archive Table, used in the SIAP NED_NVO Service, is updated now regularly.

On behalf of the Michelson Science Center, IPAC registered the W. M. Keck Observatory Archive (KOA) with the NVO. Public release of the KOA is scheduled for July 17, 2006.

5 Data Models

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

Work continued on the Spectrum model and enhancements to the Spectrum model Java library. At the Victoria Interop, we agreed to split the difference between the old approach and full STC/Characterization compatibility with a fully internal simplified equivalent of STC/Characterization. Work on this design should be complete in July and it will be reflected in new software implementations over the summer.

Discussion in Victoria elaborated the idea of "spectral data associations," which would include SEDs, time resolved spectra, echelle spectra, etc. Although these cases have different semantics we can take advantage of the fact that they have a common structure. This work will follow completion of the Spectrum 1.0 model.

The Spectral Line model is also making progress. J. McDowell reviewed and contributed to a new version of the document (by the European lead authors). Reports on implementations were presented at Victoria.

The VOEvent specification also made significant advances toward being an IVOA Recommendation (R. Williams, Caltech and R. Seaman, NOAO). A lot of effort at the IVOA Interop in May was devoted to ensuring the rapid passage of VOEvent to this state. We hope that this stable, standard packet specification will encourage more network development, and the use of VOEvent for real astronomy. In particular, we hope for VOEvent to be used as the basis for the next generation of the GCN service that reports results from NASA gamma-ray burst satellites. The event lifecycle is now clearly connected to the VO registry, and problems with the Space Time Coordinates (STC) are resolved.

5.2 Low-level (measurement, quantity, uncertainty, relationship)
The Characterization data model is progressing with the release of a V0.9 document. The common spatial, temporal, and spectral axes were called out as separate items to simplify usage and the document is open for discussion in the working group. It was agreed to define two different levels of compliance: level 1 would provide only the Location and Bounds elements while level 2 would guarantee presence of the more sophisticated Support element.

The working group agreed that the mini-models used in the Spectrum model—Curation, DataID, Accuracy, etc.—should be reused in a definitive Observation model and elsewhere, and could be issued as separate standards.

5.3 Descriptors and ontologies (UCDs)
Discussions at the Victoria Interoperability meeting included several updates to the UCD list proposed by NVO members. The UCD IVOA recommendations are now adopted, including the procedures for maintaining the list. A preliminary list of object types and other astronomical concepts is under consideration by the IVOA Semantics group. A controlled vocabulary has been developed by A. Preite Martinez (EuroVO) and R. Williams (Caltech) to describe the types of astrophysical objects and processes that might lead to observable events. The vocabulary is for the section of the VOEvent packet where the Author hypothesizes the nature of the event that caused the observed transient.

5.4 Space-Time and regions
Space-Time Coordinate metadata (STC) development concentrated initially on consolidating the changes that were developed during the second quarter of FY 2006 into version 1.30, to be readied for the IVOA Interoperability meeting, the VOEvent Proposed Recommendation, and the Spectral Data Model. Subsequently, a few changes were necessary to accommodate Characterisation, while at the same time the opportunity was taken to simplify the inheritance structure in the STC schema. Great care has been taken to ensure full, continued compatibility with all existing applications, i.e., none of the modifications will invalidate documents produced by existing applications. In parallel, Rots has worked on a new version of the UML model for STC. As part of the VOEvent-related activities, the selection of solar coordinate systems was revised in consultation with representatives from the solar community.

Following input from the Europeans, J. McDowell agreed to assign data model field identifiers (UTYPEs) to the elements of the STC model. K. McCusker has begun design work on a software library implementing the STC model.

5.5 Standard schema
Discussions at Victoria centered on the usage of the UTYPE mechanism for labeling data model field elements. A document will be prepared for the Moscow meeting in September.

6 Data Access Layer

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

Spectral Data. New versions of the Spectrum data model (J. McDowell, D. Tody, et. al.) and the SSA protocol (D. Tody et. al.) and were prepared and discussed at the May IVOA Interoperability workshop in Victoria.

The Spectrum data model discussions focused primarily on the use of the Characterization and STC data models, both of which saw a lot of work in the months leading up to the Interop. Several options for how to proceed were considered. It was decided to include separate, somewhat simplified versions of these directly in Spectrum for now. A revised version of Spectrum was prepared following the Interop based on these discussions.

Much of the work on the revised SSA protocol specification focused on details of the basic HTTP protocol. Parts of the revised specification were adapted from the OpenGIS Web Mapping Service (WMS) protocol, which is a similar HTTP GET-based protocol which is widely used within the geographic information systems (GIS) community, in particular as the basis of the WorldWind 3D mapping visualization application. Protocol elements were defined to add support for runtime version verification and management, and services with multiple operations. Technical aspects of the protocol, such as allowable formats for numbers, case sensitivity, and so forth were also clarified. A general range-list syntax was defined for query parameters where the parameter value can be a list of parameter values or ranges.

The proposed basic service profile for a DAL service now includes the following required or optional operations:

- **queryData**. Data discovery and metadata retrieval query. In SSA and all other current DAL interfaces the request is currently parameter-based. An optional query language (ADQL) based version will eventually be added, with all other service elements remaining the same (see the discussion of the proposed simple table access protocol below). The request response is a VOTable.

- **getData**. Fetch a dataset identified by an earlier query. Not currently used as at present dataset retrieval is based on direct reference via an access reference URL.

- **stageData**. An optional operation used to asynchronously generate or stage data, or request that the service deliver the data to a given location (such as a designated VOSpace). Support for polling and messaging to determine the status of a request will also be added here.

- **getCapabilities**. Query the service capabilities, e.g., the query parameters supported, whether the service supports cutout or reprojection capabilities, and so forth. The request response is a XML document conforming to the basic VOResource container defined by the registry.

Additional work is required to resolve differences between the *getCapabilities* operation and the similar get metadata operation proposed by the grid and web services basic service profile.

The SSA query parameters were also revised based on earlier discussions. Support was added to permit queries for datasets which have been added or modified recently, to permit remote clients to track changes to data collections. The semantics of the interface were refined to better support data for which certain physical attributes may not be defined, for example theory data or solar data, for which position for example may not be meaningful.
Draft versions of the V1.0 SSA protocol and data model documents, ready to go to the PR stage, are now planned for mid summer 2006 (July).

Time Series Data. A discussion of time series data was held at the May Interop. S. Dalla summarized the work being done within AstroGrid to interface synoptic solar image data and time series data to VO. Their prototype interfaces both types of data via a modified version of SIA; the current version 1.0 of SIA could not be used since it requires that position be specified, but position is undefined for whole disk solar images. We concluded that the next version of SIA, which supports queries based on any of position, time, or spectral bandpass, would work for this purpose without having to define a new interface.

The SSA technology will support tabular time series data and this is what should be used for solar (or any) time series data, but it was felt that it would be inappropriate to use something called "SSA" for time series data. Exactly how we resolve this issue is still to be determined; probably there should be an interface specifically for time series data (which follows anyway from the object oriented nature of the DAL interfaces), but it should be based on the SSA technology and common data model, which supports all tabular spectrophotometric data. To support solar or planetary data will also require more generality in the area of spatial coordinate systems.

Image Data. Image data access was discussed at the May Interop, with a focus on cube (general N-dimensional) image access. The Interop discussions were based on a study done earlier this year (D. Tody et. al.). F. Bonnarel presented an analysis of the Canadian Galactic Plane Survey (CGPS) use-case. M. Dolensky presented a use-case for the ESO SINFONI IFU instrument. S. Gibson (Arecibo), M. Kissler-Patig (SINFONI), R. Taylor (CGPS), M. Marquarding (SGPS), A. Richards, A. Rots, R. Williams, also contributed to these analyses.

It was agreed that support for cube data was a high priority in order to support radio surveys, which routinely produce spectral data cubes, and also to support optical and infrared integral field unit (IFU) instruments. Not all such instruments fit the "image" (regularly-gridded) data model, but most do and the model is considerably simplified as a result. Time cubes are very similar and can probably be supported by the same interface.

The concern was expressed that generalizing SIA to support cube data may complicate the interface to the point where a simple image access interface is no longer possible. It was agreed to split the interface into 2-D and N-D cases if addressing the general case introduces too much complexity. However, we do not yet know if this will be necessary. The current SIA interface, while restricted to 2-D data, is already based on a general N-D image model, and already supports advanced capabilities such as cutouts and projections. It is possible that a single interface could be used for both cases, with the difference being only the capability level of the service.

For example, the basic SIA query required to cutout a 3-D region of an individual spectral data cube might be as simple as this:

```
sia.QueryData (pubid=<cube-ID>,
POS=xx, SIZE=xx,      # spatial region
BANDPASS=xx/yy)       # spectral bandpass region
```

To go one step further and compute a 2-D projection of the subcube one would merely need to add NAXES=2 and possibly a NAXIS parameter specifying the dimensions of the output image. To filter the data along the spectral axis, BANDPASS would be replaced by a range list specifying the spectral regions where data is desired.

The current plan for SIA is to advance the current SIA 1.0 interface to a preliminary recommendation in the fall of 2006. Since SIA 1.0 is already in use, no interface changes would be permitted, however changes to the specification are possible to clarify details of the interface. The new version supporting cube data, an updated query interface similar to SSA, asynchronous data staging, etc., will be either SIA 2.0, or a new interface. Currently there are no plans for a SIA 1.1 as the V1.0 interface is in such wide use that a new version is not desired until enough new functionality is provided to make the upgrade worthwhile.

The issue of simplicity versus capability and sophistication evident in the data cube discussions is becoming increasingly important. Where rapid uptake and widespread adoption by the community is concerned, the simplest interfaces have been most successful. However, as people try to do more serious applications, the limitations of the simple early interfaces become more apparent, and there is a desire for more capability, rigor, and sophistication. Both points of view are important. The solution is probably interfaces that define a simple basic capability, but which also define advanced capabilities that the best implementations should support. Ultimately, easy to use reference implementations, automated service verification, and versioning will be required to achieve both ease of use and sophistication.

Catalog access. Starting with discussions in the architecture session of the May Interop, there has been increasing interest in defining a new Simple Table Access Protocol (STAP) as a DAL interface. This would be associated with, but distinct from, the existing SkyNode technology from the VOQL group. STAP would address the problem of basic access to a single table, including methods for both table data queries and table metadata queries (to discover tables, table fields, etc.) but avoiding the more complex issues of large scale distributed queries and cross-matches, which are best addressed elsewhere, probably as an application service or portal. This might call STAP at the lowest level, but issues such as execution plans for distributed queries, and statistical cross matching algorithms, would be addressed at a higher level. J. Good, R. Williams, A. Szalay, M. Nieto-Santisteban, D. Tody, and others have thus far contributed to this analysis.

While STAP would address the issue of basic access to tables, a data model based approach is also required to provide a uniform interface to astronomical catalogs, such as point or extended source catalogs, astrometric catalogs, and so forth. This would keep the basic STAP interface, but define standard query parameters and table fields based on a data model. P. Osuna and others have produced a first cut at a table container and source catalog data model, which will be reviewed in the September Interop in Moscow.

Since STAP and the related catalog query interface are to be DAL interfaces, they will need to be made consistent with the basic service profile and other conventions adopted for the other DAL interfaces. Both parameter-based and query language (ADQL) based query interfaces are planned. The STAP technology will ultimately provide an optional syntax-based query interface for all the other DAL interfaces as well. The simpler parameter-based interface will continue to be supported.

Current plans call for the existing cone search interface to be submitted as a preliminary recommendation to the IVOA in the September Interop in Moscow. R. Plante has agreed to lead this effort. Development of STAP is still in the early stages and will likely continue into 2007.

Spectral Line Data. An updated spectral line access protocol (SLAP) and associated spectral line data model was prepared and discussed in the May Interop (P. Osuna, M. Lise-Dubernet, J. Salgado, et. al.). Current plans call for this to advance to the preliminary recommendation stage in the fall of 2006. The intention is that SLAP will be consistent with the basic service interface being developed for SSA 1.0 and the other second generation DAL interfaces.

Theory Data. A simple numerical data access protocol (SNAP) has been proposed to provide data access for theoretical simulations (G. Lemson et. al. and the Theory interest group). This is still in the early planning stages. The Theory interest group will work on the design, which will ultimately be advanced as a DAL interface.

Event and Visibility Data. F. Viallefond and others within the international ALMA project are currently working on a data model for raw and calibrated instrumental visibility data from ALMA and other telescopes. The current effort includes both a general data model, and proprietary data access interfaces used within the ALMA project. Depending upon how this develops, we may eventually want to define an IVOA data access interface for radio visibility data based on this data model.

6.2 Data representation (VOTable, etc.)

The issue of how to model and represent datasets was discussed extensively in the weeks leading up to the May Interop, and in the Interop itself (F. Bonnarel, F. Ochsenbein, J. McDowell, D. Tody). Discussion focused on the use of reusable component data models which are logically associated to model complex datasets, on how to represent a component data model as a set of parameters, and on the use of the UTYPE tag to identify the elements of a data model.

It was noted that a data model is an abstraction that should be specified independently of implementation. A reference serialization defined in XML is desirable, but XML is an implementation technology and XML alone does not define a data model. It is common in data analysis applications, for example, to represent a data model as a class in memory, or to store the elements of a data model in various forms in various languages (hash table, parameter set, keyword table, etc.). Not everyone agreed; some felt that it was sufficient to merely provide the XML.

UTYPE usage was discussed extensively. Some felt that UTYPE should be like an XPath expression, with the capability to reference the *contents* of a data instance. Ultimately an agreement was reached that a UTYPE tag only references a field of a data structure (class), independent of an instance or the data in an instance, and that UTYPE strings must be simple predefined strings since they will be "wired into" applications and used via simple string equivalence to explicitly reference data model elements.

Another key issue is how to deal with complex data, that is, data that is modeled by associating simpler elements. This was discussed in a special session at the May Interop. A couple of use-cases of complex data are a radio data cube survey, where a field may consist of a spectral data cube, various 2-D projections or continuum images, and possibly extracted 1-D spectra or source catalogs, or an Echelle spectrum, which is an association of simpler 1-D spectra with common observational metadata, and possibly a precomputed combined spectrum. Two alternative approaches to dealing with complex data are to explicitly model the data by constructing a more complex model for each class of use-case, or implicitly model the data by merely describing the logical associations of simpler, fully defined elements such as individual images or 1-D spectra. This is the traditional trade-off between hierarchical and object relational data models.

The current thinking for SNAP (see above) is that data from theoretical models should be returned in XML (VOTable). An issue is what to do about associating binary data with VOTable. It is possible to include binary data directly within a VOTable, but it is not clear if this is a good idea. Past experience indicates that it is generally preferable to segregate metadata and bulk data, plus XML is naturally text-based and does not coexist easily with binary data. Other approaches, such as associating binary data with a VOTable instance within some type of container, are also being considered.

6.3 Framework (mediators, components)

Work on a scalable framework for data processing and analysis and computational services has focused recently on a prototyping effort using the OpenMPI and OpenRTE software (D. Tody, P. Grosbol (ESO), R. Castain (Los Alamos)). A high level design mapping our planned architecture onto OpenRTE was completed and presented at a face-to-face meeting with Opticon partners in Garching in June. Details on the OpenRTE software, a highly scalable run time environment from the high performance computing community, were delivered at the same meeting (R. Castain). More detailed prototyping work is planned to resume in late 2006.

A first version of the VOClient software was produced in early July, in time to include in the NVO Summer School software release (D. Tody, M. Fitzpatrick). A talk on VOClient was given in the implementations session of the DAL working group at the May Interop, where similar work on the Astro Runtime software from AstroGrid was also presented.

VOClient continues on from the dalclient Java library released last year. VOClient wraps the dalclient library plus related client-side software (a registry interface, a name resolver interface, etc.) to produce a software daemon that runs locally on a client workstation or other hardware such as a small cluster. VOClient implements the client-side functionality required to interface to the VO framework and access remote data and services. Client applications talk to the VOClient daemon via a high performance, stateful, connection-oriented interface. The applications programming interface (API) defined by VOClient is a high level object oriented interface, and hides most of protocol-level details required to talk to VO. Bindings of the VOClient API are available for most popular scripting and compiled languages and application environments, including Python, IRAF, C/C++, Java, FORTRAN, and so forth.

6.4 Data provider/consumer implementations and end-to-end testing
A special session on implementations was included for the first time in the DAL working group meeting at the May Interop in Victoria. Among other things this included talks on the DAL-related software from AstroGrid and NVO, e.g., the AstroGrid Astro-Runtime (J. Taylor, N. Winstanley), and the VOClient software from NVO (D. Tody). K. McCusker described a Java library implementing the SSA Spectrum data model, and F. Bonnarel presented techniques for describing dataset footprints in DAL queries.

CADC and CDS (D. Durand, F. Bonnarel) have extended the prototype DAL interface to the CGPS radio HI data cube survey to include dynamic extraction of 1-D spectra. Discussion continues with Arecibo and others on how to interface radio data cube data to VO.

Discussions continue with STScI (R. Thompson and others) on how best to interface MAST spectral data to VO. The chief issue is how to deal with the complexity of data from multi-segment instruments such as an Echelle. After much discussion the current plan is to interface spectral segments as individual 1-D spectra, and use higher level techniques to associate individual spectra to model more complex observations. It may also be possible to generate and publish a precomputed composite spectrum representing an entire observation. This approach provides a simple, uniform basic interface but may still be able to deal with complex associations without having to physically package all the data in some custom fashion.

CADC (P. Dowler) has built a prototype image indexing service which using caching techniques to provide richer data discovery capabilities than are currently provided by VO. An attempt to implement this on top of SIA was not successful due to insufficient metadata being returned by first-generation SIA services. The experience gained will be reflected in the design of the second-generation SIA; however, getting data providers to actually provide good metadata is likely to continue to be a problem.

A group within AstroGrid in the UK (S. Dalla) has adapted SIA to access solar synoptic imagery and time series data.

7 Query Language

7.1 Low-level: Astronomical Data Query Language

During the third quarter of this year a new version of the ADQL WD V1.04 was presented and discussed at the IVOA Interoperability meeting held in Victoria. The VOQL WG pushed toward ADQL V1.0, and agreed on several simplifications to the language. Specifically, the unit syntax was removed from the ADQL grammar. It was agreed to reduce the number of time data types from 17 to 2 and maintain the Region syntax as it was instead of changes in syntax and complexity.

The main debate of the meeting remained in the syntax and operation of the cross-match function. Up to now, ADQL included a spatial cross-match function based on a maximum likelihood estimator that assesses the probability that two or more objects belonging to different catalogs are in the same position. At the IVOA, there was a lot of discussion about what such a function does and voices were raised in favor of a cross-match function based only on distance. Although a distance-only cross-match appears simpler, in fact it may bring a number of issues. A paper on spatial cross-matching using the maximum likelihood estimator versus a basic distance computation is under preparation with the goal of clarifying and documenting the advantages and disadvantages of each approach.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

The updated version of the SkyNode WD V1.02 specification was also presented and discussed in Victoria.

Raised by the discussions on the XMatch function a similar debate was opened in the SkyNode arena. Currently a basic SkyNode supports ADQL queries against itself only (i.e., supports the basic language, with REGION but not XMATCH). This is clearly a useful capability and database curators should be encouraged to publish at least basic SkyNodes. But should we push for basic SkyNodes to include support for cross-match? And if so, what cross-matching algorithm?

Arguments are also made that the SkyNode specification is really the purview of the Data Access Layer WG, while the ADQL language specification falls to the VOQL WG. The VOQL WG has been dealing with both ADQL and SkyNode since the two specifications basically grew up together. Indeed, the XMATCH controversy noted above indicates that the ADQL and SkyNode specifications must be coordinated.

Another issue arising is the need to provide a Simple Table Access Protocol, STAP, in parallel with the other DAL interfaces. STAP would define the minimum interface that a service willing to publish tabular data should implement. It is now under consideration whether the basic SkyNode should be renamed into STAP or they are different services.

An additional controversial point is the need for methods in the SkyNode interface allowing queries expressed as strings (ADQL/s) instead of as a XML document (ADQL/x) only.

In early August a meeting to discuss all these open questions will be held at JHU.

A new implementation of AXIS-free Java SkyNode implementation is under development.

7.3 High-level: Complex queries

Nothing to report in this area.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)

The VOSpace specification is nearing completion as M. Graham (Caltech) has worked with an international team. This distributed storage system is designed to allow astronomers to store and share data requested from NVO services, or from elsewhere. VOSpace is a SOAP interface to metadata and access to big data storage through high-performance grid protocols. It will be easy to implement using the SRB (Storage Resource Broker), and will also have a "local" implementation, so that it will be as easy to make a VOSpace node as it is to make a web server. The local and federated versions of VOSpace are now split into versions 1.0 and 2.0. The specification document is now well unified to the formal WSDL description.

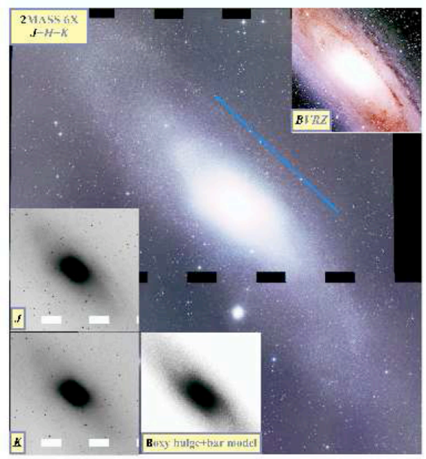
8.2 Grid Services (OGSA)

In this quarter, work on an interoperable authentication framework continues to progress. R. Plante (NCSA) demonstrated at the May IVOA Interoperability Meeting a working user authentication server (UAS) which can be used to log users into the NVO for secure access to restricted resources. In this framework, users have a global VO identity that they can log into through any compliant portal. The portal manages X.509 certificates on behalf of the users, which it uses to gain access to restricted services. Plante and his team at NCSA (M. Freemon and B. Baker) are now working on packaging the server and portal software to put the framework to real use. We will deploy the UAS to allow users to create logins, and we will work with a few portal developers to use the logins. Early target include the NESSSI portal for submitting grid computing jobs and the NOAO Science Archive for accessing proprietary data. We plan to have this fully working and ready to support users by the fall NVO Summer School.

Development is now shifting to supporting a notion of "weak" and "strong" certificates. Weak certificates are intended to be quick and easy to get and allow low-level access to protected resources right away. Some providers—such as those providing access to high-end computing resources—will require tighter security; in particular, they will want greater confidence that the user is who they say they are. Through a collaboration with the Globus team at Argonne, we are working on a identity verification framework that can supports multiple methods of verifying a user's identity.

8.3 Computational resource management
Nothing to report this quarter.

8.4 Virtual data
Beaton et al. (2006; ApJLett; in press, 2006astro.ph..5239B) used the Montage on-request image mosaic service to generate a 2.8×2.8 deg mosaic of M31 in the J, H and Ks bands. Rectification of the sky background to a common level in all three bands brought out the "boxy" shape of the bulge of M31, and analysis confirmed that M31 is a barred spiral like the Milky Way.



8.5 Application and service integration with Grid
The NESSSI project (NVO Extensible Secure Scalable Service Infrastructure) has advanced toward a production deployment, with the prototyping of three services in this quarter. This Caltech project involves R. Williams, M. Graham, and collaborators R. Plante (NCSA) and others at JPL. A Cutout service takes a list of RA, Dec points in the sky, and a number of NVO Image Access Services, and attempts to make a cutout image of each point from each image survey. Currently supported surveys include SDSS, 2MASS, DSS, GALEX, FIRST, and the Palomar-Quest survey. A coaddition service builds deep mosaic images from the synoptic Palomar-Quest survey. A mosaicing service uses the DPOSS survey to create arbitrary mosaic images of the northern sky.

Users can access services either by clicking on a web portal, or by coding against the NESSSI client API. The web portal has advanced in this quarter, with reliable, service-oriented fetching of certificates from the UAS at NCSA, Xforms-based portlet structure so that service developers can quickly make a clickable access for their service.

In each case (click or code) a certificate is needed, either a strong or weak version (see 8.2 above). NESSSI uses the security infrastructure implemented at NCSA. Users can create a "weak" certificate in a few minutes, by filling in a form and proving that they have a valid email address (this is obviously not a strong verification of identity). Or users can present a well-validated "strong" certificate that has come from one of the trusted Certificate Authorities. Each of these can be done either through the web portal or the client API. The request and certificate are presented to the service together—meaning that small requests are honored even if the certificate is weak, but large requests need a strong authentication.

Documentation of NESSSI has also improved this quarter (see http://us-vo.org/nesssi).

9 Applications

9.1 Data location services

R. Plante (NCSA) has coordinated effort for a major revision of all registry services to use a new version of the registry schema. Both the old and new services will run in parallel for some months until registry clients have been updated.

A new version of the NVO DataScope has been released with many new features. This uses a completely independent code-base in Java, AJAX, and JavaScript from the previous Perl based version. Tabbed panels are used to organize the complexity of the responses. Much more information is given on failed requests. Users see an optical image of the region requested and get a summary of the selected files and services. Display of large tables is now feasible: tables are displayed in chunks. Previously very tables required many minutes of rendering time. A few problems have been encountered with the new versions. CfA's DS9 had incorporated access to the previous version internally but cannot emulate the sophisticated JavaScript used in the new version. Thus the old DataScope will not be completely removed. Incompatibilities with OASIS were not completely addressed, but the HEASARC and IRSA have worked to understand where there arise and fix them.

Software has been released by the VOEvent group (M. Graham, A. Drake, Caltech) for subscribing to streams of events. The software uses the popular Jabber protocol for instant messaging, is available in Java and Python, and is available at http://voeventnet.caltech.edu.

9.2 Cross-correlation services

Staff at the HEASARC have developed a simple full SkyNode interface compatible with the OpenSkyQuery portal. Initial tests have been successful, but extensive testing is required. It is anticipated that this interface will be used to introduce students at the Summer School to bringing up SkyNodes within the NVO.

The University of Pittsburgh continued to maintain and augment the WESIX (Web Enabled Source Identification with Cross-Matching) service along with integration with GESTALT (http://frank.phyast.pitt.edu:8080/wsext/). The improvements include

- Detection on one image and measurement on a second
- Detection with weight maps
- Measurement of vector parameters (i.e. Aperture magnitudes)

- Transfer of images is done via MTOM. This removes restrictions on byte array length.

A client interface is anticipated for use in the Summer School.

9.3 Image combination, registration

M. Fitzpatrick continued work on NOAO's WCSFixer web application. Updates were made to improve the stability of the derived solution as well as general enhancements to make the application more robust to problematic images. These changes were installed in the released version of the code on the backend. Additional work was done on new features not released during the reporting period, specifically a new capability for interactively processing lists of images such as the VOTable returned by a SIAP service, and a web-service programmatic interface. These new capabilities were demonstrated at the IVOA meeting along with user-interface enhancements and are expected to be released in the next quarter.

J. Brewer and S. Krughoff have deployed the U. Pittsburgh WCS fixer as a SOAP web service. An example of the WCS fixer client will appear in the NVO Summer School package.

9.4 Visualization tools and services

The new SkyView Java application was delivered to the NVO CVS area as an example of an application that uses the SIAP and Cone search applications to access remote images and catalogs.

9.5 Theory

The IVOA Interoperability meeting in Victoria (May 2006) resulted in significant new steps forward in the development of well-defined protocols for theoretical data and simulations and to their integration into the overall VO structure. In particular, the IVOA Theory Interest Group met several times in order to establish detailed and formal liaison with other IVOA Working Groups. These meetings served to define the requirements for theory datasets in terms of both the metadata and the data products that will ensure compatibility with other VO structures. The principal areas addressed by these meetings are in the areas of semantics, data modeling, and data formats. Discussions have been held with the Semantics, Data Access Layer, and Data Models Working Groups of the IVOA to assist in defining the requirements and next steps. Specific milestones to be achieved in each of these areas over the next six months to one year have been laid out in a roadmap for the Theory Working Group. In addition, milestones have also been defined for the development of a Simple Numerical Analysis Protocol (SNAP) over the next nine months. If all of these milestones are met, then within a year's time significant progress will have been made in providing usable and effective tools for the astronomy community that will enable large scale comparisons between observations and theoretical predictions and simulations. In addition, the creation of "virtual telescopes" to assist in the design of new instruments and observing facilities will be possible in the near future.

9.6 Statistical analysis

Nothing to report this quarter.

9.7 Data mining, outlier identification

Nothing to report this quarter.

9.8 Interfaces to/from legacy software systems

M. Fitzpatrick, in collaboration with D. Tody (NRAO), began work on a “VO Client” interface library as part of the NVO Summer School software distribution and general IRAF legacy-system integration work. The package consists of a multi-threaded Java “VO Client daemon,” preliminarily built on the DAL client code developed for the 2005 NVOSS as well as new code for accessing other VO services, and a C-language API used to access the VO functionality in the daemon via simple IPC messaging. The Java-based daemon is able to easily use existing VO code and standards while having a C-based API makes it possible to create bindings to a common interface for languages traditionally used by legacy code (e.g. IRAF, C or Fortran), or not yet used in the VO (e.g. Ruby, Tcl, Lisp). Demonstration tasks currently exist for a half-dozen different languages including the native IRAF SPP compiled language. Work has begun on integrating this interface with the IRAF ‘CL’ scripting language to allow the development of higher-level applications built on VO data and services. The first public release of the interface will be part of the NVOSS software distribution; release of the IRAF integration work is expected late in the next quarter. NOAO also expects to offer next quarter a public service version of the daemon to which VOClient-enabled applications can connect as a proxy to VO data services.

10 Community Engagement

10.1 Documentation

Plans are being made to publish the proceedings of the NVO Summer School in a more formal volume, and for this volume to be a complete reference for the design and implementation of the NVO framework. An agreement is pending with the Astronomical Society of the Pacific. The volume will include both print and on-line content, and will be inexpensive enough so that graduate students and post-docs can easily afford a personal copy.

10.2 Web site

The Spectrum Services from the JHU group have been added to the NVO web page, as well as pages for the NESSSI project (Section 8.5). The web site continues to be maintained by S. Bunn (Caltech).

A number of NVO services deployed at Caltech have been separated from the web server (us-vo.org) where they have been prototyped, and redeployed on a new, more powerful cluster machine. The domain name for these services (nvo.caltech.edu) has been redirected. For the past two years, both of these have been hosted on the same machine, causing worry that the service component could overwhelm and affect reliability for the web component. Now the web server runs on its own machine with, we hope, much better reliability. The redeployed services are: Carnivore registry, VOEvent database, NESSSI portal, and Hyperatlas services.

10.3 Technical training initiatives
Preparations are now underway for the third NVO Summer School, in Aspen, Colorado. M. Graham (Caltech) has volunteered to coordinate writing a formal proceedings from the Summer School material, and has also been coordinating the software collection from faculty and download to the students. S. Bunn (Caltech) has been organizing many aspects of the meeting; made and distributed posters to all astronomy departments, and sent announcements to many wide mailing lists; also built and run the application forms and refereeing process, and sent letters to applications of their status.

10.4 Advocacy
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." This will be held during the IAU General Assembly in Prague in August. Also during the IAU there will be a discussion on archiving and data management in astronomy, just following the VO Special Session. Team members also submitted abstracts for both oral and poster presentations.

G. Fabbiano (SAO), as SOC chair, concluded the preparations for the workshop on "High spatial and spectral resolution challenges in space astronomy" for the 36th COSPAR Scientific Assembly, Beijing, China, 16-23 July 2006. This workshop consists of three sessions, one of which is entirely dedicated to the discussion of data analysis challenges, with a strong Virtual Observatory flavor. Both Dr. Ohishi (Japan), current chair of the IVOA, and J. McDowell (NVO and IVOA Data Model working group lead) were invited to give talks.

11 Education and Public Outreach
The NVO Education and Outreach Program continues to pursue mechanisms to enable its partners to become more active users of the NVO in their established programs.

11.1 Strategic partnerships
Our submission for a special session at the Seattle AAS has been accepted. This session is entitled "Education with Large Astronomical Surveys" and will comprise presentations by contributors on how the large surveys can be utilized for meaningful science education. Heretofore, education has largely concentrated on small, culled data sets. We envision that the NVO will contribute to such efforts, especially as use of the Sloan Digital Sky Survey, the STScI Digitized Sky Survey and eventually Large Synoptic Survey Telescope data is included, with links to real data.

We have also submitted a request for a special session on "Visualizing the Sky" at the September Astronomical Society of the Pacific meeting. This is a collaboration among several groups interested in using astronomical data (not just catalogs) in visualizing the sky.

11.2 Formal education
C. Christian and K. Borne participated in the in the "Project AstroData" working group at the May 2006 DLESE Data Services Workshop "Using Data in the Classroom." The

purpose was to examine data sources and develop strategies for producing classroom lessons that access astronomy data via NVO tools and services.

11.3 Informal education

We submitted a proposal to the NSF Informal Science Education (ISE) program, to develop a training program involving NVO data and services for museum personnel. The workshop proposed is based on the NVO Summer School, The DLESE Data Services Workshops, and the NASA Explorer Institutes. This proposal, entitled MyVOICE, is in collaboration with the NVO technical team, the NVO EPO group, and a wide variety of informal science educators.

11.4 Outreach and press activities

The NVO EPO website regularly undergoes improvement and touch up. Plans are in progress for developing a wiki and a blog page for NVO EPO.

The feedback from scientists attempting to use the NVO interfaces for their research was obtained through a structured, widely used approach. The observations and interactions with scientists in the STScI/JHU community and some informal feedback from scientists at the AAS were obtained by a “Science Advisory Resource” group at STScI. The report was completed and disseminated to the Executive Committee and to the NVO technical working group.

11.5 Technical development

We interacted with colleagues who are participants with the International group working on metadata definitions for distributed collections of EPO images and videos. This is particularly relevant for incorporating these materials into the NVO and eventually into meta-data driven searches through visualization interfaces.

C. Christian and A. Conti are collaborating with U. Pittsburgh and others on visualization of the sky using SDSS data, DSS data, and other ancillary interesting data such as HST pointings. R. Williams (Caltech) has been loading and supplying the DPOSS image survey data for this purpose (1.2 Tbyte so far). The activity is aimed to produce a public interface that allows individuals to browse and find astronomical objects and information through highly visual interfaces.

The project to create meta-data for Hubble Space Telescope press release images so that this collection can be registered in the NVO continues by augmenting the original 150 images with a new set. The data have been tagged with their corresponding World Coordinate System data. Production of the Registry itself is continuing and should be complete very soon.

Activities by Organization

Caltech–Astronomy Department and Center for Advanced Computational Research (CACR)
Activities at Caltech have focused on time-domain astronomy (VOEvent) and access to major computational resources for the astronomical community (NESSSI).

- In VOEvent, there has been a major thrust to standardize the VOEvent packet definition, to build a prototype network (with J. Bloom, Berkeley, R. Seaman, NOAO, T. Vestrand, LANL, and A. Allen, Exeter UK). The network includes a publisher and repository, software for publishing and subscribing to events, interaction with astronomers who can be event authors, and interaction with keepers of robotic telescopes who may be willing to engage in follow-ups of transients.
- The NESSSI project will allow astronomers to both build and use high-performance computing services that can run on the NSF TeraGrid. Service developers will be able to use the national cyber infrastructure to deploy services, and clients will have access to these services and enormous resources, but without having to go through complex security procedures.
- Other activities include: major part in specification and prototyping of VOSpace, responsibility for maintaining the project web server, a publishing registry (Carnivore), for administrative and faculty involvement in the NVO Summer School

Caltech–Infrared Processing and Analysis Center (IPAC)
J. Good of IRSA continues to chair the NVO Technical Working Group. The work there has been focused on finalizing the initial set of NVO services/protocols and in particular the Registry and SkyNode catalog search mechanisms and formal logging of NVO-enabled service instances. A draft specification of a *Simple Table Access Protocol* has been prepared.

NED supported maintenance of its NVO-registered, including corrections and enhancements to existing services. In particular, the NED Image Archive Table, used in the SIAP NED_NVO Service, is updated now regularly.

On behalf of the Michelson Science Center, IPAC registered the W. M. Keck Observatory Archive (KOA) with the NVO. Public release of the KOA is scheduled for July 17, 2006.

IPAC supported the NVO effort to harvest and compile access statistics. NED and IRSA delivered their Apache logs to JHU/STScI for analysis, and added system key requests to the logs to provide all the information needed by the log harvester.

NED staff member O. Pevunova has worked closely with users of the NED NVO services, especially the "Advanced All Sky by parameters" Service, to correct defects and support development of distributed services. We cite three instances:

- J. Lucey (Durham University, UK) reported defects in the VOTable of Basic Data for two 2MASS objects, out of 13650; these were corrected and released.
- The Multimission Archive at STScI (MAST) NED Search Interface that exploits NED's Advanced All Sky Service. This Service permits a user to search the NASA/IPAC Extragalactic Database (NED) and cross correlate the results of the NED search with observations from the MAST archive. The interface is now available at http://archive.stsci.edu/ned.php. NED worked closely with MAST in correcting defects and advising MAST on how to use the NED services.
- NED user P. S. Chen asked, "How can I put a file containing the list of targets to searching the Galactic extinctions?" In response, NED provided two scripts that accessed the NED NVO services.

The Montage image mosaicing package now includes a module for making images using remote archives that provide SIA interfaces. Currently it has only been tested with 2MASS, SDSS and DSS data but in principle should work with any such data source. This can include ad hoc user datasets since the only real requirement imposed on them is that the metadata be returned in the correct format. IPAC is building a dedicated cluster in support of a generic mosaicing service.

IRSA is in the process of upgrading ROME (used to manage requests to image mosaic service) to the new EJB 3.0 standard. This should rectify some of the operational problems encountered with JBOSS, notably what was presumably a memory leak (JBOSS required restarting every few weeks).

The Visible and Infrared Survey Telescope for Astronomy (VISTA), scheduled for commissioning in 2007, is evaluating Montage as part of its post-pipeline processing to stitch the individual image "paw prints" together to create a fully sampled tile. Montage computes the image geometry to high precision and co-adds the individual images and compensates for variability in sky conditions. J. Good has provided support for this effort, including correcting a defect reported by VISTA that led to singularities in the background rectification process.

B. Berriman gave a presentation on "Applications of Montage" at the Earth Sciences Technology Conference 2006, held in College Park, Maryland, June 26-29. 2006.

Canadian Astronomy Data Centre/Canadian Virtual Observatory

The main VO-related development projects at CADC continue to be (i) the search engine prototype and (ii) re-tooling basic archive infrastructure to support current and future VO requirements.

We have released Octet, a Java application that supports sophisticated querying of the CVO search engine. With Octet, users can perform simple queries and retrieve observations from a variety of data sources; in addition, users can formulate multiple simple observations queries and then find all intersection of these query results through our custom spatial overlap indexes. These queries are very fast and allow all-sky spatial joins of observation metadata.

CADC continues to work on improving our internal infrastructure to support VO and upstream data provider requirements.

Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt)
S. Krughoff, R. Scranton, and A, Connolly have begun collaboration with A. Ptak (JHU) on deploying the X-Ray reduction package XAssist as a WESIX style web service. This is in response to Ptak receiving an NVO small research grant.

J. Brewer, S. Krughoff, and A. Connolly have been looking into XML-RPC as an alternative to SOAP web services. A working SExtractor interface using XML-RPC is in place with a similar interface to WCSFixer. These interfaces will not deprecate the SOAP or web interfaces. This is in response to ongoing problems with web service packages such as .NET and Axis.

Fermi National Accelerator Laboratory (FNAL)
N. Sharma conducted a survey of several modules in the publicly accessible NVO CVS source code repository to determine which, if any, modules could be checked out, installed and utilized by a non-expert. There has not been any requirement that this be possible (other than for modules used by the Summer School, which were not examined), thus the exercise was mainly for information purposes. One module, VOTable, worked properly. The others can be considered as works in progress.

N. Kuroupatkine started development of a new web service interface to the SDSS Data Archive Server, which provides access to flat-file versions (including images) of the SDSS data archive. The current interfaces, which were developed pre-NVO, were custom designed.

High Energy Astrophysics Science Archive Research Center (HEASARC)
The HEASARC continued its leadership role in the preparations for the third NVO Summer School. T. McGlynn leads the faculty team. In collaboration with other NVO personnel particularly at CalTech and STScI, the overall schedule for the summer school was worked out in detail, the software and test plans developed, and the call for applications sent out. A review of the 56 applications was organized by a subcommittee of the faculty and the student body was selected.

As part of these preparations the HEASARC has developed a version of the full SkyNode that is much easier to install than the Java-AXIS based version used previously. A preliminary version of this SkyNode has been delivered to the NVO CVS site and testing is beginning. In addition to being useful to the summer school, this experience may be useful in exploring alternatives to the powerful but cumbersome Web services frameworks that have been developed for Java.

During the past quarter a major new release of the DataScope VO service has been made. This new release uses an entirely new code base in Java and JavaScript. It is considerably faster than the previous versions and easier to use, addressing a number of the recommendations of reviews by DataScope users. It also addresses a complaint that arose by allowing a user to make a DataScope request without having this request made visible to other users.

After release two issues have arisen. While we had intended to remove the old Data-Scope, this service is used within the DS9 image application that cannot use the new version. To retain compatibility with this version, the old version of DataScope will be retained but with minimal maintenance. Incompatibilities with OASIS were also uncovered and some of these have been addressed in collaboration with staff at IRSA.

The HEASARC has continued its efforts to understand the applicability of UCDs to its table holdings. M. Preciado of the HEASARC has begun a systematic review of all tables at the HEASARC, assigning a UCD to every column. This is a manpower intensive operation. Currently about 10% of the highest priority HEASARC tables have been reviewed and have assigned UCDs. The basic HEASARC database metadata have been modified to allow a UCD field for each column. Work is now beginning on providing tools to allow users to select tables based upon UCDs. For example, a user wishing for tables with both optical and X-ray information will be able to search for those tables with columns that contain both em.opt and em.x-ray (or some more specific UCD word).

The HEASARC delivered four packages to the NVO CVS area: the nom.tam Java FITS library, the SkyView Java applications, the new version of DataScope and the simple SkyNode.

The HEASARC continued its participation in the weekly Technical working groups, and participated in the NVO team meetings and discussions.

Johns Hopkins University
A. Szalay has worked with T. Budavári and G. Fekete on developing footprint services and making enhancements to the hierarchical triangular mesh (HTM).

T. Budavári continued his work on the public footprint repository portal, developing the IVOA specifications and services.

M. Nieto-Santisteban attended the IVOA Interoperability meeting in Victoria, where she presented and discussed ADQL 1.04 WD and SkyNode 1.02 WD specifications, and also chaired the VOQL sesions. Nieto-Santisteban, R. Hanisch (STScI), and A. Szalay (JHU) summarized issues regarding ADQL and SkyNodes and prepared a report for distribution among the NVO community. Nieto-Santisteban kept working to support and help users reporting problems with SkyNodes and Open SkyQuery. She also continued working on the parallel implementation of Open SkyQuery as an engine for large-scale access and cross-matching.

G. Fekete defined and implemented better algorithms for computing intersections between HTM mesh and regions. HTM coverage maps (trixels that completely cover a given region) can be created for a variety of purposes, such as an arbitrarily fine approximation of a region, or for creating HTMID ranges that act as a coarse filter for spatial queries. The criteria for what is considered as an adequate coverage map depend on the application. The quality can now be quantified, and the exposed API allows the user to specify a halting condition for the search for the "optimal" coverage map.

A. Thakar collected, filtered, and ingested three months worth of log data from various NVO sites into a web log DB at JHU and created a NVO web log preview page at http://skydev.pha.jhu.edu/nvologs/. The preview contains several snapshots of the web-log data collected for the demo. Later, he also added a SQL query form to the NVO logs page to allow users to submit free-form SQL queries against the web log data.

This was intended as a first step toward an NVO log harvesting system. A common format and delivery method has been agreed upon for all the NVO sites, and Thakar has communicated the expected format to the logging contact at each site. Harvesting of the logs will begin as soon as all the sites have made the necessary changes to provide the web logs in the common format.

J. Raddick continued to maintain and track the usage of the NVO's education and outreach site, www.virtualobservatory.org. He worked with C. Christian (STScI) on an NSF Informal Science Education proposal to run a summer school aimed at museum and planetarium education staff. He also successfully applied for the NVO's third Summer School, to be held this September in Aspen, where he will learn how to use NVO services to create tools for the general public to view and search data through the NVO.

Y. Luo wrote user documentation on using OpenSkyQuery and fixed several bugs in ADQL. He has continued his development of a FITS cutout service, developing a Java web service for FITS cutout.

J. Matias, an undergraduate student at the University of Puerto Rico-Mayagüez, is working at JHU through the Far Ultraviolet Spectroscopic Explorer (FUSE) Research Experience for Undergraduates (REU) program. His summer project has been to create a SkyNode for FUSE data. He has studied SQL and the FUSE file structure, and has received the data to be put into a database.

N. Li worked on the chunk-processing version of OpenSkyQuery. In this system, each SkyNode pipelines its cross-match workload and sends the results to the next node in the execution plan in chunks. This system parallelizes the work, thereby increasing the speed of the cross-match.

Microsoft Research
No activity this quarter.

National Optical Astronomy Observatories (NOAO)
R. Seaman continues his activities associated with the VOEvent WG. The VOEvent V1.1 specification has been released as a proposed IVOA recommendation with a large group of co-authors. A smaller group continues to work on an IVOA Note describing VOEvent's reference transport protocol. R. Seaman attended the first VOEventNet work-

shop in Berkeley in May, and presented a paper entitled "VOEvent and sky transients in the NOAO science archive" at the Observatory Operations: Strategies, Processes and Systems conference of the SPIE later that month. NOAO is now an official, if unfunded, member of VOEventNet. P. Warner attended the IVOA Interop meeting in Victoria and participated in VOEvent sessions. Seaman and Warner continue to develop the VOEvent standard, to build conforming technology, and plan to begin capturing NOAO transient alerts, particularly SNe resulting from the Essence survey, starting with the 2006B observing semester.

C. Miller, E. Fuentes, and D. Gasson , in collaboration with S. Krughoff and A. Connolly (U. Pittsburgh) continued development on the primary release of the NOAO NVO Portal (http://www.nvo.noao.edu). This new version moves development out of the current prototype phase and into a fully functional release. The new portal contains connections the Pittsburgh World Map Server (WMS), which serves up SDSS images. The new portal also talks directly with the WESIX web services running at Pittsburgh. D. Gasson, E. Fuentes, and C. Miller have also produced VORuby: a set of Ruby libraries for working VO services and data. The current library contains tools for VOTable manipulations, SIAP and Cone Services calls. C. Miller participated in NVO Summer School 2006 faculty phone-cons and is scheduled to provide a new set of VO IDL libraries and give 4 talks at the Summer School.

A. Egana, A. Cooke, E. Fuentes, and C. Miller finished the NOAO ADQL image search service and a newly revised SIAP service. These services will be made available to the public in the next Quarter. The ADQL service allows user to search the entire NOAO Science Archive holdings using the power ADQL. The much simpler SIAP service extends the current NOAO NSA R2 version, by allowing a greater number of search and retrieve parameters. A. Egana attended the VO Interop meeting in Calgary.

D. De Young continues to carry out his responsibilities as US-NVO Project Scientist and member of the NVO Executive Committee. This includes continuing emphasis on efforts to integrate the NVO into the US astronomy community in a manner that will allow the NVO to become a part of the essential infrastructure needed to carry out astronomical research. In addition, he is a member of the IVOA Executive Committee and a member of the IVOA Theory Interest Group. De Young was also has been asked to become a member of the Science Advisory Committee for the newly formed EURO-VO, and he attended the first EURO-VO SAC meeting in April 2006.

M. Fitzpatrick and D. De Young attended the Sprint NVO Team meeting in Pasadena. M. Fitzpatrick, P. Warner, A. Egana, and D. DeYoung participated in the IVOA Interop meeting held in Victoria, BC. M. Fitzpatrick and C. Miller participated in several planning telecons for the NVO Summer School.

National Radio Astronomy Observatory (NRAO)
DAL-related work (D.Tody) was especially heavy this past quarter, what with the IVOA Interop in May, followed by a scalable data analysis framework workshop in June, and

preparations for the third NVO summer school. This work is documented in detail in Section 6 of this report.

A meeting was held in Charlottesville in June to discuss NRAO participation in the NVO Facility, as well as continue planning science archive and VO related activities at NRAO. This was organized by N. Radziwill and D. Tody, with management representatives from ALMA and the North American ALMA science/regional center, as well as EVLA and GBT and NRAO management. R. Hanisch attended representing NVO. A goal was to broaden NRAO participation in NVO, and in particular explore how ALMA should be involved. A series of follow-up meetings have since been held. In particular we are evaluating options for partnership with outside groups such as CADC and ESO archiving, to bring more experience and resources to the NRAO efforts.

D. Tody participated in the Arecibo skeptical review, helping to review primarily ALFA-based observing programs proposed for Arecibo, and in particular plans for these projects to publish their work to the VO. Related discussions are underway with various radio data surveys to determine how best to interface radio spectral data cubes to the VO.

Raytheon/ADC (University of Maryland and George Mason University)
George Mason University (GMU) staff K. Borne attended and participated in the NVO project team meeting at Caltech in April 2006. Primary efforts during this reporting period focused on four activity areas. First, Borne contributed to the NVO's education/public outreach (EPO) plans. In particular, Borne contributed to the Stage 2 NVO MyVOICE ("My Virtual Observatory for Informal Collaborative Education in Science") proposal submitted in June 2006 to the NSF ISE (Informal Science Education) program. Second, Borne continued his collaboration with the LSST Project, providing NVO-LSST synergy in several areas, particularly, VOEvent scenarios, NVO-enabled distributed data access, visualizing the sky, and NVO-enabled access to astronomy data for use in the science classroom. Related to latter item, Borne participated in the May 2006 DLESE Data Services Workshop "Using Data in the Classroom"—Borne was a major instigator and contributor to the "Project AstroData" breakout group, which developed a classroom module that could potentially employ NVO access to the data. Third, Borne and UMBC collaborators presented three conference talks related to their work on distributed data mining: (1) a paper presented in April 2006 at the SIAM Scientific Data Mining Workshop: "Distributed Data Mining for Astronomy Catalogs," (2) a paper presented at the May 2006 AGU society meeting on "A Paradigm for Space Science Informatics," and (3) a paper presented at the INTERFACE 2006 meeting on "Astronomy Data Collections of the Future: Massive Data Mining Opportunities." In tandem with the INTERFACE conference, Borne participated in the NASA Data Mining Workshop at the same location (Pasadena, CA). A paper submitted earlier has been refereed and accepted for publication, in the refereed IEEE proceedings of the JPL SMC-IT 2006 conference on the Space Mission Challenges for Information Technology. The above-mentioned papers focus on 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. Finally, Borne contributed NVO expertise to NASA's space science data center in the use of the VO registry for registering unique legacy astronomical data sets held by NASA. In

addition, Borne served as an uncompensated consultant to several NASA VxO (Virtual "Space Science" Observatory) projects, to advise them on NSF's NVO activities related to registries, metadata, distributed data mining, VOSpace, and more.

B. Thomas attended the IVOA Interop Meeting in Victoria CA in May. At the Interop, he discussed various W3C efforts for standardizing distributed work flow and process descriptions. He has been studying and actively gaining experience with the use of OWL-S for work flow to handle High Level Astronomical Data Query. Thomas also updated code on the NVO CVS for VOORML, the VO object-relational markup language. This provides a common method for using XML extracts of a relational database such as a VOCatalog (object-based) and to have elements in these point to columns in the database tables.

E. Shaya has been developing a high level query based on OWL (W3C Web Ontology Language). A query for an astronomical object with multiple and perhaps complicated constraining properties is seen as a subclassing of that astronomical object. An OWL description of this owl:Class with owl:Restrictions on Properties can be passed to any datacenter subscribing to a common ontology, where proper instances of the subclass can be collected. An abstract of this work entitled "OWLviper – A Semantics based Analysis and Query Tool" was submitted for the August IAU meeting in Prague in the session SPS3: The Virtual Observatory in Action: New Science, New Technology, and Next Generation Facilities." Shaya also presented work on an ontology based workflow application at the NVO meeting in Caltech in April.

San Diego Supercomputer Center

SDSC is still working with other groups on the replication of additional sky surveys onto the TeraGrid.

- SDSS. SDSC has also been slowly replicating the SDSS sky survey onto the TeraGrid using an Rsynch server at Fermi Lab. We can speed up the process by moving the data through NCSA and then to SDSC.
- USNO-B. SDSC has provided 1.3 TB of disk for the staged replication of the image collection. However, no images have been replicated yet onto the TeraGrid.

SDSC currently houses 106 Terabytes of image collections for the NVO. This consists of 14 million images, which includes a count of the replicas that are created to minimize possibility of data loss. We plan to reduce the amount of data stored at SDSC to just two copies for each survey. At the moment, the data stored per survey is:

Survey	Collection size (TB)
2MASSdr3	17.4
CTIO-KPNO	0.3
DPOSSdr1	2.8
SDSSdr1	8.0
Hyperatlas	43.7

We are tracking down the difference between the total storage used and the amount per collection.

SDSC is collaborating on the Large Scale Synoptic Survey Telescope project. The LSST team has selected the Storage Resource Broker as the data grid technology that will be used to managed distributed data collections. An approach similar to that used by NOAO for shipping images from telescopes in Chile to Tucson will be explored. Data grids will be installed at the University of Texas (representing the telescope), SDSC (representing the base camp), and NCSA (representing the archive). Images will be pulled from the telescope to the base camp to the archive and processed along the path. NCSA has released a press announcement on the project at http://access.ncsa.uiuc.edu/Releases/07.11.06_Framework.html.

SDSC has released version 3.4.2 of the SRB. This release addresses security vulnerabilities that were found in a security audit conducted by Barton Miller at the University of Wisconsin.

Smithsonian Astrophysical Observatory
J. McDowell and K. McCusker attended the Victoria Interoperability meeting.

A. Rots attended the team meeting in Pasadena, 3-4 April, and the IVO Interoperability meeting in Victoria, BC, 15-19 May.

K. McCusker has been working on the Spectrum data model software implementations.

G. Fabbiano, as SOC chair, concluded the preparations for the workshop on "High spatial and spectral resolution challenges in space astronomy" for the 36th COSPAR Scientific Assembly, Beijing, China, 16-23 July 2006. This workshop consists of three sessions, one of which is entirely dedicated to the discussion of data analysis challenges, with a strong Virtual Observatory flavor. Both Dr. Ohishi (Japan), current chair of the IVOA, and J. McDowell (NVO and IVOA Data Model working group lead) are going to deliver invited talks. Fabbiano also participated, as SOC member, in the planning of the VO and Data Management sessions for the IAU meeting (Prague, August 2006).

Fabbiano also published a paper, discussing the VO, in the concluding remark of the IAU Symposium 230, Populations of high energy sources in galaxies, eds. E.J.A. Meurs and G. Fabbiano, Cambridge University press, 2006 (page 481).

A proposal was submitted to the Harvard Initiative for Innovative Computing (PIs: Fabbiano & Elvis, Co-I Goodman) for the development of a VO portal at the Harvard-Smithsonian Center for Astrophysics. The proposal has successfully passed the pre-selection phase; final notification of the results is expected early in the fall.

At the Chandra X-ray Center, work continues in ensuring up-to-date connectivity between the Chandra Archive and the NASA ADS, so that Chandra archival data and related literature are cross-referenced. Also, work continues in the design and prototype testing of the Level 3 Chandra data processing, aiming at producing the Chandra source catalog database and related data products for user analysis and cross correlation.

Fabbiano attended the IVOA meeting in Victoria, Canada, and discussed the SAO NVO effort for the coming year with R. Hanisch.

Minor fixes have been implemented and deployed for the Chandra SIAP service.

The development of the Chandra Source Catalog is taking considerable care to make the catalog and its public interfaces VO-compatible.

Space Telescope Science Institute
The STScI APT application has completed integration with the CDS Aladin application. This provides HST science proposal tools full access to the VO data archives and services. The APT incorporates the VO SkyNode interfaces to several key data archives and uses the Aladin interface for catalog and image visualization.

The STScI MAST has implemented several prototype SSAP (Simple Spectral Access Protocol) services (HST, FUSE, IUE, WUPPE, EUVE, HUT) all of which are accessible by the NVO registry and Spectral VO tools. With the implementation of these services, R. Thompson has contributed valuable feedback to the ongoing development of a VO SSAP standard.

G. Greene has actively participated in the IVOA Registry Working Group for the recommended proposal of VOResource V1.0 schema and a Registry Interface Web Service Description Language IVOA standard. The NVO Registry at STScI and mirror at JHU are currently upgrading database schema and search services to support the new standard. Registry resource metadata descriptions will include multiple capability elements to facilitate more efficient cross-referencing of data services and curation/service validation level indicators. The registry interface standard will provide uniform exchange between all IVOA registries in support of inter-registry resource harvesting and client-server interface protocols.

G. Greene and S. Lubow have developed a prototype all-sky footprint service for HST WFPC2 and ACS instruments in coordination with JHU. The service incorporated the Space Time Coordinate V1.3 standard and also has contributed the overall development of footprint standard definitions. An IVOA proposal is in preliminary draft to define the footprint specifications.

United States Naval Observatory
Catalogue search and image cutout services at USNO have been upgraded. The VO cone search access to the USNO-B1, USNO-A2 and NOMAD catalogues has been modified to allow a user to request a larger region at the expense of a brighter limiting magnitude. Work by S. Levine is ongoing to upgrade the current web based image cutout service to be compliant with the VO SIA protocol, and to enable the service to access and serve the entire USNO digital plate archive. 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 remains as NVO's primary representative to the IVOA Registry Working Group and primary author of the XML schemas for resource metadata.

Plante is also leading a collaboration to develop an interoperable framework for VO user authentication. In addition to working with NVO partners M. Graham and R. Williams, the collaboration includes contributions from the NSF-funded NMI/Grids Center team at NCSA and Globus developers at the Argonne National Laboratory. We demonstrated a new working User Authentication Server at the IVOA Interoperability Meeting. We are now deploying a new version of the server, so that we can support users at the NVO Summer School. We are also working with Globus developers to build a framework for verifying user identities. In this framework, user certificates will indicate what organizations have verified the identity; services and portals can use this information to decide how much to trust the identity and assign access rights accordingly.

Plante has also contributed to the ADQL and SkyNode development efforts. This has primarily been via a Java library for translating ADQL between string and XML formats.

University of Southern California (USC/ISI)

ISI has continued to provide support for the building of the production-level Montage service. This work involved supporting the continued deployment of software and services at IPAC. Among such services are the Replica Location Services (RLS), Condor, and GridFTP. The installed software also included Pegasus.

Publications and Presentations

"Digital data preservation and curation: a collaboration among libraries, publishers, and the Virtual Observatory," R. Hanisch, Library and Information Services in Astronomy, Harvard University, 19 June 2006.

"Digital data preservation and curation: a collaboration among libraries, publishers, and the Virtual Observatory," R. Hanisch, Scholarly Publishing Academic Resources Coalition/Association of College Research Libraries Forum, American Library Association meeting, New Orleans, 24 June 2006.

"The Virtual Observatory: core capabilities and support for statistical analyses in astronomy," R. Hanisch, Statistical Challenges in Modern Astronomy IV, Penn State University, 13 June 2006.

Virtual Observatory Articles in the Popular and Technical Press

"Montage Software Fuels Astronomical Advances," Quarterly Newsletter of the NASA Computational and Information Sciences and Technology Office (CISTO) (Spring 2006). http://cisto-news.gsfc.nasa.gov/06_spring/Montage.html

Acronyms

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

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

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