Quarterly Report
April-June 2004

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

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

Period covered by this report: 1 April—30 June 2004
Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

Scientific. A number of astronomical catalogs were set up as SkyNodes, enabling on-demand cross-correlation among over a dozen major collections. A prototype VOTable output service for NED queries was developed and made available for testing within the project team. The project web site was redesigned with an emphasis on the end-user community, making it much easier for astronomers to find science applications and tools.

Following the suggestion of the NVO Advisory Committee, we have created a Science Steering Committee to help us set science priorities and assist us in gaining broad community support for the NVO. The first meeting of the NVO SSC will be in July 2004.

A proposal for a special session on VO-enabled science was submitted to the AAS Council, and has been accepted for the January 2005 meeting (San Diego).

Technical. The NVO project hosted the May 2004 IVOA Interoperability Workshop in Cambridge, MA. Our thanks to our SAO team members for their hard work in organizing the meeting. The NVO project manager led the program organizing committee and chaired the plenary sessions. Over 70 people representing VO projects worldwide attended. The next Interop workshop will be held in Pune, India, in September 2004, in association with a regional VO meeting. VO Japan will host the Interop workshop in May 2005, in Kyoto.

A draft specification for the Simple Spectral Access Protocol (SSAP) was developed and presented at the May 2004 IVOA Interop meeting where it was endorsed with only minor modifications. The draft is now being finalized. Minor updates to the Simple Image Access Protocol (SIAP) were also discussed at the Cambridge Interop meeting; updates will be released in V1.1.

International technical collaboration continues to be very active. Technical working groups have been established in six areas: registries, data models, UCDs (Uniform Content Descriptors), VO query language, data access layer, and VOTable. In addition, two new Interest Groups were formed: Applications and Theory, the former led by an NVO team member. The NVO proposed a third Interest Group on data preservation at the IVOA Executive meeting in Glasgow (June 2004); we expect this group to begin this fall.

Substantial progress has been made in the VO Registry, with international agreement reached on the resource metadata definitions and the associated VOResource XML schema (which defines the encoding standard for the resource metadata). The Open Archives Initiative (OAI) standard has been adopted internationally for exchange of metadata records. The NVO project now has two publishing registries and a mirrored full registry (publishing and searchable) containing some 10,000 astronomical resources. The NVO and international collaborators are now investigating techniques for fully automated replication and synchronization of worldwide registries, and tackling the problem of registry metadata curation.

International agreement was also reached on UCD (Uniform Content Descriptors) V1+. UCD 1+ is a more generic system for describing the contents of astronomical tables than the original UCDs.

Substantial progress has been made in defining a standard query language for the VO, the Astronomical Data Query Language (ADQL). ADQL is basically SQL with astronomical extensions. Queries may be encoded either as text strings or as XML documents, allowing for compatibility with XML-based database systems and parsing toolkits. Agreement was reached on OpenSkyQuery and OpenSkyNode, extensions of the SDSS-developed SkyQuery/SkyNode system but without operating-system dependencies. The first OpenSkyNodes were implemented in the past quarter.

NVO team members led the IVOA data model discussions and brought forward proposals for Quantity and Observation data models. The data model effort has been particularly important this year, with a spectral data model forming the basis for the Simple Spectral Access Protocol definition.

Education and Public Outreach. We have developed a partnership with Project Lite at Boston University. Project Lite includes an interactive, web-based tool for teaching undergraduate students about astronomical spectroscopy. Through use of NVO interfaces, Project Lite will have access to hundreds of thousands of spectra rather than just ~50 samples. NVO team members joined a UC-Berkeley based proposal (also including ManyOne) to NASA’s Applied Information Systems Research Program to implement a general public portal to NVO data collections.

Much effort is now going into preparations for the NVO Summer School, to be held in September 2004 at the Aspen Center for Physics. The project successfully applied to NSF and NASA for supplemental funding to underwrite the costs of the Summer School. Forty-eight applications were received for forty available spots. Many of the applicants are professional software developers working at astronomical organizations, though about one-third are graduate students. One-third of the participants are from outside the United States. Twelve NVO team members comprise the Summer School faculty.

Management. We worked with the AVO and AstroGrid projects in developing a position paper on the Virtual Observatory for the Organization for Economic Cooperation and Development's Global Science Forum. The report includes strong endorsement for VO activities.

A draft transition plan describing how the NVO project might move from development into operations, including an estimated budget for ongoing operations, was prepared and submitted to NSF.

The project is in good fiscal health. Spending is at originally planned levels, though with a slow-start backlog continuing to roll forward.

A team meeting was held in Urbana (April 2004), and our summer meeting is planned for Aspen (July 2004).

Activities by WBS

1 Management

1.1 General (planning, reporting, communications, team meetings, etc.)
Regular telecons continue for the NVO Executive Committee (weekly), Metadata Working Group (weekly), and project status reviews (biweekly, with Project Manager and WBS leaders). A spring team meeting was held on April 29-30, hosted by NCSA in Urbana, Illinois. The US NVO project hosted this year's major IVOA Interoperability Workshop, May 24-28, at Harvard (thanks to our SAO colleagues for providing all logistical support!). A summer team meeting will be held on July 29 in conjunction with the first meeting of the NVO Science Steering Committee, July 28, in Aspen, Colorado.

A draft transition plan describing how the NVO project might move from development into operations, including an estimated budget for ongoing operations, was prepared and submitted to NSF.

1.2 Science
Recruitment of members for the NVO Science Steering Committee (SSC) was completed in the past quarter. The membership of the committee is as follows: D. De Young (NOAO, chair), G. Djorgovski (Caltech), M. Donahue (Michigan State University), M. Haynes (Cornell University, after 1 November 2004), F. Hill (National Solar Observatory), P. Pinto (University of Arizona), J. Ulvestad (National Radio Astronomy Observatory), and B. Wilkes (Smithsonian Astrophysical Observatory). The first meeting of the SSC will be held in Aspen on July 28th, immediately preceding the NVO Team Meeting. The purpose of this first meeting will be to familiarize the SSC with the current status of the NVO, to present the committee with the state of NVO thinking about the scientific tools and applications needed to engage the participation by the US astronomical community, and to solicit the advice of the SCC about optimal strategies for achieving this engagement.

1.3 Technical (including standards, configuration management)
R. Williams led a review of the VO architecture for the IVOA Executive. The goal of the review was to make sure that no major components of the VO have been overlooked, and that all interfaces are understood and documented. The architecture review was discussed by the IVOA Executive at its meeting in Glasgow in June, and participants were satisfied with the analysis.

The NVO is making effective use of other NSF-sponsored information technology developments, including

- Grid software (GSI, Pegasus, Chimera, Globus, Condor)
- Data grid software (Storage Resource Broker)
- Portals
- Digital library technology (Open Archives Initiative, DSpace, Fedora)
- Computational resources (TeraGrid)

In addition, there are a number of NVO developments that could quite easily be taken up or adapted by other groups:

- Resource and web services registry
- mySpace, myDB, VOSpace remote data management services
- OpenSkyNode/OpenSkyQuery

A few NVO developments are quite astronomy specific, such as UCDs, the space-time metadata definitions, and data models, though their encodings are utilizing IT standards such as XML schema.

1.4 Financial

The project is in good fiscal health. Spending is at originally planned levels, though with a slow-start backlog continuing to roll forward.

1.5 International coordination/collaboration

NVO participation in IVOA coordination and collaboration remains high. Members of the Executive Committee attended the June meeting of the IVOA Executive in Glasgow, Scotland (held in conjunction with the SPIE conference on Astronomical Telescopes and Instrumentation). NVO senior personnel continue to lead IVOA Working Groups (J. McDowell, Data Models; D. Tody, Data Access Layer; R. Williams, Unified Content Descriptors; R. Hanisch, Document Standards), and T. McGlynn convened the first meeting of the IVOA Applications Interest Group during the Interoperability Workshop in Boston (May). International cooperation is strong and growing, with renewals or increases in funding for several VO partners and the addition of a 15th VO organization, the Spanish VO, to the IVOA.

R. Hanisch collaborated with P. Quinn (AVO) and A. Lawrence (AstroGrid) on a white paper for the OECD Global Science Forum entitled "The Management, Storage, and Utilization of Astronomical Data in the 21st Century." This white paper was strongly endorsed, and the OECD GSF issued recommendations that are very supportive of the international VO initiatives:

> *Findings.* The Virtual Observatory concept is a bold community-led response to the challenges the astronomical community faces in data management and storage. Impressive progress has been made and the momentum of the International Virtual Observatory Alliance will ensure sustained progress, provided the agency level support and funding is available.
>
> *Recommendations.* Agencies and governments should consider adopting the IAU resolutions as the basis for progress in this field. Agencies should recognize that this is an important long-term issue and should coordinate plans, provide adequate funding on a long-term basis, and support development and maintenance of the needed infrastructure. Agencies should encourage the broadening of the existing VO collaboration into a fully representative global activity. New projects and facilities must take the date management, storage, maintenance, and dissemination

into account at the earliest planning stages, consulting potential users in the process.

2 Science Requirements

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

Work on usage scenarios is slated to occur at the summer NVO team meeting, in conjunction with the newly constituted Science Steering Committee.

Planning continued on the integration of the theoretical astrophysics community into the US NVO and into international VO programs. At the Interoperability Workshop in Boston a special session was held by the IVOA Theory Interest Group. This entailed presentations of "Theory VO" activities in several countries (France, Germany, India, Italy, UK, US), followed by a discussion of general strategies and next steps. It was decided that leads in TVO activities in each country should act to coordinate their respective VO theory efforts, and at the same time the group will continue to develop common interfaces and standards for incorporating theory datasets into the VO. In the US, efforts are now underway to obtain participation from several theory groups in a common TVO web page that will contain links to results that each group wishes to post. This is seen as a modest first step in generating interest in the NVO by US theorists, to be followed by their active participation in this project.

2.2 Requirements analysis

The July 2004 team meeting will be focused on setting requirements for the coming year's science applications development.

2.3 Demonstration definition and review

The July 2004 team meeting will be the venue for defining new science applications and establishing implementation teams.

3 System Architecture

The components for the system architecture are being designed to support researchers in the astrophysics community and simplify their access to sky surveys. Of equal importance is the development of new capabilities that will enable research projects that are currently not feasible. This requires architecture components for which a consensus has not yet been achieved. The components include digital libraries, workflow systems, and bulk processing systems. Efforts are now underway in all three areas.

The Astronomy Working Group of the Global Grid did not meet at the GGF meeting held in Hawaii. For the next GGF meeting to be held in Brussels, the AWG will conduct an initial data grid assessment.

3.1 System design, components; relationships to Grid components

The document summarizing the IVO architecture (developed by R. Williams et al.) has been accepted by the International Virtual Observatory Alliance. The consensus approves an architecture based on three types of services:

- Web services (http)
- Stateful services (WSDL)
- Grid Services (evolving from WSDL to WSRF).

The overall architecture is summarized in the diagram in Figure 1.

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

3.2.1 Processing

The TeraGrid is now in limited production. In addition, a data intensive computing platform is available at SDSC, called DataStar. The TeraGrid provides 13 Tflops of computing power distributed between four sites (NCSA, SDSC, Caltech, and ANL). The DataStar is a peak 10-Tflops system with 4 TBs of memory and high-speed fiber channel connections to a 500-TB SAN.

In support of the NVO, L. Brieger at SDSC installed, ran, and verified the latest versions of Montage on DataStar. A script was produced to generate 1734 Montage mosaics that fill the entire sky for the 2MASS sky survey. Montage has also been installed on the TeraGrid cluster. The current status is that the large runs will be done from Caltech. Other large mosaicing runs are being done with the Atlasmaker code by R. Williams, with the purpose of federating many astronomical sky surveys.

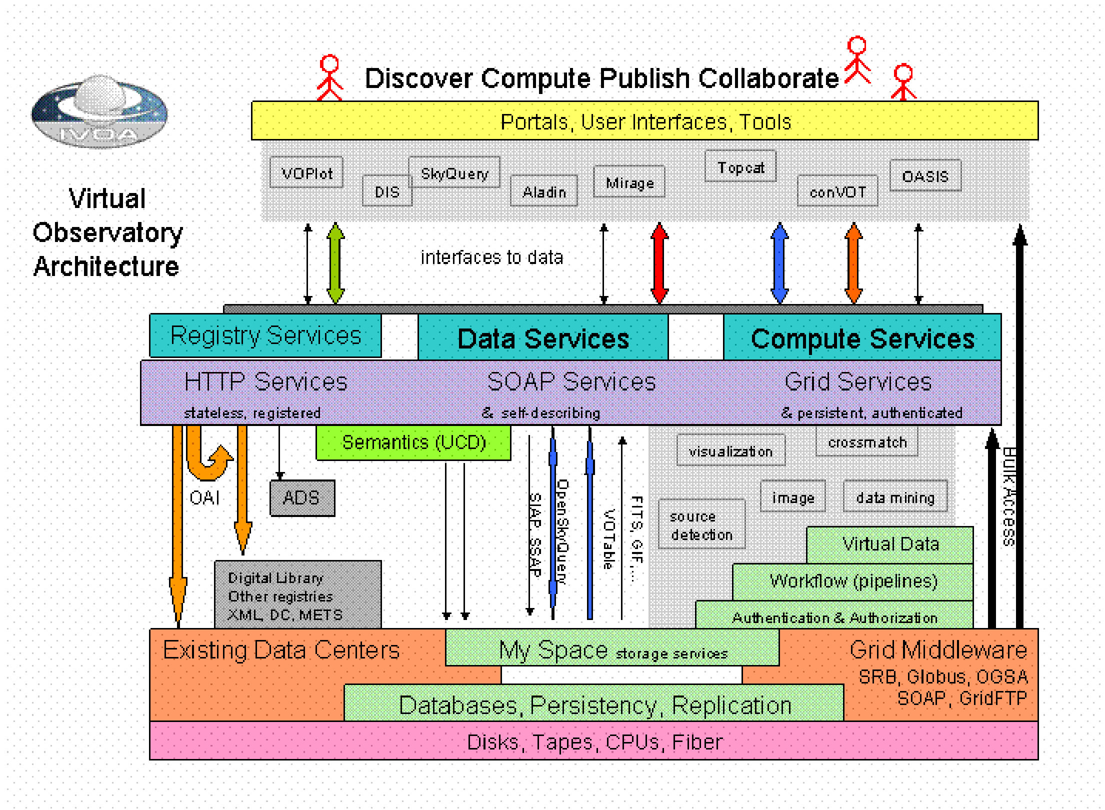

Figure 1. Overview of the Virtual Observatory architecture. Different coloured vertical arrows represent the different service types and XML formats by which these portals interface to the IVOA-compliant services. In the IVOA architecture, we have divided the available services into three broad classes:

- Data Services, for relatively simple services that provide access to data,
- Compute Services, where the emphasis is on computation and federation of data, and
- Registry Services, to allow services and other entities to be published and discovered.

These services are implemented at various levels of sophistication, from a stateless, text-based request-response, up to an authenticated, self-describing service that uses high-performance computing to build a structured response from a structured request. In the VO, it is intended that services can be used not just individually, but also concatenated in a distributed *workflow*, where the output of one is the input of another.

3.2.2 Bulk data storage

Attempts to replicate the MACHO and USNO-B sky surveys onto the TeraGrid are still waiting on interactions with the originating institutions. The issues include management of access, management of data transfer, and management of the FITS header associated with each image.

In the interim, the SDSS DR2 release is being replicated onto the TeraGrid in support of large-scale analyses. Metadata services (SIAP) at Johns Hopkins are being upgraded to expose these grid-replicas to the VO community.

3.2.3 Workflow systems

A system architecture component that is needed to support collection building, image processing, and bulk analyses is a workflow or dataflow system. The typical example is the movement of multiple files between multiple services under process control

This topic was raised at the IVOA meeting, but was viewed as a future requirement. An effort to track standards within workflow environments was encouraged as part of the Astronomy Working Group of the Global Grid Forum.

Within the NVO, multiple workflow environments are available, ranging from ROME (for tracking the processing of tasks), to Chimera/Pegasus for managing workflow within the grid, to Matrix for supporting interactions with collections while managing state information associated with the data flow.

New approaches to workflow build on the emerging service infrastructure of NVO. The NSF/GRIST project is creating libraries of derivative data products, which can be created on demand, through batch jobs, or through grid scavenging. An upcoming international workshop at Caltech (Service Composition for Data Exploration in the Virtual Observatory) will resolve different approaches.

ROME (Request Object Management Environment) is a job control system to enable long-running requests to VO services. It is currently undergoing formal testing, covering load-balancing, failure recovery, loading tests, and user feedback. ROME is being used for building 2MASS images for each of the sources in the MAST Scrapbook project.

ISI is executing the Montage image service to create custom mosaics on demand within a computational grid through use of the Pegasus/Chimera workflow system. ISI has run Montage on the TeraGrid using the Pegasus framework. A new Montage pipeline was developed to exploit the full parallelism inherent in montage processing. In addition, several improvements were done over the initial version.

- Generation of an abstract DAG to describe the workflow.
- The re-projection of the images is done in parallel.
- The overlap between the images is decided before starting the projections.
- Unique names are provided for the images files.

The new montage pipeline is depicted in Figure 2.

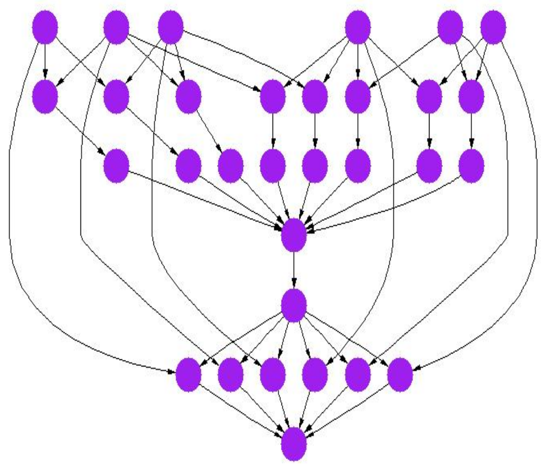

Figure 2. The new Montage pipeline.

Montage has been run on single and multiple condor pools and on the TeraGrid. The runs have achieved significant speedup compared to the sequential version of montage even in the presence of scheduling overheads. ISI has been running large region mosaics. As a result of the code stress testing, several bugs and weakness were found in the Montage application code, the Pegasus code and related software. For instance, the run time memory requirement in mAdd and other modules did not become obvious until large region mosaicing was attempted. Also, the problem of using predetermined parameter input file for mAdd did not appear until now. This has lead to the addition of another step in the abstract DAG that is yet to be added into mDAG service. The critical load problem on a GridFTP server used for data movement lead ISI to update the data transfer client in Pegasus. In addition, several feature improvements are currently pending to satisfy large Montage workflow. The condor/condor-G has been updated due to the problems ISI discovered during the stress test, i.e., the periodic release/held problem and the postscript problem in the rescue DAG. The large-scale testing has also highlighted problems in the Replica Location Service which has lead to eventual bug fixes.

3.2.4 Authentication

The current focus is on access. The major design issue is how to support authentication across multiple participating sites. The IVOA is discussing whether accounts will be needed for access to controlled resources, or whether identification as a member of a community will be sufficient. The resolution will depend upon how much control is required for access, as opposed to controls applied for management. The access controls only require identification of remote users for non-public data, and identification as a member of a community may suffice. Management controls require identification of individuals allowed to add material and update material (both data and metadata). An identification system that is compatible with the Grid Security Infrastructure will be sought, probably based on X.509 certificates. The first challenge here is to identify suitable certificate authorities or empower communities to create their own CA. This will be followed by a much greater effort to encourage professional astronomers to use the technology. Within the NVO, GSI authentication is already used for access to TeraGrid resources.

3.3 Digital library integration

Large sky surveys are based on catalogs and portals that provide most of the functionality of digital libraries (collection organization, search, access, and manipulation). For researchers, additional technology is needed to enable small collections, in which images and results can be published for use by the broader community. The approaches that are being tried within the NVO are based on technology from the grid community and the digital library community.

3.3.1 Grid libraries

The Globus project at ISI is developing a Metadata Catalog Service for registering metadata about user files. The system will provide interfaces based on the OGSA-DAI framework (http://www.ogsadai.org.uk). Details about the current version of MCS are found at http://www.isi.edu/~deelman/MCS.

The data grid project at SDSC uses the Storage Resource Broker technology to support distributed collections. The access interfaces include Open Archives Initiative metadata harvesting protocol, WSDL services for metadata and data management, and web browser interfaces to support search. Both projects are designed to scale to millions of digital entities. The SRB technology supports federation of independent data grids, making it possible for a major sky survey to implement their own data grid, and then at a later time when the data is publicly available, publish their data grid into the NVO data grid. SRB version 3.2 was released on July 2, 2004 (http://www.npaci.edu/DICE/SRB).

3.3.2 Digital libraries

A second approach uses the standards implemented within the digital library community. The DSpace technology from MIT focuses on preservation, and the Fedora technology from Cornell focuses on manipulation. Both approaches are being evaluated within the NVO. Johns Hopkins is collaborating with Fedora to build mySpace for storage of small collection. SDSC is collaborating with MIT to integrate the DSpace technology as an access mechanism on top of the SRB data grid. Both projects will provide web-based interfaces for accessing collections that are deposited by researchers. The DSpace/SRB integration will also provide the ability to federate multiple independent digital libraries, making it possible for an institution or researcher to establish their own collection, and then federate that collection with a larger publication environment.

This points towards two possible approaches to support for mySpace collections:

- NVO central repository. This would be similar to the NSF National Science Digital Library, with a single repository pointing to all material, backed up by a persistent archive for preservation.
- NVO federation. This is similar to the approach being followed by the Worldwide Universities Network for the sharing of research data between institutions. The WUN serves as the policy management system for federating multiple independent digital libraries. NVO can provide a similar role, serving as the institution that defines and

manages the federation policies required for publication of researcher data into the NVO name space.

Equivalent metadata catalog services are incorporated in sky survey catalogs, the SRB data grid, NVO portals, and the SRB federation environment. At some point, NVO will need to do a comparison between the approaches to analyze which system provides the best support for astronomy collections. The comparison needs to address federation of name spaces between independent data grids, support for preservation environments, and support for NVO portals.

4 Registries

The major goal of activity in this area over the last quarter was to advance the interoperable registry standards to a level that will be useful for the NVO Summer School. In particular, we worked to refine the metadata schema standard to ease development, and we made significant progress drafting the standard programmatic interface to registries. The Registry Interface specification (being led by K. Benson of the AstroGrid project) incorporates both searching and harvesting; G. Greene (STScI), W. O'Mullane (JHU), R. Plante (NCSA), and M. Graham (Caltech) have contributed significantly to this specification. An important checkpoint for these standards development came in May at the semi-annual IVOA Interoperability Meeting in Boston where the IVOA Registry Working Group (RWG) reviewed and formed consensus on various proposals.

4.1 Resource metadata

R. Plante presented the IVOA Identifiers Working Draft version 1.1 to the IVOA Working Group at the May Interoperability meeting at which time it was elevated to a Proposed Recommendation. This deprecates version 1.0 of this standard. After a comment period, this will be considered for elevation to full Recommendation status by late summer 2004.

4.2 Resource metadata schema

In April, Plante hosted a "videocon workshop" to consider revisions to the VOResource XML schemas which was attended by four IVOA projects from seven institutions (from NVO: Caltech, NCSA, NASA Goddard, and STScI). The purpose of the workshop was to address issues raised in the IVOA Note entitled "Lessons Learned Using the VOResource XML Schemas in the NVO" (Plante et al. 2004, http://www.ivoa.net/Documents/latest/RMExp.html) which has made the schemas difficult to use. This workshop generated by consensus a short list of recommended changes. Plante generated updated versions of the schemas—VOResource version 0.10—along with set of examples for a rapid period of testing so that they could be presented and discussed at the May Interoperability Meeting. During this latter meeting, the IVOA Registry Working Group agreed to adopt the recommended changes as Working Drafts and the basis for the January 2005 demos. Since then, the revised schemas have been undergoing heavy testing at NVO registry sites. Initial tests using the .NET framework and XForms have revealed only minor issues. Plante and Ramon Williamson (NCSA) have generated XSL style sheets for updating version 0.9 documents to version 0.10. With their official posting to IVOA web site, registry maintainers will begin upgrading their systems accordingly.

A. Rots released an updated working draft of the Space-Time Coordinates data model and associated XML schemas (see Data Models). In anticipation of its incorporation into the resource metadata schema, Rots has been examining how we make describing common coordinate systems simpler by referring to standard definitions. After some prototyping, we decided to adopt the W3C XInclude standard for referencing XML in external documents.

Now that several XML schemas are becoming more mature and in greater use within the IVOA community, long-term standardization issues are becoming more apparent. One issue that has come up a number of time regards the extent to which we attempt to reuse standards—particularly our own—and understanding when existing standards are a good match and when they aren't. One particularly difficult example is whether and how we might use components of the VOTable standard within the resource metadata schemas to describe tables. While at the moment we have chosen not to do this, we have to monitor this practice and prevent the proliferation of multiple standards that do the same or similar things.

4.3 Publishing and harvesting protocols

Building on NVO's experience using the Protocol for Metadata Harvesting (PMH) from the Open Archives Initiative (OAI; http://www.openarchives.org), R. Plante developed a proposal for adopting this digital library standard across the IVOA as part of the standard Registry Interface specification. This proposal included a definition of a Web Service version of the protocol that built on the prototyping work of C. Cowart (SDSC), Plante, W. O'Mullane (JHU), and G. Greene (STScI). The Registry Working Group voted to adopt the current HTTP version of OAI as the mandatory harvesting interface and the Web Service version as an additional, optional interface.

Plante has been working closely with K. Benson (AstroGrid), the lead editor for the IVOA Registry Interface Specification, to get that specification to a useable state. Work on the first prototype implementations of the standard will begin this summer.

M. Graham (Caltech) has released updated style sheets that convert XML schemas into XForms documents that can be used to generate fill-in forms that providers can fill in to create VOResource documents and publish descriptions of their resources. This is the main mechanism for registering resources with the Carnivore Registry system. The purpose of the system is to convert arbitrary XML schema to human-usable fill-in forms to create instances of the schema. Thus Carnivore is well-placed to respond to changes in registry schemas, as well as having flexibility for novel uses of the registry.

The full searchable registry at STScI (Figure 3) now does automatic re-harvests from the publishing registries based on the harvesting date. We now support the OAI mechanism for marking resources as deleted and have successfully used this to "re-register" resources under a new identifier. A number of improvements have also been made to the interactive web interface to make it easier to use.

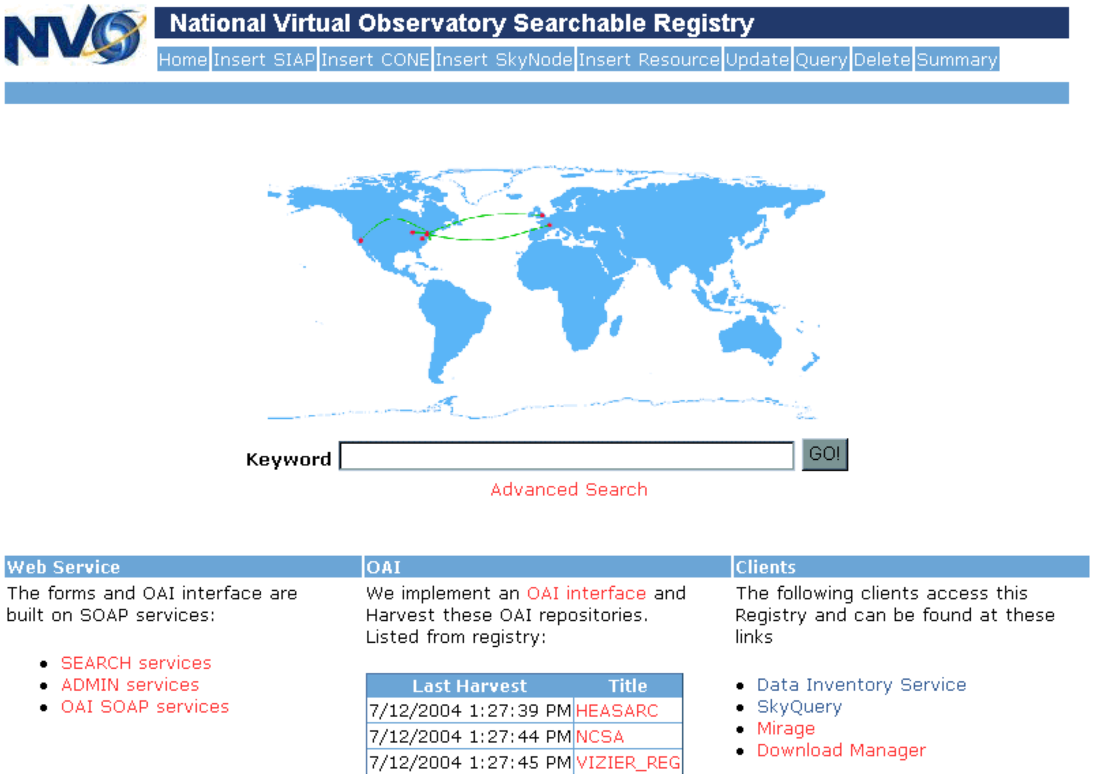

Figure 3. The interface to the full searchable registry at STScI.

4.4 Search protocols

To support generalized registry queries, the Registry Interface specification combines two complementary standards. The first is the emerging Astronomical Data Query Language (ADQL) standard (a working draft from the VOQL working group): registry queries will be expressed as a restricted form of ADQL that contains only a "where" clause. Column names are expressed as a simple XPath (containing no "[]" structures; e.g. "curation/publisher") that points to a particular entity in the VOResource schema. This restricted form of XPath makes it straightforward to support queries against any entity in a VOResource record by either XML or relational databases. Early implementations will begin this summer.

The OpenSkyPortal (which allows users to execute database queries across distributed databases) is the latest application to actively use the NVO registries. It uses the current non-standard interface to locate available SkyNodes, of which there are now thirteen.

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

A proper and permanent URL has been selected for the searchable registry namely: http://nvo.stsci.edu/voregistry. The STScI (nvo.stsci.ed/voregistry) and JHU (voservices.net/registry) registries now replicate between each other on a regular basis, thus making them a proper fail over for each other.

The discussion of registry maintenance was elevated to the IVOA level when R. Hanisch discussed at the May Interoperability the issues we've experienced thus far. Further work in this area is expected in the next year.

5 Data Models

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

Extensive e-mail discussions on the Observation Model led to the decision to issue the model as an IVOA Note and prepare the component parts for the IVOA Recommendation route. The Observation model describes general metadata for observations, divided between the description of the actual data values, the Characterization, which describes the context of the data in terms of physical variables free of instrument signature (e.g., the sky area coverage, spectral resolution etc.), and the Provenance, which describes the details of the observation and data analysis process. It was agreed that the Characterization is the highest priority element needed by other VO activities. Work by UK and Australian collaborators on radio astronomy observations has verified that the existing Observation model can be extended to cover their data.

A new draft of the Simple Spectral Access SED data model, developed by J. McDowell and D. Tody, was issued prior to the Boston meeting and will be put to the recommendation process. The document defines an SED (Spectral Energy Distribution) model as an assembly of segments consisting of individual spectra, time series, or photometry measurements, and handles two-sided flux errors, upper limits, instrument apertures, and other critical details. It also proposes XML-schema, VOTABLE, and FITS serializations for the model.

The near-term plan is to push the SED and Space-Time Coordinates models to recommendation level this summer, followed by Observation Characterization. This will support demos in the January timeframe.

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

The Quantity model progressed with extensive email discussions on the IVOA list. At the Boston Interop meeting it was decided to more clearly separate the basic Quantity ideas from the more sophisticated multiple-axes representations desired by some working group members. A document on Mappings was also sent to the mailing list but has had relatively little review to date.

5.3 Descriptors and ontologies (UCDs)

The latest version of the UCD1+ scheme seems to have broad support in the VO community. UCD1+ elements have been defined for existing UCD1 elements, and have been proposed for the elements of the spectral data model.

5.4 Space-Time and regions

The revised Space-Time Coordinates document now includes a fully general coordinate description and is accompanied by a full XML schema implementation. It was decided at the Boston Interop meeting to put forward the current version of the STC document for IVOA Recommendation as version 1.0, with the understanding that a later version 2.0 might include more commonality with the Quantity model. Given concerns about the complexity of the model, the STC author (A. Rots) was given actions to define standard instances for common cases. With the action, the Space-Time Coordinate schemas were thoroughly revised to (schema) version 2, which will correspond to document version 1:

- Transition to substitution groups
- Improved extensibility for incorporating new coordinate reference frames
- Inclusion of standard solar, lunar, and planetary frames
- Adoption of XInclude will make life much easier for users and developers
- Various features requested at the IVOA meeting were incorporated:
    - Offset positions
    - Region specification through sky indexing schemes
    - More unified astronTimeType
- A proper document has been prepared
- The UML diagram has been updated
- The document will include a set of example instances

At the time of this writing (early July 2004) we are at version 0.6 of the document and the schemas and document have been circulated among the NVO metadata group. By the middle of July solar coordinate systems will have be incorporated, the documentation expanded, and a fair number of examples added, including examples of XInclude usage. At that time the document and schemas will be submitted to the IVOA, to be followed by a library of XInclude coordinate system instances.

The remaining question is whether use of UCDs should be prescribed for coordinate names.

5.5 Standard schema

Work is underway to fully define the data model representation in VOTable, and in particular the syntax of the UTYPE element that will help encode the schema.

6 Data Access Layer

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

6.1.1 Catalog access

A goal for the next year is to add general query language support to all the DAL services. The first step will be to add a DAL version of the SkyNode catalog access service to replace the older cone search. More generally, we need to integrate query language capabilities based on ADQL into all the DAL services. This will take the form of extending the query interface to provide both a simple parameter (forms) based interface as at present, plus a more general ADQL interface to provide syntactic queries as well as support for advanced VO functionality such as regions. Only the query part of a DAL service will be affected. The query response will be the same for both parametric and syntactic queries. The *getData* method for each service, used to perform data model-based data access, will be unaffected.

For example, integrating ADQL into SSA would mean adding a second ADQL-based query method. The simple parameter-based query would continue to be available, and the query response would be the same in both cases. Data access would be unaffected, returning a SED, spectrum, or time series object conformant to the SSA data model, in any of the defined representations.

A potential problem with using a query language to access data via DAL is support for virtual data. Services which can return virtual data can potentially return an almost infinite number of distinct data objects, hence a conventional database table cannot be used to statically describe all possible virtual data objects. A possible solution to this problem has been proposed by Y. Shirasaki. Rather than query an actual database table the query would be against a "virtual table", allowing the service to match data objects to the client query as in the parameter-based query approach.

Our original DAL service architecture from the IVOA Cambridge meeting in 2003 defined both a set of type-specific interfaces (for catalogs, images, spectra, etc.) as well as a general "dataset" query that could be used to query for any type of data. ADQL is what is needed to implement the general dataset query. The query response would return only general dataset metadata, and would include an indication of the type of data available.

6.1.2 Image access

A new version of SIA is required to update the interface to reflect developments in the underlying VO technology over the past year, and to add new functionality to address issues identified by the first round of client applications using SIA. Plans for the new version were reviewed in the DAL sessions at the IVOA workshop in Boston in late May. The technology upgrades affect mainly the query response table (QRT). The changes under consideration include:

- Upgrade the query interface to add UCD1+ UCDs for all values in the response VOTable.
- Add UTYPE tags for all SIA interface elements.
- Revise the image metadata to use the emerging data model standards as for SSA. This includes uniform data characterization, coverage, dataset identification, and so forth.

Functionality enhancements include:

- Logical name proposal (R. Williams). This would provide a way to say that a subset of the images in the QRT are in some sense the "same" image, e.g., differing only in the file format or the spectral band.
- Ranking proposal (A. Rots). This would add a means to grade candidate images in the query response with the "best match" datasets receiving the highest grade or ranking.

This new version of SIA is required to upgrade the interface to reflect current standards, particularly in the data model and VOTable areas, and to provide compatibility with SSA.

The next major release, probably some time in 2005, will add support for ADQL-based queries.

6.1.3 Spectrum access

Considerable work on the Simple Spectral Access (SSA) data model was done by J. McDowell and D. Tody in this past quarter. A draft SSA query interface was developed by D. Tody and others. Both of these documents were discussed extensively in the DAL sessions of the IVOA workshop in Boston. While good progress has been made, it has been difficult to finalize the SSA data model as many components of the model, e.g., for data characterization, dataset identification, coverage, etc., are not specific to SSA and require broader agreement.

The scope of SSA is spectrophotometric data in tabular form: 1D spectra, time series, and SEDs. It does not include spectral image cubes, long slit spectra, or synoptic imagery, which will be handled by the image service since they are gridded data. Much of the work being done on the SSA data model will however carry over to image data.
The components of SSA are:

- The SSA data model.
- The query interface, including the query and the response table.
- Data access, including standard representations of spectral and time series data in various formats.

All aspects of SSA were discussed in Boston and working draft versions of the interface documents are in preparation. Much of this work is being done in collaboration with the other working groups, e.g., the SSA data model is a joint effort of the DM and DAL groups, we are working with the VOTable WG on data representation issues, and with VOQL on queries.

6.1.4 Time series

The latest version of the SSA data model is now based on representing a SED as a series of segments, each of which implements a general spectrophotometric component data model. This model may be subclassed as a Spectrum object (1D spectrum), where time is constant, or as a TimeSeries object, where the spectral coordinate is constant.

Although SSA includes support for time series we have limited experience with actual time series data. It is not yet clear if a general model will suffice for actual time series data, which can be idiosyncratic. Our strategy is to include provisional support in SSA and interface some actual time series data on a trial basis.

6.1.5 Event and visibility data

Most of the effort to deal with event and visibility data is currently focused on developing standard metadata and data models. P. Lamb and A. Richards have developed a high level model for interferometric radio data that incorporates VO standards for dataset characterization. A. Richards and D. Tody have proposed an extension to the data characterization model to describe the spatial bandpass of a dataset; this is needed to characterize the limited UV coverage of interferometric radio observations and allow such data to be distinguished from other data (e.g., optical images) that uniformly samples all spatial frequencies.

6.2 Data representation (VOTable, etc.)

J. McDowell has proposed rough data formats for spectral data that would map the SSA data model into equivalent VOTable and FITS formats. T. Budavari as done an initial mapping of the SSA data model to native XML.

Following the Boston meeting we have agreement on the usage of UTYPE to tag interface elements in VOTable, however we do not yet have a specification for what to put in UTYPE tag, for the connection if any between UTYPE and GROUP, and so forth. We are trying to work this out in time to include it in the VOTable 1.1 specification.

F. Bonarrel (CDS) and P. Osuna (ESA) have suggested that a mechanism be added to allow data providers to extend the SIA VOTable query response with non-standard metadata. This is already possible in a limited way by adding nonstandard FIELD extensions to the base table. The proposed new mechanism would add more extensive metadata in additional RESOURCE elements following the main QRT. This could take the form of either additional table elements, or structured data of some form. It was agreed that a subgroup would define some use cases and propose an extension mechanism. This is a good opportunity to experiment with mechanisms for extending VOTable.

6.3 Framework (mediators, components)

Work continues to determine what is needed for a common reference framework to implement both the server and client side of the DAL services. A technical architecture for this “scalable data analysis framework” has been developed by D. Tody and others at NRAO. This defines a logical system model including a Distributed Virtual Machine (DVM) to manage scalability, and a Package Manager to manage service interfaces including parameters and parameter sets. An analysis of what it would take to implement the core framework in CORBA (for computation within high performance clusters) has been performed. We are in the process of performing a similar analysis for a SOAP implementation. Possibly a common component model can support both protocols. This work is being coordinated with a similar effort being funded by OPTICON in the UK. D. Tody represents VO in the OPTICON-funded project.

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

Planning for SSA implementations was a topic in the Boston meeting. A quick survey of those present found 8 sites planning to implement spectrum or SED services for atlas (pre-generated) data, 2-3 sites which would implement services to generate spectral data on the fly, and 2-3 sites which would put up time series data. Likely early test client software includes SpecView, IRAF, DIS, VOPlot, and many others via the FITS interface.

T. McGlynn and others, in connection with the DIS service, have manually run compliance tests on SIA 1.0 services.

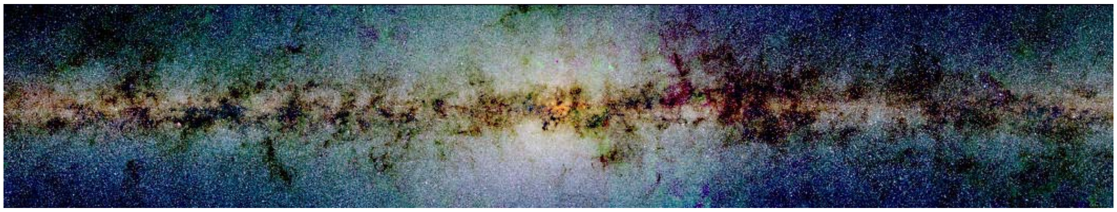

Figure 4. A 3-color mosaic of the Galactic Plane as measured by 2MASS. The image covers an area 44 x 8 degrees (representing 158,800 x 28,900 pixels). The mosaic of each color is 36.5 GB in size.

Implementation of the XML-based interface to NED holdings is well underway. At the request of the NVO project, the scheduled deployment was adjusted to allow early deployment of the component dealing with photometric information as an example of spectral energy distribution (SED) data. This will need to be adjusted when a formal spectral data specification is finalized but it was felt that the NED example was needed as input to that process.

IPAC has developed a fully functional SIA-compliant mosaicing service for 2MASS Quicklook data (which will be converted to the photometric Atlas data when this is on-line). This service runs on a cluster of five Sun Linux machines, with 1-TB of disk attached as a staging area. The service can produce mosaics with any WCS-compliant projection and any scale, though only when used through ROME (*i.e.* as a "background" job) can fields larger than about half a degree be handled. Figure 4 shows an example of a large-scale mosaic computed in this fashion.

This application is also a prototype for a grid service to produce custom mosaics for users, but maintaining the service for general community use is at present beyond IPAC's budget.

J. Good and others at IPAC have also implemented a NED XML photometry data service that makes the SED data from NED available via XML. While this service does not yet implement the SSA interface, it will help provide a real world test case for SSA.

7 Query Language

7.1 Low-level: Astronomical Data Query Language

In response to the many postings on ADQL0.7.1 and in time for the IVOA meeting, ADQL0.7.4 was published. This seemed to be well accepted at and prior to the IVOA meeting. There was till some discussion about additions to this but finally no more discussion of changing the XML structure. Again C. Page and M. Hill were most vocal in this area. The main additions as listed on the IVOA TWiki are:

- Functions : User defined funtion support in ADQL
- DML : Insert, Delete, Update (Create, Drop ?)

- Comments : SQL style comments should be in ADQL
- Sub-queries: Currently this is not suported
- JOIN : Use Explicit JOIN syntax
- INTO : MyDb, MySpace for both select and insert
- Top: Tie down semantics
- GCDIST: in ADQL or just another function ?
- UNITS: Add to ADQL/s, ADQL/x (need more metadata on table in SkyNode)
- XMATCH : Move sigma inside the bracket (or drop it ?) Add + for outer join and ? for uncovered areas

As many of these as possible will be rolled into ADQL0.8 in July. Units are still a contentious issue and although not many spoke out against units at the IVOA meeting later several people expressed reservations in private. The agreement is that units would be optional – in this way they should be acceptable. This would allow those who wished to experiment with units to do so and others to simply ignore them.

The VOQL group is also now talking more to the DAL group to see where ADQL fits in that area. An interesting presentation by Y. Shirasaki showed how having a pseudo table would allow ADQL to augment SIAP and conesearch very easily without actually needing a database behind it. This is a very interesting idea and should be worked on further in the DAL working group.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

Work continued on the OpenSkyQuery (OSQ) portal at JHU. This was up and running for the IVOA meeting and was improved several times since. A permanent URL was bought for this, namely, www.openskyquery.net. (.org was also purchased and points to the same site). As reported previously this was a major effort and integration of VOTable1.1 proved non trivial but was achieved.

A pre-team meeting at NCSA proved very fruitful and meant that the ADIL SkyNode in Java, though not finished for that team meeting, was up for the IVOA meeting. R. Williamson has put a great deal of effort in to making this work. Turing the Java node into a FullSkyNode with the Xmatch ability has proved less successful. This particular dataset is of images and does not lend itself easily to the extant Xmatch algorithm. This is also not well documented. Further work is needed on the specification for Execute plan. This will be worked on in the coming months. The Java node code and documentation will also be forthcoming soon.

A. Conti (STScI) made available GOODS, GALEX and some other datasets as OpenSkyNodes bringing the number of nodes to around thirteen.

OSQ uses the registry to find SkyNodes and uses the spectral information to list them and color code them according to frequency in the query screen. The exception to this is the registry itself, which implements a basic SkyNode but of course does not fall in any bandpass.

OSQ also allows upload of a table of sources that may then be used for cross-matching against any of the other cross-match nodes.

Much tidying up and documentation is needed in this area now that a working system is in place.

7.3 High-level: Complex queries

The interaction of VOQL and DAL groups may be seen as an initial foray in to this area. Seeing how ADQL may be used for more than tabular query and its position in the DAL should lead us to have some more complex queries. As mentioned above the pseudo-table notion is interesting in this area.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)

Caltech (M. Graham, A. Mahabal) has produced a set of statistics services for the VO at the AAS meeting in Denver, supported also under NSF-Astrostatistics project (PI G. J. Babu, Penn State). These implement sophisticated statistical services for data in the VOTable format. Caltech (R. Williams) has also produced a collection of SOAP web services to implement the xy2sky and sky2xy tools that are part of the WCSTools package (D. Mink, SAO).

R. Plante has been prototyping some best practices for defining “Standard Web Services” —Web Services that are defined by a WSDL file that can be standardized via the IVOA standardization process. Among the issues being address are optional arguments and operations, inheritance/inclusion of other IVOA standards, and non-standard extensions. These issues have become relevant with the emerging standard Web Services for registries, OpenSkyNodes, and Support Services.

R. Williamson has implemented a basic OpenSkyNode Web Service against the ADIL image catalog using Java, Apache Axis, and PostgreSQL. Williamson and Plante are turning this prototype into an exportable package that helps providers connect an OpenSkyNode interface to their existing JDBC-capable catalogs.

R. Plante has developed classes that make it easier to plug existing XML software (e.g., JAVOT for VOTables) into Axis-based Web Services.

JHU has built a spectral service supporting SSA that was demonstrated for the IVOA meeting and for the NSF when they came to JHU June 29.

IVOA Support Services Specification 0.2 is under discussion with other IVOA partners. This specification provides the means to request availability and uptime information for registered services, and information about planned downtime.

8.2 Grid Services (OGSA)
Caltech is planning joint activity with the UK e-Science project for a sequence of "SC4DEVO" workshops (Service Composition for Data Exploration in the Virtual Observatory). The workshop will be July 12-15, covering topics such as formal description of grid services, persistent services, authentication and security, and ways to bring professional astronomers to the emerging service-oriented architecture.

ISI has created a new MCS release v3.1 (http://gaul.isi.edu/mcs/v3.1) that is compatible with OGSA-DAI release 4.0 and Globus Toolkit v3.2. The new release provides the same functionality as the previous release and has been tested on Redhat Linux OS. ISI has also updated the Montage version used by the Pegasus Grid Portal (v2.0). An application programming interface to Pegasus that can be used by NVO Portals is under discussion. The API interface would provide functionality for workflow mapping and submission.

8.3 Computational resource management
ISI is investigating the possibility of achieving greater speedup on the Montage runs by doing aggregation of nodes in the workflow into larger clusters. This is an attempt to redefine the granularity and parallelism of the workflow in the presence of scheduling and other overheads.

8.4 Virtual data
SDSC released version 3.2 of the Storage Resource Broker data grid on 2 July 2004. This version has an associated web services interface that is available through the Matrix system, also version 3.2. Matrix is used either as a grid workflow management system or as SRB Web Services. Matrix uses a Data Grid Language (DGL) as the medium to specify interactions between Matrix Web Service clients and the server. The software development was funded through the NSF Grid Physics Network project, and is an example of the interchange of software technology between ITR projects.

New features include:
- Ability to manage data-flows over a very large number of files/data sets. Matrix can in parallel start a new dataflow based on the number of files (increasing the speed of execution) even for non-parallel applications.
- Provenance tracking of gridflows for all data and process during the execution and even after the execution.
- Developer friendly Java client API.
- Declaration of scoped variables within the workflow to help in dynamic computational steering based on the previous results.
- DGL developer guide

8.5 Application and service integration with Grid

MaxBCG, Finding Galaxy Clusters using databases systems technologies. The Maximum-likelihood Brightest Cluster Galaxy (MaxBCG) application searches for galaxy clusters that span a large dynamic range of cluster masses and provides good redshift estimates. MaxBCG was developed originally in TCL by J. Annis (Fermilab) and ran on the Terabyte Analysis Machine (TAM), a 10-CPU cluster specially tuned to solve this type of problem. The same application code was grid-enabled and used to test the Chimera Virtual Data System created by the GriPhyN project. As is common in file-based grid applications, the TAM and Chimera implementations use thousands of files that travel from the SDSS Data Archive Server to the computing nodes. SkyServer is the relational database hosting the SDSS catalog data. All the data required to run the MaxBCG is available in the SkyServer database. At Johns Hopkins University, we have implemented a version using SQL on a three-node Microsoft SQL Server 2000 cluster that processes the same sky area 20 times faster than TAM.

In grid environments, databases and database systems are typically used only to access and integrate data, but not to perform analytic or computational tasks. Limiting usage in this manner neglects the main strength of database systems, which is their ability to efficiently index, search, and join large amounts of data. It is a mistake to move large amounts of data to the query, when you can move the query to the data and execute the query in parallel. For this reason, we are investigating how the grid nodes could benefit from supporting Database Management Systems so that SQL applications could be deployed as easily as traditional grid applications coded in C, FORTRAN, etc.

We are also developing an efficient grid-enabled implementation of MaxBCG that takes advantage of databases technologies and does not require transferring thousands of files over the network. When the user submits the MaxBCG application, upon authentication and authorization, the SQL code (about 1000 lines) is deployed on the available data-grid nodes hosting the SkyServer database system. Each node analyzes a piece of the sky in parallel and stores the results locally or, depending on the policy, transfers the final results back to the origin. At the moment, two different organizations host the SkyServer database and the CasJobs system: Fermilab (Batavia, IL, USA) and The Johns Hopkins University (Baltimore, MD, USA). In the near future, the Inter-University Centre for Astronomy and Astrophysics (IUCCA) in Pune, India, will also host the system.

We expect MaxBCG to provide us with a use case to implement a workflow system based on database technologies.

9 Applications

The first meeting of the IVOA Applications Interest Group (AIG) was held at the Cambridge Interoperability Meeting in May. One of the main purposes of this group is to facilitate the flow of information about new and revised applications. Demonstrations were given of services including several developed under the NVO aegis.

The AIG also agreed to look over existing and new use case scenarios developed by the NVO and other organizations with a view towards ensuring that the standards being developed in the virtual observatory community are appropriately matched to the applications needed by the astronomical community.

9.1 Data location services

A new version of the Data Inventory Service was released in April that provides the capability of searching all images from archival SIAP services. The internals of the DIS are being re-engineered to use a set of simple portable Perl tools that others will be able to use to build VO applications.

The team is also looking at how the DIS may form part of the Education program. A more catch title is probably needed; suggestions include NVO Telescope or SkyFinder.

9.2 Cross-correlation services

The yourSky service, developed at JPL, was described on the VO web site. yourSky provides a way to access customized images from a number of survey datasets.

Work has begun on the development of a query builder interface that will enable users to conveniently access SkyQuery sites. SkyQuery is currently being tested by trusted non-developers.

Related Activities. The NASA AISR funded ClassX project continued to use NVO developed technologies to help in its efforts to classify all detected ROSAT sources. The German Astrophysical Virtual Observatory (GAVO) project correlation tool is being linked to classification tools developed at the HEASARC and STScI in a non-NVO funded project. The GAVO activity has highlighted areas of concern with the incomplete adoption of VO standards by some groups.

9.3 Image combination, registration

Integration of the Montage image visualization service with the TeraGrid is discussed in the Architecture section, WBS 3.

9.4 Visualization tools and services

The latest versions of the JHU Spectral and Filter services were demonstrated at the Cambridge interoperability meeting. These services are available to the public.

9.5 Theory

No new theory-related applications were developed this quarter. See WBS 2.1 for activities related to the Theory Interest Group.

9.6 Statistical analysis

Related Activities. The VOStat project (NSF Focused Groups Research grant DMS-0101360) at CMU/U. Pittsburgh aims to provide a rich suite of web-based statistical services for exploration and analysis of astronomical data sets, especially in the context of a Virtual Observatory environment. Functionality currently offered includes multi-resolution k-dimensional trees for clustering and outlier detection and more traditional techniques such as principal component and survival analyses. We are currently updating the help files. We plan to include in the help files information on what particular program does, how the output looks, the code for the program, and links to additional resources for advanced users. We are also in the process of adding many more functionalities. Science applications using VOStat were demonstrated at the Denver meeting of the American Astronomical Society. We are also planning on adding recently developed streaming techniques to estimate/update the results as the data streams through. We are testing to add an easily extensible distributed web services-based framework transparently accessed via an open-source client GUI.

9.7 Datamining, outlier identification
Related Activities. The NSF ITR supported GRIST team has begun to explore the opportunities afforded by the VO to perform data mining and data exploration in astronomy. The GRIST team web site is at http://grist.caltech.edu.

9.8 Interfaces to/from legacy software systems
Work continued at several sites to provide access to existing archive and catalogs services using new Web protocols. Many new services were made available, notably NED and other JPL tools.

10 Community Engagement

10.1 Documentation
NVO team members have been active in the development and promotion of international VO standards, leading or making major contributions to the following standards documents and usage notes in the IVOA document collection:

- IVOA Identifiers Version 1.1 (R. Plante). http://www.ivoa.net/Documents/PR/Identifiers/Identifiers-20040621.html
- Virtual Observatory Architecture Review (R. Williams). http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch.html
- VOTable Format Definition Version 1.1 (R. Williams, R. Hanisch, T. McGlynn, A. Szalay, contributors). http://www.ivoa.net/Documents/PR/VOTable/VOTable-20040608.html
- Simple Image Access Specification Version 1.0 (D. Tody and R. Plante). http://www.ivoa.net/Documents/WD/SIA/sia-20040524.html
- Resource Metadata for the Virtual Observatory Version 1.01 (R. Hanisch). http://www.ivoa.net/Documents/REC/ResMetadata/RM-20040426.html
- Lessons Learned Using the VOResource XML Schemas in the NVO Version 1.0 (R. Plante, M. Graham, G. Greene, R. Hanisch, J. Lee, W. O'Mullane, R. Williamson). http://www.ivoa.net/Documents/latest/RMExp.html
- Data Model for Quantity Version 0.2 (J. McDowell). http://www.ivoa.net/internal/IVOA/IvoaDataModel/qty.v0.2.pdf
- Data Model for Observation Version 0.2 (J. McDowell). http://www.ivoa.net/internal/IVOA/IvoaDataModel/obs.v0.2.pdf
- IVOA Astronomical Data Query Language Version 0.7.4 (A. Szalay, W. O'Mullane, V. Haridas). http://www.ivoa.net/internal/IVOA/IvoaVOQL/ADQL-0.7.4.pdf

A major additional effort in documentation will come in conjunction with the NVO Summer School (WBS 10.3 below).

10.2 Web site

The NVO Project web site (http://us-vo.org/) was redesigned and reimplemented, changing the orientation from in-team communications to reaching the astronomy user community. There is an emphasis on practical use of NVO tools and applications, and a series of graphics boxes that highlight events or projects. The new design is shown in Figure 5.

10.3 Technical training initiatives

Funding requests to NSF and NASA to support an NVO Applications Software Summer School were approved, and planning began in earnest this past quarter. The Summer School was announced through a variety of electronic channels, including mailing lists, the AAS e-Newsletter, and the CSWA (Committee on the Status of Women in Astronomy) e-Newsletter. Forty-eight applications were received for 40 places. The final list of participants includes a dozen graduate students, and a third of the participants are from outside the United States. A faculty of twelve NVO team members has been busy crafting the program. Facilities have been secured with the Aspen Center for Physics, with accommodation at the nearby Aspen Meadows Hotel. The JHU travel office is arranging travel for those who requested support. Lectures and other course materials will be made available on a public Summer School website. The Summer School will be held the week of September 13-17, 2004.

Figure 5. The new NVO Project website, http://us-vo.org. The design emphasizes use of NVO facilities, tools, and applications by the astronomy community rather than communication among team members (which is now handled largely within the IVOA working groups).

10.4 Advocacy

R. Hanisch submitted a proposal to the American Astronomical Society for a special topical session on VO-enabled science, to be held during the January 2005 meeting of the AAS in San Diego. The AAS Council reviewed and approved the proposal at its June 2004 meeting.

R. Hanisch collaborated with P. Quinn (AVO) and A. Lawrence (AstroGrid) on a white paper for the OECD Global Science Forum entitled "The Management, Storage, and

Utilization of Astronomical Data in the 21st Century." This white paper was strongly endorsed, and the OECD GSF issued recommendations that are very supportive of the international VO initiatives (see Section 1.5 for details).

11 Education and Public Outreach

11.1 Strategic partnerships

We have developed a partnership with Project Lite at Boston University. Project Lite includes an interactive, web-based tool for teaching undergraduate students about astronomical spectroscopy. Through use of NVO interfaces, Project Lite will have access to hundreds of thousands of spectra rather than just ~50 samples. NVO team members joined a UC-Berkeley based proposal (also including ManyOne) to NASA's Applied Information Systems Research Program to implement a general public portal to NVO data collections.

11.2 Formal education

J. Raddick (JHU) developed a set of use cases in which NVO services could be utilized for school curriculum at various grade levels. He has begun work on implementation. Use cases include:

1) Adopt an Object. A class of middle school general science students participates in an "adoption" program for deep sky objects. Earlier in the school year, they had "adopted" a species of wild animal; now they are doing a similar activity to learn astronomy. Each student picks a famous star, galaxy, nebula, or comet and researches it in depth. They view images of the object as seen in different wavelengths (radio to gamma ray) at different times. Each student would prepare a written report and oral presentation summarizing what astronomers know about the object they adopted.
2) A New Nebula. High school honors astronomy students examine many images of the variable reflection nebula McNeil's Nebula (http://seds.lpl.arizona.edu/messier/more/m078_mcneil.html), which was discovered by amateur astronomer Jay McNeil in 2004. Can students see it in pre-2004 images? What did it look like for the past 75 years? What did it (and does it) look like in different wavelengths and in spectra? Can students use small telescopes and CCD cameras to find the next new nebula?
3) Image Catalog (with Captions!). A retired physics teacher with a strong interest in space gives talks several times a year astronomy clubs, teachers, students, and adults with emphasis on astronomy. The talks are usually power point slides of images, alternating with demonstrations. He is looking for a photo album-style catalog of beautiful images WITH DESCRIPTIVE ANNOTATIONS (his emphasis) to include in these presentations.
4) Olympic Coordinate Conversions. A middle school geography teacher and a high school physics teacher collaborate on a tutorial and competitive event for the national Science Olympiad, focused on using mapping coordinates to locate positions on Earth and in space. The tutorial would introduce students first to latitude and longitude on Earth, then to RA and dec in the sky. The event would have students use NVO to locate the same objects using various representations (RA/Dec vs. alt-azimuth, HMS vs. decimal degrees, J1950 vs. J2000, etc).

5) Color-Magnitude Diagrams. A high school honors-level astronomy teacher searches stellar data to make two color-magnitude diagrams: one a random sample and one for a particular star cluster. The class would retrieve magnitudes in several wavelengths (not just visible) and produce several diagrams. Students would compare diagrams in all wavelengths and between the two diagrams.
6) The Moon and Planets. A 5^{th} grade special education teacher studies the moon and visible planets (Venus, Jupiter, Saturn) in her class. The students have severe learning disabilities, but recognize the visible planets from their own naked-eye observations. The teacher shows them the moon and planets through an inexpensive 6" refractor, then shows them Hubble Space Telescope images retrieved through the NVO. She tells them that they are seeing the same planet with naked eye, the refractor, and the HST at increasing magnification/resolution.

11.3 Informal education
No activity this quarter.

11.4 Outreach and press activities
No activity this quarter by NVO, but the AVO project issued a press release on their first science results. Team members led by P. Padovani (ESO) describe the discovery of new Type 2 (obscured) quasars in a paper accepted for publication in *Astronomy and Astrophysics* (http://xxx.lanl.gov/PS_cache/astro-ph/pdf/0406/0406056.pdf).

11.5 Technical development
C. Christian completed a review of NVO EPO opportunities, a census of potential partner organizations/projects, and an assessment of NVO infrastructure required to support EPO product development. This review will form the basis for partnership development and will help guide development efforts.

Activities by Organization

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

The Palomar-Quest synoptic survey is a major new sky survey that is being built from the ground up with VO standards. The survey now has 5 TByte of imaging data, collected in 112 nights observation at the Palomar 48" telescope. Data is archived at NCSA, with catalog processing and image reprojection (Atlasmaker) being done on the Caltech TeraGrid machine. The NSF TeraGrid project provides high bandwidth between these sites. The catalog product of the Palomar-Quest survey is being deployed to a new database server as OpenSkyNode.

Participated in the definitions of the Registry web service interfaces and the Standard Service web service interface, particularly at the IVOA Interop meeting.

Presented of a set of statistics services for the VO at the AAS meeting in Denver, supported also under NSF-Astrostatistics project (PI G. J. Babu, Penn State).

A collection of SOAP web services has been deployed at Caltech to implement the xy2sky and sky2xy tools that are part of the WCSTools package (D. Mink, CfA).

Caltech – Infrared Processing and Analysis Center (IPAC)

- Formal test of the ROME request management system is underway.
- A SIA-compliant general mosaicing service for 2MASS data has been put on-line for NVO evaluation.
- Generalized VOTable output for NED queries have been partially completed. One of these components is also a prototype spectral/SED services.

Canadian Astronomy Data Centre/Canadian Virtual Observatory

No activity reported this quarter.

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

The VOStat project (NSF Focused Groups Research grant DMS-0101360) at CMU/U. Pittsburgh aims to provide a rich suite of web-based statistical services for exploration and analysis of astronomical data sets, especially in the context of a Virtual Observatory environment. Functionality currently offered includes multi-resolution k-dimensional trees for clustering and outlier detection and more traditional techniques such as principal component and survival analyses. We are currently updating the help files. We plan to include in the help files information on what particular program does, how the output looks, the code for the program, and links to additional resources for advanced users. We are also in the process of adding many more functionalities. Science applications using VOStat were demonstrated at the Denver meeting of the American Astronomical Society. We are also planning on adding recently developed streaming techniques to estimate/update the results as the data streams through. We are testing to add an easily extensible distributed web services-based framework transparently accessed via an open-source client GUI.

Fermi National Accelerator Laboratory (FNAL)
Working with other Fermilab scientists, V. Sekhri completed loading of a final version of DR3 (data release 3) Catalog Archive Server (CAS) with SDSS data for testing. The CAS is the backend database for the SDSS SkyServer. The bottleneck in the process turns out to be the building of indices to speed up the query speed. Final release to the public and the NVO community is scheduled for October.

V. Sekhri completed his work installing gsiftp and a gsirft (reliable file transfer protocol) interface to the SDSS Data Release 2 Data Archive Server (DAS), which contains a flat file version of the data, including the full pixel data. gsiftp and gsirft use certificates to provide authentication and provide a more robust mechanism than web services for transferring bulk data, particularly when authentication mechanisms are needed. Argonne National Laboratory (ANL) and NCSA intended to replicate the DR2 dataset for use on the TeraGrid (WBS 3).

V. Sekhri left the project in May, and a search for a replacement hire is in progress.

High Energy Astrophysics Science Archive Research Center (HEASARC)
HEASARC activities in the past quarter focused in three areas. T. McGlynn chaired the new IVOA Application Interest Group, developing its TWiki site, and planned and chaired the AIG's inaugural session at the Cambridge Interoperability meeting. A report was made to the IVOA steering committee describing the consensus findings of that meeting. Notable issues were the need for reference implementations of VO standard interfaces, and the desirability.

A new version of the Data Inventory Service was released which provides more complete access to archival data. The DIS is being reengineered on top of a set of small, simple and portable Perl modules. Other applications at the HEASARC and elsewhere will be able to use these modules to access VO services including SIA, Cone, SSA, and Registry services. We also hope to include access to SkyQuery services. In addition to facilities for querying and displaying services, the new framework includes simple capabilities for exploding queries to a number of services and for caching results. Documentation of the current capabilities and the new modules is being developed.

Revisions and documentation of the OAI publications capabilities for HEASARC services is the third major element of the HEASARC activities in the past quarter. Currently HEASARC services are described in a static text database independent of the HEASARC's catalog databases. Work is underway to ensure that HEASARC scientists can quickly and conveniently update the entries to be published in OAI and to ensure that the HEASARC's internal metadata matches what is published externally.

Preparations for the NVO Summer School are proceeding.

HEASARC personnel attended the weekly NVO metadata telecons.

Johns Hopkins University

T. Budavari continued to work on Simple Spectral Access protocol and data model with the team. Spectral services were moved to the web cluster at JHU and made generally more robust.

V. Haridas fixed some bugs in the ADQL translation service. Working on new FITS cutout service that will also provide JPG-format files using the asinh stretch.

W. O'Mullane has continued to work on the OpenSkyQuery portal with V. Haridas and N. Li. This was demonstrated at the IVOA meeting in Boston to show ADQL0.7.4 in action and for the NSF June 29 at JHU. Work continued with G. Greene (STScI) on the searchable registry, mainly concerning enhancements of the interface and incorporating the new VOResource v0.10. Continued to provide feedback on the nascent registry specification document. Produced the spec for Support Services.

N. Li worked on the OpenSkyQuery portal. This included a new schema browser for nodes and enhanced the feedback from the individual nodes during operation. There are still some user interface issues to deal with. Worked on changes to CasJobs to make it more amenable to implementation on the Grid.

M. Nieto-Santisteban has continued to work on the SQL version of the Finding Galaxy Clusters algorithm (MaxBCG) developed originally in TCL by J. Annis (Fermilab). Preliminary results showing a 2X speed-up improvement over the grid implementation were presented in the January-March 2004 Quarterly Report. During this period, we have enhanced the algorithm and now the application runs an order of magnitude faster (20X faster) than the original TAM TCL-file-based implementation. This work has demonstrated that Database Management Systems have a good place in the grid environment and may significantly improve performance in some cases. A paper has been submitted to the IEEE /ACM International Workshop on Grid Computing 2004 to be held jointly with the Super Computing 2004 conference.

Work has also been done to design a grid version of CasJobs that will allow running SQL applications on different sites holding the SDSS SkyServer database. A demonstration of the ImgCutout application was presented to NSF.

A. Thakar attended a Grid Performance Workshop in London and spoke about logging on SkyServer. This was briefly presented to NSF also. Logging is part of the Support Services Specification. Working with D. Wittman and P. Gee at UC Davis to get their DLS (Digital Lensing Survey) DB hooked up to a SkyServer-like interface. They also want to implement web services like ImgCutout and (maybe) CasJobs, and they are already a SkyNode

G. Fekete, A. Szalay, J. Gray, and W. O'Mullane have worked on the HTM and region support. An HTM paper in the form of a Microsoft Technical report has been released.

J. Raddick has written some use cases for EPO and is starting to implement some of them.

S. Carliles has been working on Mirage. A new image module that does not rely on JAI is now almost ready for release. This will be released built into Mirage but also as a separate Java module for others who need a Java image module that handles FITS. This module will also contain the new asinh scaling for the image.

Microsoft Research
No activity this quarter.

National Optical Astronomy Observatories (NOAO)
Members of the NOAO Data Products Program attended the NVO Team meeting in Champaign, IL (M. Fitzpatrick) and the IVOA workshop in Cambridge, MA (M. Fitzpatrick, R. Hiriart, R. Shaw, P. Warner), which were held during the past quarter. NOAO management is preparing a revised statement of work for NVO related activities, consistent with a smaller level of effort and the reduced funding level provided by NVO project management. NOAO Archive development staff are preparing their domain data model for review and incorporation into the NVO data modeling effort."

National Radio Astronomy Observatory (NRAO)
In the past quarter D. Tody contributed to NVO in the following areas:

- Chaired the DAL sessions at the IVOA workshop in Boston in May.
- Lead development of the SSA data model and query interface.
- Further work on VOTable and FITS data representation for SSA.
- Scalable data analysis framework research and design.
- Participated in IVOA system architecture WG.
- Worked with VOQL WG on ADQL integration into DAL services.
- Fostered development of data models and formats for visibility data.

Our attempt (lead by J. Ulvestad) to produce reference images for archival VLA data has stalled for the moment due to lack of resources to implement and operate the pipeline. We continue to look for a way to make this happen. In the meantime E. Greisen is working on autoflagging code for AIPS for use in an AIPS-based VLA processing pipeline.

The “E2E architecture and oversight committee,” co-chaired by W. Cotton and D. Tody, has begun work to define an overall dataflow and data management system for all NRAO telescopes including ALMA. Data access to the VO is a component of this, but the scope includes all aspects of observing and post-processing of data through to the final data products in the NRAO archive. Part of this effort includes specifying science data models and export data formats for ALMA, EVLA, VLBA, and GBT. We are coordinating this work with VO data models development.

We have started shipping bulk data to NCSA via a disk box, as the first step in setting up an automated replica of the NRAO archive at NCSA.

Raytheon/ADC (George Mason University and University of Maryland)
GMU personnel (K. Borne) contributed to the ideas and planning for the new Applications Interest Group and Theory Interest Group. We attended the IVOA Interop Meeting at Harvard in May 2004 and participated in various working group activities. We continue to investigate algorithms for distributed data mining in the VO. With computer scientist collaborators at UMBC, we are implementing a test case for finding outliers in distributed databases. We are exploring the multi-dimensional photometric parameter color-color space defined by the combined 2MASS-SDSS colors. We are testing a new theorem for finding outliers in the full multi-database by first identifying outlier candidates in much smaller subsets of the distributed data, thereby greatly reducing the volume of data to be retrieved and shipped. We hope to extend this work to include additional distributed databases: the ROSAT All-Sky Survey (RASS) and NRAO VLA Sky Survey (NVSS). We will publish the results and present them at a future Data Mining Conference.

Data Model work on Quantity, STC, and Observation continued at U. Maryland (E. Shaya). We are near completion of specialized code to read Quantity Objects. This code parses and interprets the XML Schema of a document and thereby is aware of which elements inherit from either of the 3 forms of a Quantity or is a QuantitySet. The general concept of Schema understanding by applications is new and may provide a powerful tool for object-oriented applications. Now a totally new object can be properly dealt with by treating it as one would treat a familiar superclass of the object.

We are exploring OWL (Web Ontology Language) for use in UCDs and Units. The idea is to have URI strings in Fields that point to an OWL class that describes the UCD or Unit and provides information on its properties and its relationship to similar entities.

We attended a four-day conference on Protégé Ontology Engine, which is the de facto standard for OWL development and will be reporting to the technical committee on this soon.

We provided a short report on XInclude, a W3C candidate recommendation for inclusion of XML documents within each other. This was adopted as a promising way to include space-time coordinate system declarations into XML data files such as VOTable.xml.

We attended a two-day NVO team meeting and took part in Metadata WG weekly teleconferences.

San Diego Supercomputer Center
In support of the NVO, L. Brieger at SDSC installed, ran, and verified the latest versions of Montage on DataStar. A script was produced to generate 1734 Montage mosaics that fill the entire sky for the 2MASS sky survey. Montage has also been installed on the TeraGrid cluster. The current status is that the large runs will be done from Caltech.

SDSC has initiated a project with MIT on the integration of DSpace and Storage Resource Broker technology. The goal is the creation of a digital library environment that will scale to the size of NVO collections. This project is funded by NARA, but the results will be made available for use in the NVO.

R. Moore participated in NVO Executive telecons and attended the IVOA Interoperability workshop in Boston and the IVOA Executive meeting in Glasgow.

Smithsonian Astrophysical Observatory
At SAO efforts continued via email on the Observation model, which describes general metadata for observations. It is divided between the actual data values (Characterization) and the observation and analysis process (Provenance). Further defining the Characterization was deemed higher priority. A new draft of the Simple Spectral Access SED data model was developed in collaboration with D. Tody (NRAO). It defines an SED model and proposes XML-schema, VOTABLE, and FITS serializations for the model. Progress was also made on the Quantity model via email and discussion at the Boston IVOA meeting. The Space-Time Coordinate schemas were thoroughly revised.

The CfA/SAO VO activities also include development of a CfA-VO prototype to test the IVOA Data Model design as part of the VO CfA (Harvard) NSF proposal (PI: A. Goodman). People involved in these activities are J. McDowell, A. Rots, I. Evans, J. DePonte Evans, M. Cressitello-Dittmar, and M. Harris. Optical and Chandra archive federation is a first step; extension to other archives comes next.

This quarter, the group worked to expand the design of the DataContainer object to include the concept of Accuracy. We have also created instance diagrams of several types of data, and sample serializations of them.

The bulk of our effort has gone into implementing code to create, read, and display several prototype DataModel objects. Efforts include a Units Module, UCD Module, Frames, Container interface, and the Core and Simple Container classes.

Meetings:

- A. Rots and J. McDowell attended NVO Team meeting, NCSA - Urbana, IL, April 29, 30
- A. Rots, J. McDowell, G. Fabbiano, M. Cressitello-Dittmar, M. Harris attended and presented at the IVOA Interoperability meeting, Cambridge, MA, May 24-28
- G. Fabbiano represented SAO (as `expert' in Digital Archives) in the Smithsonian Institution Collection Task force. Recommended A. Rots to serve in the SI metadata working group, to increase SAO participation and foster a VO-friendly set-up.

Presentations:

- G. Fabbiano gave presentation on the VO and CfA VO activities to a visiting Congressional Science Committee group.

- G. Fabbiano gave presentation on the CfA Archival and VO activities to the CfA Visiting Committee.

Initiatives:

- G. Fabbiano oversaw the logistics organization of the IVOA meeting in Cambridge, and ensured funds for the same. CfA people who helped out in this effort were R. Harnden, R. Brissenden, the HEA administration, and A. Goodman (Harvard).
- G. Fabbiano made VO points in science talks when appropriate (e.g. at AAS special session on X-ray populations in Denver).
- G. Fabbiano successfully lobbied CfA director for the need of improving and making V0-compliant the multi-wavelength CfA archives, and was appointed to lead this activity. As chair the CfA archive committee, Fabbiano successfully urged the Director to publish a CfA data policy, with limited data proprietary periods for non-NASA data. Wrote proposal to SI for archive funds to support this effort.
- G. Fabbiano continued to advocate NASA VO funding.
- G. Fabbiano submitted NASA AISR proposal for SED tool suite (IRIS), in collaboration with J. McDowell, A. Rots, I. Evans, and G. Fazio and M. Ashby (Spitzer Science Center).
- G. Fabbiano is a member of the support group for the Harvard Computing Initiative (Major proponent: A. Goodman).
- G. Fabbiano advocated the establishment of a standard X-ray photometric system (with J. McDowell; no proposal funded so far, but these obvious new ideas always take time).
- G. Fabbiano, as Head of the CXC Data System Division, is ensuring that we have a forward-looking, VO-friendly outlook. Relevant efforts, besides the Data Model and the Chandra Archive, include the development of Level 3 pipelines and subsequent analysis tools. Level 3 will produce source and source property lists for catalogs as well as source-based data-cubelets for archival. Fostered Chandra/ADS archive/ literature connection, implemented by A. Rots and G. Eichhorn.

Space Telescope Science Institute

Substantial progress was made on the NVO Registry, led by G. Greene and W. O'Mullane (JHU). The registry database was re-indexed so that keyword-based searches can be completed very quickly. A much-improved user interface to the registry was written and made available. An automated harvesting service was also implemented to check other VO OAI registries hourly for updates to resource. A SkyNode interface was implemented via a Registry web service such that the Registry is now accessible directly through the OpenSkyQuery web portal. This includes the services that accept the VO ADQL (Astronomical Data Query Language) as part of the emerging query standard for the IVOA community. Automated replication between JHU and automated harvesting of all VO registry sites was developed such that we have two viable mirror sites for failover protection and resources updated on an hourly basis. Changes that occur at remote sites will be propagated rapidly to the ST VO Registry.

The search interfaces now support paging between large numbers of queried resources to eliminate browser buffering difficulty. The new standard VOResource XML schema has been successfully tested with the VORegistry web services. In the process, missing element definition was reported to the IVOA team and will be used for further enhancements to the standard.

MAST staff (A. Conti, B. Shiao) implemented SkyNodes for the GALEX, HDF, and GOODS source catalogs. OpenSkyQuery is currently in beta testing mode and will also host these SkyNodes. The STScI HST archive has continued to expand the publication of VO compliant services, including SIAP and cone search services. These services have been published in the STScI NVO registry and can be harvested to other IVOA registry sites. The SkyNodes are now publicly available through the OpenSkyQuery portal and support ADQL.

An NVO TWiki site was established at STScI to support collaborative development between JHU, STScI and other VO developers.
http://nvotwiki.stsci.edu/twiki/bin/view/Main/WebHome.

R Hanisch worked with A. Szalay (JHU) on a major redesign of the NVO team website. The new site takes a more user-oriented view and features VO-enabled research. See http://us-vo.org/.

NVO staff met with representatives of the GOODS science team on two occasions to discuss opportunities for collaboration. We would like to make greater use of the GOODS data as a test bed for VO applications and to get feedback from GOODS scientists on their needs/expectations for VO tools. An image cutout service developed by GOODS scientists is currently in the process of adapting VO SIAP standards. Several GOODS catalogs are planned for OpenSkyNode services.

R. Hanisch attended the IVOA Executive Committee meeting, held in conjunction with the SPIE Conference in Glasgow. International cooperation remains strong. The IVOA has grown to 15 member projects with the addition of the Spanish VO.

United States Naval Observatory
No activity this quarter.

University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA)
R. Plante has led the revision of the VOResource metadata schemas. To this end, he organized the April videocon workshop and coordinated comments and testing. He has organized the documentation on the TWiki to make it easier to find current schema information.

R. Williamson continued work on the SkyNode interface to the ADIL using Java, Apache Axis, and PostgreSQL. He attempted to add “Full” SkyNode capabilities on top of the existing “Basic” functions; this primarily means implementing a function to enable object cross-identification. Since our goal was to test the SkyNode specification and prototype software, the fact that the ADIL is not an object catalog (but rather an image catalog) was not important to us. However, given the limits encountered in stretching the cross-identification intent and ambiguities in the current draft specification, we have tabled further work on this "full" compliance. We are not packaging the work we have done to allow other sites to quickly connect a SkyNode interface to an existing relational database.

In connection to the development of the standard Registry Interface specification, Plante has been experimenting with techniques of defining standard Web Service Description Language (WSDL) documents. The intent is that IVOA would publish a single WSDL document that represents a standard interface that many providers would implement. Once a client has been implemented to work with one compliant service, it should be able to work with any other compliant services without further code modification. A number of issues must be addressed including optional methods, optional arguments, standard inheritance/inclusion, and non-standard extensions. Plante is writing up his results as an IVOA Note.

Finally, NCSA hosted the spring NVO team meeting in April. Taking advantage of the face-to-face time, V. Haridas and W. O'Mullane (JHU) arrived early collaborate on the SkyNode implementation and the ADQL specification.

University of Pennsylvania
P. Protopapas has worked on the following:

- Creating a web site that integrates all tools build for light curves for searching and comparing light curves.
- In this quarter we tested our algorithm to detect extra solar planets by analyzing light curves. We implementing this algorithm as a web service and we are integrating it in our light curve web site.

University of Southern California (USC/ISI)
The work during the three-month period has been on

- Developing a new MCS release
- Updating the Montage version on the Pegasus Grid Portal
- Investigating the possibility of higher speedup through aggregation of montage jobs

ISI has created a new MCS release v3.1 (http://gaul.isi.edu/mcs/v3.1) that is compatible with OGSA-DAI release 4.0 and Globus Toolkit v3.2. The new release provides the same functionality as the previous release and has been tested on Redhat Linux OS. ISI has also updated the Montage version used by the Pegasus Grid Portal (v2.0). An application programming interface to Pegasus that can be used by NVO Portals is under discussion. The API interface would provide functionality for workflow mapping and submission. ISI is also working on getting our accounts renewed on the TeraGrid for doing Montage runs.

ISI is also investigating the possibility of achieving greater speedup on the Montage runs by doing aggregation of nodes in the workflow into larger clusters. This is an attempt to redefine the granularity and parallelism of the workflow in the presence of scheduling and other overheads.

University of Wisconsin
No activity reported this quarter.

Publications

"There Goes the Neighborhood: Relational Algebra for Spatial Data Search", A. S. Szalay, G. Fekete, W. O'Mullane, A. R. Thakar, G. Heber, A. H. Rots, MSR-TR-2004-32, April 2004

“The International Virtual Observatory Alliance: Recent Technical Developments and the Road Ahead,” P. J. Quinn et al., 2004, SPIE 5493-13.

“The Simple Spectrum Access Protocol,” M. Dolensky & D. Tody, 2004, SPIE 5493-47.

“Montage: A Grid-Enabled Engine for Delivering Custom Science-Grade Image Mosaics on Demand,” G. B. Berriman et al., 2004, SPIE 5493-22.

“The Digital Universe Coalition: Building a Prototype NVO E/PO Portal,” Mendez, B.; Craig, N.; Haisch, B.; Lindblom, J.; Hanisch, R.; Summers, F.; and Abbott, B., AAS Meeting 204, 78.05.

“Spectral Energy Distributions in the Virtual Observatory,” McDowell, J. C.; Fabbiano, G.; Evans, I., 2004, AAS Meeting 204, 76.04.

“Doing Science with VOStat,” Mahabal, A.; Djorgovski, S. G.; Graham, M.; Williams, R.; Feigelson, E.; Babu, G. J.; Nichol, R.; Vanden Berk, D.; Wasserman, L., 2004, AAS Meeting 204, 76.03.

“VOStat: a distributed statistical toolkit for the Virtual Observatory,” Graham, M. J.; Djorgovski, S. G.; Mahabal, A.; Williams, R.; Feigelson, E. D.; Babu, G. J.; Nichol, R.; Vanden Berk, D.; Wasserman, L., 2004, AAS Meeting 204, 76.02.

“Multiwavelength Astronomy Education with the Sloan Digital Sky Survey,” Raddick, M. J.; Sparks, R., 2004, AAS Meeting 204, 19.04.

“SDSS as a Cornerstone of the Future National Virtual Observatory,” Ivezic, Z.; Sloan Digital Sky Survey Collaboration, 2004, AAS Meeting 204, 14.04.

Virtual Observatory Articles in the Popular and Technical Press

"The Virtual Observatory," by Suzanne Ross, Microsoft Research News and Highlights, 2004. http://research.microsoft.com/displayArticle.aspx?id=217

"New Black Holes Found in a Virtual Observatory," Universe Today, 28 May 2004. http://www.universetoday.com/am/publish/black_holes_virtual_observatory.html

"Mining the Mountain of Data from Deep Space," Space Today Online, 2004. http://www.spacetoday.org/DeepSpace/Telescopes/Observatories/VirtualObservatory/AVO.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)
GOODS	Great Observatories Origins Deep Survey
GRB	Gamma Ray Burst
GriPhyN	Grid Physics Network
HEASARC	High Energy Astrophysics Science Archive Center
HTTP	HyperText Transport Protocol
IPAC	Infrared Processing and Analysis Center (Caltech)
IRAF	Image Reduction and Analysis Facility (NOAO)
IRSA	Infrared Science Archive (IPAC)
ISI	Information Sciences Institute (USC)
ITWG	Information Technology Working Group (NASA data centers)
iVDGL	International Virtual Data Grid Laboratory
IVOA	International Virtual Observatory Alliance

JDBC	Java Data Base Connectivity (Sun, Inc., trademark)
JHU	The Johns Hopkins University
MAST	Multimission Archive at Space Telescope (STScI)
MB	megabyte
MOU	Memorandum of Understanding
MWG	Metadata Working Group
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputer Applications
NOAO	National Optical Astronomy Observatories
NPACI	National Partnership for Advanced Computational Infrastructure
NRAO	National Radio Astronomy Observatory
NSF	National Science Foundation
NVO	National Virtual Observatory
OAI	Open Archives Initiative
OASIS	On-line Archive Science Information Services (IRSA)
OGSA	Open Grid Services Architecture
OIL	Ontology Inference Layer
OWL	Web Ontology Language
PB	petabyte
PMH	Protocol for Metadata Harvesting (of OAI)
Q	quarter
QSO	Quasi-Stellar Object
RC	Replica Catalog
RDF	Resource Description Framework
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
WSDL	Web Services Description Language
XML	Extensible Mark-up Language
2MASS	Two-Micron All Sky Survey