Quarterly Report
January-March 2004

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

IVOA

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

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

Period covered by this report: 1 January—31 March 2004
Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

Following the successful science demonstrations at the January AAS meeting (reported on in the previous Quarterly Report) attention turned back to development and refinement of core system elements. Progress was made in the following technical areas:

- Registry: V1.0 of the Resource Metadata definitions has been forwarded to the IVOA Executive for promotion to Recommendation, the highest level IVOA document. Publishing registries at NCSA, Caltech, CDS, and AstroGrid have been successfully harvested into a master registry at JHU/STScI. The registry schema (the XML encoding of registry metadata) is now being reviewed, as many experienced developers have found it to be unduly complicated.
- A more modest but tractable revision of Unified Content Descriptors (UCDs) has been completed.
- First drafts of the Quantity and Observation data models have been released and will be refined prior to the May 2004 IVOA Interop Workshop.
- Extensions to the Simple Image Access Protocol are being discussed, and a first draft of the Simple Spectral Access Protocol is nearing completion. Both will be discussed at the May 2004 IVOA Interop Workshop.
- The Astronomical Data Query Language is advancing rapidly. OpenSkyNode development is also moving forward, with experience gained in implementing additional (.NET-based) SkyNodes for the HST observation catalog and the GALEX object database.

NVO team members submitted a proposal to the TeraGrid for 257,000 computational service units (1 service unit = 1 CPU-hour). The proposal was successful.

Proposals were submitted to NSF and NASA for financial support for the NVO Applications Software Development Summer School, which will be held in September at the Aspen Center for Physics. NSF funding has been secured.

Following the advice of the NVO Advisory Committee, the Executive Committee is establishing a Science Steering Committee to help us set science priorities and select additional demonstration projects.

The IVOA Executive created Applications and Theory Interest Groups. These groups will have their first meetings at the May Interop Workshop.

Carol Christian is assuming the position of Education and Public Outreach Coordinator for the NVO project. Carol has extensive experience in developing EPO programs and working with diverse partners.

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 has been scheduled for April 29-30, to be hosted by NCSA in Urbana, Illinois. The US NVO project will also host this year's major IVOA Interoperability Workshop, to be held May 24-28 at Harvard.

1.2 Science
In its December 2003 report, the NVO Advisory Committee strongly encouraged the formation of an NVO Science Steering Committee, and efforts are currently underway to constitute this committee. The mandate of the SSC would be to assist the NVO in formulating policy issues that affect the scientific capability of the NVO, both short term and long term. In addition, it is expected that the SSC will be deeply involved in monitoring the implementation of these policies and in providing essential advice during the course of this implementation. Because enhancing the scientific capability of US astronomy is the basic goal of the NVO, this committee is clearly of key importance. At this time a preliminary list of committee members has been identified, and recruitment is in progress. The committee will be chaired by the NVO Project Scientist.

1.3 Technical (including standards, configuration management)
R. Williams was asked by the IVOA Executive to lead a VO architecture review. Working with the chairs of the IVOA technical working groups, Williams will assure that all major components of the VO framework have now been identified and that their relationships are understood.

An infrastructure has been set up on the NVO web site that allows team members to upload and maintain project descriptions. This will help to distribute the responsibilities of web site maintenance.

NVO has led in the promotion of the Resource Metadata standard to the status of Recommendation, the highest level in the IVOA document standards process.

1.4 Financial
The overall project remains in solid fiscal health. Spending for the past quarter remained steady compared to the previous quarter, at $662,842. Our underrun of $1.2M will continue to be reduced as we continue project work at or near current levels well into Year 4.

1.5 International coordination/collaboration
NVO participation in IVOA coordination and collaboration remains high. Members of the Executive Committee attended the January meeting of the IVOA Executive in Garching, Germany. 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 has assumed leadership of a new IVOA Interest Group in VO-based applications. International cooperation is strong and growing, with renewals or increases in funding for several VO partners and the addition of a 14th VO organization, VO-Hungary, to the IVOA.

2 Science Requirements

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

The past quarter saw an increase in the level of effort in integrating the activities in theoretical astrophysics into the NVO. In addition to maintaining communication with the currently interested theorists performing large scale N-body calculations, additional progress was made in strengthening the links to theory activities in the international arena. The International Virtual Observatory Alliance (IVOA) has established a theory "interest group" at the last IVOA Executive Committee meeting, and US participation in this group is being encouraged. In addition to recruiting the active N-body simulation theorists, efforts will be made to include theorists working in Large Scale Structure, gravitational lensing, and hydrodynamic and MHD simulations. Collaborations with observers involved in large-scale surveys are also essential, and these connections will be pursued as well.

2.2 Requirements analysis

No activity this quarter.

2.3 Demonstration definition and review

No activity this quarter.

3 System Architecture

3.1 System design, components; relationships to Grid components

A document summarizing the IVO architecture has been developed by R. Williams. The architecture is based on the development of services to support interoperability across existing image catalogs, image archives, access portals, and grid resources. The components are similar to those proposed for NVO:

- Portals – provide common interface to multiple sky surveys, composition tools, display mechanisms, and query mechanisms
- Services layer – provide unifying services for connecting resources to portals. Three approaches are used, based on http Get/Put, SOAP, and WSRF protocols.
    - Registry service – based on OAI protocol for harvesting metadata about each service published into the IVO
    - SIAP, SSAP, Cone Search, VOTable – data access services for small sets of files
    - Grid Services for bulk processing and large scale analysis – visualization, cross-match, source detection, data mining

- Digital library – support metadata registries based on UCD semantics. This includes a MySpace temporary storage area for managing data.
- Grid Middleware – support for workflow pipelines, virtual data, authentication, authorization, distributed data management, distributed job execution.
- Resources – includes databases, storage systems, CPUs, networks.

ISI has worked on hardening Grid software in support of Montage and Pegasus, and has continued development of a Metadata Catalog Service. ISI has been running large region mosaics that have not been done before. As a result of the code stressing, several bugs and weakness were found in the Montage application code, the Pegasus code and related software. Upgrades have been made to Montage, Pegasus, Condor/Condor-G, and are proposed for the Replica Location Service.

The Astronomy Working Group of the Global Grid forum met in Berlin and discussed the robustness of Grid software, the experiences within the IVO community on attempts to use Grid software, and set criteria for more extensive use. The chairs of the Working Group are N. Walton and M. Ohishi. Equivalent GGF working groups were identified for each of the IVOA working groups. Two additional IVO working groups should be created to track the GGF efforts in workflow systems and preservation environments.

The consensus on use of Grid software is that two approaches were appropriate:

- Use Grid functions at individual sites along with GSI and GridFTP as pervasive services, but use IVOA technology at higher levels in the software architecture.
- Encapsulate Grid functions as web services. An example is Atlasmaker, which uses the TeraGrid to generate mosaics, but is accessed as a web service.

The Open Grid Services Infrastructure in being modified to conform to the Web Services Resources Framework. WSRF is the standard for managing stateful services that the grid community, W3C, and commercial vendors have agreed to support. The changes are:

- WSRF does not use service factories to create service instances.
- WSRF allows an additional level of indirection between the name of the service and the location where it is executed.

We expect WSDL to continue to be the description language in which services are described and expect the WSDL interfaces provided by OGSA-DAI to persist. The technology that is used to support the services will be modified to conform to WSRF.

For the next GGF meeting, the Astronomy Working Group will conduct an initial data grid assessment. And Newcastle will implement a SkyServer portal for searching for white dwarfs, based on grid services.

A Metadata Catalog Service for data intensive applications is being implemented by ISI, based on the OGSA-DAI framework (http://www.ogsadai.org.uk). Details about the new version of MCS can be found at http://www.isi.edu/~deelman/MCS.

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.

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

The TeraGrid is scheduled for production use in April 2004. The request for a TeraGrid allocation, submitted by R. Williams, A. Connolly, and J. Gardner, was approved. The allocation will be used to support installation of the NVO testbed on the TeraGrid, replication of NVO selected surveys onto TeraGrid storage systems, execution of Atlasmaker to create standard image projections for surveys, execution of the Montage mosaicing software, fitting of quasar spectra, development of n-point correlation functions of galaxies, and development of a Cosmic Microwave background grid, for a total of 257,000 service units. Participants in the proposal will be given accounts on the TeraGrid.

A collaboration with the ANU Supercomputer Facility is underway to replicate the MACHO collection onto the TeraGrid. S. McMahon at ANU is coordinating access to the data. The Storage Resource Broker data grid has been installed and tested at ANU. ANU is now able to make the MACHO data available to the SRB data grid from their mass data store. The data is being converted to the "world coordinate system" (WCS) along with some other data conversion and addition of some metadata. The data will be provided once the conversion is completed. We have the opportunity to provide access to the existing data sooner, if desired. ANU will probably start by serving a subset of the data to see how the transfers would impact their server and if there are any other issues.

A collaboration with S. Levine at USNO got set back a bit at the beginning of the year when they had a multi-disk RAID failure (the effects of which they only recently finished repairing). SDSC has provided a disk for starting the replication process via data transport by grid bricks.

The Remote Object Management Environment (ROME) has been developed to provide a stateful interface to computational intensive applications initiated through a web interface. The ROME components can be subdivided into three types:

- A set of JAVA objects which reside in any standard EJB container, communicate with any JDBC-compliant DBMS, and implement the "business logic" associated with managing processing requests, user identities, "processor" identities (see below), and messages flowing from the running applications and the users.
- "Processors" that directly manage the execution of end applications (CGI-like programs with the addition of optional intermediate messaging). Only one instance of a processor has been written: a standalone multi-threaded JAVA application. (This is not to be confused with the EJB work above: these processors run in a simple JAVA VM environment on any platform and contact the ROME EJB infrastructure remotely in search of tasks to run)

- A task monitoring tool that contacts "registered" applications (programs listening on a socket) and passes messages concerning applications (using a variety of filters to determine which applications and which messages). The ROME task monitor is a stand-alone JAVA GUI application to act as such a tool.

The first two classes, mostly EJBs and Servelets, employed the commercial EJB container WebLogic, and in this quarter have been ported to the open-source EJB container JBOSS. The third class has been developed at IPAC this quarter and runs under WebLogic and JBOSS. Informix is being used as the JDBC DBMS, primarily to avoid bottlenecking problems under load. All developer testing of these classes has been completed.

In order to properly test ROME under operational conditions, IPAC has written a test plan, and has begun implementing it. IPAC is setting up a test bed that runs collection of real world applications with varying run times (2MASS mosaicing taking minutes to days; database queries taking seconds to hours). Using a simple set of "client" programs, IPAC will simulate the submission of requests at various rates (up to whatever is the maximum – presumably limited by DBMS throughput) and monitor their execution with regard to reliability (no jobs lost or unprocessed), load balancing across multiple processors, etc. This testing activity will begin shortly and is scheduled to take approximately three months.

3.3 Digital library integration

R. Williams has continued work with our AVO collaborators to develop UCD1+, a more modest revision to the UCD naming system than had been previously considered. UCD1+ uses a familiar name-space format, with period-separated fields, and a simplified hierarchical structure. Comparisons between UCD1 and UCD1+ are available at http://vizier.u-strasbg.fr/UCD/lists/ucd1-ucd1p.txt
and
http://vizier.u-strasbg.fr/UCD/lists/ucd1p-composed.txt .

4 Registries

In the last quarter, two areas of effort related to registries cut across several of the sub-work packages. First, members of the NVO registry team (M. Graham, G. Greene, W. O'Mullane, and R. Plante) have been collaborating with K, Benson of the AstroGrid Project on an IVOA working draft for Registry Interfaces (components of which are discussed below). Second was the demonstration of the Carnivore publishing and searchable registry (described in Section 4.5).

4.1 Resource metadata

R. Hanisch moderated the IVOA review of the "Resource Metadata" standard document and edited it accordingly; as a result the document was elevated to the highest standard level, an IVOA Recommendation (the first of its kind). R. Plante submitted Identifiers 1.1 (a minor revision) for IVOA review; its elevation to Proposed Recommendation is imminent, at which time it will deprecate version 1.0.

4.2 Resource metadata schema

R. Plante is leading work on a revision of VOResource, the XML Schema based on the Resource Metadata document. This process began with the generation of a document (submitted as an IVOA Note) entitled "Lessons Learned Using the VOResource XML Schemas in the NVO" (http://www.ivoa.net/Documents/latest/RMExp.html) that discusses the major issues that can make using the XML schemas difficult. The next step is a virtual workshop (to be held in April) to address these issues with IVOA partners. The goal is to rapidly test solutions to these issues and present them at the IVOA interoperability meeting in May.

We feel it is vital to address the use-ability issues surrounding VOResource this spring. This will allow us to establish a stable Registry environment for the NVO Summer School. Since an important goal of the Summer School is to get developers from outside to use the NVO infrastructure, it is likely that we will become a bit more entrenched in the schema due to the large number of existing resource descriptions in use. This will make a substantial change in the schema model more difficult.

4.3 Publishing and harvesting protocols

C. Cowart (SDSC) and R. Plante (NCSA) have been working to define a Web Service version of the Protocol for Metadata Harvesting (PMH) from the Open Archives Initiative (OAI; http://www.openarchives.org). The goal is two fold. First, we hope to provide a prototype standard that is as universally useful to the digital library community as the original OAI-PMH standard; we plan to submit the Web Service definition (WSDL document) back to the OAI community for consideration (and further refinement) as a standard. Second, we wish to define a specific use of this proto-standard for VO registries. Plante is contributing the Harvesting section of the IVOA Registry Interfaces working draft, which is based on web service-enabled OAI.

The Carnivore Registry system also addresses issues of publishing (see Section 4.5). As part of the effort to deploy this system, Carnivore developer M. Graham worked with other registry maintainers to address discrepancies in the use of OAI.

4.4 Query protocols

The Registry team is now beginning to address the issue of standard query mechanisms for Registries as part of our work on the IVOA Registry Interfaces working draft. First, we have defined the notion of a simple keyword search and how it is implemented. W. O'Mullane has implemented this feature in the full, searchable registry at JHU/STScI. Second, we have developed a preliminary approach to more complex searching. It features an SQL-like search constraint against an XML component referred to by a simplified XPath. Other querying options, including ADQL and XQuery, are possible, as well. The Carnivore Registry system also addresses issues of searching with its support for XQuery (see Section 4.5 for details).

4.5 Replication, synchronization, maintenance, revision control, and curation
M. Graham (Caltech) demonstrated a working version of Carnivore, a publishing/ searchable registry built on an XML database. This gets around the difficulty of storing the object-oriented XML model in a relational database. It also allows us to experiment with XQuery and XPath as registry search mechanisms. At the moment, no serious performance issues have been encountered. On the publishing side, it makes use of XForms to generate custom fill-in forms on the fly based on the XML Schema for the metadata. In collaboration with R. Plante (NCSA), he is prototyping expanded in-line documentation capabilities drawn directly from the schema.

5 Data Models

5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality)
J. McDowell chaired a DM technical meeting in Garching to discuss the Observation data model with AVO team members M. Louys, D. Giaretta, G. Lemson, and F. Bonnarel. A draft document on the Observation data model was released to the WG in March after extensive email discussions following the Garching face-to-face meeting. The document describes the division of observation metadata into data, characterization and provenance and discusses each component in moderate detail. Further work on Observation is required, particularly on integrating the Space-Time coordinate work.

5.2 Low-level (measurement, quantity, uncertainty, relationship)
B. Thomas and E. Shaya (Raytheon) together with P. Dowler (CVO) and D. Berry (Starlink/UK) held extensive e-mail discussions and telecons this quarter. A document with a draft definition of the Quantity model and its serialization was released to the DM WG in March. The Quantity model describes a generic container for data, from simple metadata equivalent to a FITS keyword up to arrays with multiple world coordinate axis mappings. We have defined a layered design with BasicQuantity (single value), CoreQuantity (simple array of values, listed or algorithmically defined), and StandardQuantity (with array axes and mappings on the axes). However, work is still needed on elaborating the Mapping and the Accuracy (error/uncertainty) models that Quantity uses. A draft document on Mapping is currently being circulated in the Quantity subgroup and should be sent around the WG mailing list in mid-April.

5.3 Descriptors and ontologies (UCDs)
Recent discussion has centered on the use of updated UCDs in the SIAP protocol.

5.4 Space-Time and regions
A. Rots has made progress converting the STC schema to using substitution groups, rather than choice groups. Since the STC metadata are now appearing on the radar screens in more areas of development, there have been more substantive discussions on the design in this quarter, leading to some generalizations and simplifications. A full working draft is expected by the next team meeting.

Both J. McDowell and A. Rots are working on documents to more fully describe the STC design. A. Rots has prepared a UML representation of the STC schema, and extended the STC model to include general coordinates rather than just the specialized space, time, Doppler, and spectral coordinates. The next step is to integrate the STC model into the appropriate places in the Observation model.

5.5 Standard schema
Recent work has focused on the relationship between UCDs – used as a data dictionary for astronomy concepts and UTYPEs – proposed as a vocabulary for expressing data model schemata in the VOTABLE format. J. McDowell has presented a proposal for how UTYPEs might be used in SIAP to express data model structure in the SIAP return VOTABLE.

The Quantity draft document includes a proposed XML serialization and is accompanied by XML schemata and instance examples developed by B. Thomas. These are now under review and should form the basis for a general approach to non-VOTABLE XML serializations for data models.

T. Budavari and colleagues have implemented a spectral schema based on the proposal worked out by D. Tody and J. McDowell at the Baltimore team meeting.

6 Data Access Layer

6.1 Data access services (catalog, image, spectrum, time series, visibilities, …)
Catalog access. An initial version of the ADQL and OpenSkyNode technology developed by W. O'Mullane and others at JHU is undergoing initial trials, primarily by R. Plante and by the AstroGrid group in the UK. These efforts are focused primarily on refinement of the SQL and XML versions of ADQL and on developing a Java toolkit for SkyNode. D. Tody, M. Ohishi, and W. O'Mullane have discussed requirements on ADQL for use in DAL, in particular the need to encapsulate ADQL and to retain the capability for simple parameter-based queries in the DAL services. Our intention is to proceed with the early trials within the VOQL working group, then agree in the May working group meeting upon a plan for producing a DAL version of SkyNode for catalog access.
Image access. A second version of the Simple Image Access (SIA) interface is planned for this summer. This will upgrade the interface to integrate the standards developed over the past year and add some new capabilities based on our experience using the first version of the interface. Work this past quarter has focused on specifying these new interface capabilities. Thus far these include the following:

- Image provenance and identification (D. Tody, R. Plante). This is an application of the work on resource identifiers in the registry group to dataset identification in the DAL. A global dataset identifier, similar to the ADEC identifiers used for electronic publishing, is assigned by the dataset creator and used to uniquely identify archive data products. A separate registry identifier may identify the dataset uniquely within a particular archive or namespace. Global dataset identifiers are preserved if a dataset is published into a new collection, i.e., replicated.

- Logical name proposal (R. Williams). Candidate images described by the SIA query response may optionally be assigned a logical name to describe each image. Images that have the same logical name are assumed to be in some sense the "same" image, with the other image metadata being used to determine how the images differ. For example, the same image might be available in both the FITS and JPEG formats, or the same field might be available in several different spectral bands.

- UCD normalization and introduction of UTYPE (R. Williams, S. Derriere, J. McDowell, D. Tody). The temporary UCDs used in the initial SIA interface will be replaced by a combination of conventional UCDs, used to describe the meaning of each interface element, plus the newly introduced UTYPE tags, used to directly identify interface elements. Following a great deal of discussion about the relative merit and usage of UCD and UTYPE, a first cut at a uniform set of UCD and UTYPE values has been made.

- Image ranking (A. Rots). The data provider or service implementer may optionally assign a "score" to each image in the SIA query response, indicating how good a match each candidate image is to the query. The client can use this ranking, with or without analysis of other image attributes, to help select the "best" images from the list of candidate images returned. This is analogous to the default rank ordering seen in a Google query.

Some work has also been done on uniform image characterization, however this needs to be based on a standard data model and seems unlikely to develop fast enough to make it into the next version of SIA.

Spectrum access. We are working towards having a draft specification of the Simple Spectral Access (SSA) interface by mid-May, suitable for trial implementations and for discussion in working group meetings at the end of May.

Since this will be the first of the DAL interfaces that is based upon a formal data model, there has been some question how to coordinate the work between the DM and DAL groups and what documents to generate. The decision was to have two documents, one documenting the SSA data model and the other the DAL interface which implements this data model. The DM document will define the SSA data model, standard metadata to describe the model, and a default serialization in XML. The DAL document will define the query, the query response, and the representation of data returned by the interface. Currently we are looking at VOTable, FITS, native XML, and possibly simple text as output formats for spectral data.

Time series. The current plan remains to address time series data with the SSA interface. The general SED data model includes both spectral and time information. A time series is a projection of the general SED model at a constant spectral index.

Event and visibility data. D. Tody is fostering an effort within NRAO and ALMA to define a science data model (SDM) for visibility data. There will probably be separate models for interferometry data and single dish data. For each SDM there will be a default export data format (EDF) in which data is archived and made available for data mining via the VO. The EDF will be based on some combination of FITS, XML and VOTable. This effort is being coordinated with similar efforts by P. Lamb (CSIRO) and A. Richards (Jodrell), and with the DM group.

6.2 Data representation (VOTable, etc.)

How best to represent science data in the VO is still a controversial issue. This continues to be discussed at length in the VOTable, DM, and DAL forums without a clear consensus. We have general agreement on the following points:

- All data objects internal to the VO should have a formally defined data model.
- Data models should be defined independently of data representations. Multiple representations of the same data are permissible.
- XML in some form should be supported for data representation, although this should not be the only form of representation.

There appears to be a consensus on the following points although there may still be some controversy:

- The specification of a data model should include a standard schema and serialization. Data representations are not required to use either explicitly, but should be defined in terms of these, ideally with an explicit transformation, to allow data objects to be extracted from some arbitrary representation so that the objects could be verified with a schema or instantiated with some class code.
- FITS should continue to be used for transmission of bulk binary data, at least for the foreseeable future. In general we probably do not want to use FITS for metadata transport except as necessary to self-document a binary data element.
- VOTable should continue to be used with evolutionary enhancements, at least until there is some clearly better XML-based alternative.

The use of standard table mechanisms for the transmission of data objects containing tabular data remains controversial. Some believe that only a native XML encoding should be used, to make it easier to implement Web services and to make such data easier to process with XML tools. Others believe that VOTable (or some comparable standard table mechanism) should be used for tabular data in order to allow generic table-based tools to be used upon such data. Some believe that an "open" and easily extensible document-centric approach using a generic container to hold multiple component data models is required for datasets which are too complex to describe with a single standard data model.

6.3 Framework (mediators, components)

D. Tody and colleagues have begun an effort to design a scalable data analysis framework to link VO with general astronomical data processing and aid reuse of technology and actual software between the two areas. Within NVO this would provide a reference framework for implementing server-side data access services and for client-side development of analysis software that would use the VO services. The client-side version would be packaged for desktop use on a user's workstation as a Virtual Observatory Workbench, providing an integrated environment for both data analysis and user development of analysis software that would integrate both VO and non-VO components and services. The same computational framework would also be used for data processing, e.g., to pipeline process instrumental data to produce archive data products conformant to VO standards.

It is expected that most of the resources for developing this software will come from outside NVO, but it is critical that the work be coordinated with VO. It is premature to say who will be involved in this project from outside the NVO or what role they might play.

The basic architecture is a component framework, with most of the astronomical data processing functionality being implemented as components. Components could be used either individually, integrated into any execution environment, or via the reference framework. While the focus of this project is to define a standard component interface and implement the framework, most of the actual functionality would come from wrapping existing astronomical software as components. The major platforms targeted by this project are user workstations and clusters. Integration with the grid could occur either at the top, where a framework-hosted service could be used within a grid dataflow, or at the bottom, where data access could directly access the data grid infrastructure or invoke remote VO services. Ideally we would be able to dynamically deploy astronomy components to any node on the grid.

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

J. Good and others at IPAC have implemented SIA services for IPAC image collections including 2MASS. All of the major data centers within the US now provide some degree of support for data access via NVO.

T. McGlynn and others at HEASARC have tested many existing SIA services as part of testing the Data Inventory Service. NVO needs a more formal way to register and test new DAL services.

7 Query Language

Work is progressing on ADQL and OpenSkyNode. An informal meeting was held at JHU March 11 about the Microsoft SkyNodes in existence. GALEX, GSC, and HST Sky-Nodes were implemented. The main purpose of the meeting was to brief M. Westhead (Edinburgh e-science center) on the SkyNode technology. His group is trying to implement OpenSkyQuery on the Grid infrastructure.

7.1 Low-level: Astronomical Data Query Language

ADQL 0.7.1 specification and schema (XSD) were released early in the quarter, engendering substantial comments from the IVOA community. In general people found the schema a little more verbose than they expected. The schema was not documented and had some redundant tags that came from making a parse tree of the SQL statement.

There was also a discussion about Boolean logic and how databases handle such expressions.

The specification document raised fewer comments but definitely requires more work to make it more of a specification and less of a requirements document. The version of SQL (e.g. SQL92) needs to be tied down and the BNF (Backus Naur Form) provided for the string form of ADQL. The BNF has been posted on the IVOA TWiki at this stage and was an omission for the specification. There remains contention as to why we need both an XML and String representation of ADQL. As the group is fairly well divided on this, it seems the only approach is to keep both. Since ADQL is based on SQL the BNF should form the main definition of ADQL. XML Schema cannot express language in a manner similar to BNF. Hence the XML Schema should be able to represent any valid string from the BNF. Words to this effect will be placed in the new ADQL specification.

The simplification of the schema became a major issue. M. Hill (AstroGrid) proposed SADQL (Simple ADQL – not "unhappy" QL!). In general the group was not in favor of this as it was not rigorously produced in the manner ADQL 0.7.1 was. A simplified version of the posted schema was offered by E. Shaya. JHU finally responded with ADQL0.7.3, which was generally accepted by all.

There was a detailed discussion on the mailing list regarding how closely the XSD of ADQL should match the String representation. For Example, T. Linde (AstroGrid) suggested allowing Boolean operators such as AND to have infinite arguments. This would effectively allow us to think of AND as a prefix operator rather than an infix operator, hence we could have (AND a b c). This is rather nice but is incompatible with SQL, which only allows (a AND b AND c). The general consensus of the group was that the ADQL schema should stick close to SQL, thus the cardinality of operators such as AND should remain as 2.

R. Williamson produced XSL style sheets to convert ADQL to strings for a given database implementation. These were quickly modified when new schemas were produced.

Finally, some parsing and serialization problems were encountered with ADQL 0.7.3 forcing another minor revision. This will not yet be widely publicized until we are certain all kinks have been ironed out.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

A good deal of work has gone into the OpenSkyQuery portal at JHU. This has been hampered somewhat by trying to reach agreement on the ADQL specification. ADQL forms an integral part of the new system, hence changes to the schema impact both portal and nodes.

The nodes also require integration of VOResource and VOTable schemas. In essence this is the culmination of many specifications in the VO community. Integration of these schemas in a usable SOAP-based system has not been easy. At present we cannot get the portal going owing to serialization problems with VOTable, and possibly VOResource. Undoubtedly these issues will soon be resolved; they are, however consuming more time than might have been expected.

A rudimentary site is starting to take shape and may be ready for the next team meeting. NCSA is planning to be the first builders of a JAVA OpenSkyNode and have been prototyping with the schemas and WSDL as they are being developed. There have been several discussions between NCSA and JHU. We will meet prior to the spring team meeting to work on this.

7.3 High-level: Complex queries
No activity this quarter.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)
Web Services are being used increasingly in the NVO project. Web service interfaces are available to the JHU/STScI resource registry and are used routinely by the Data Inventory Service. The JHU spectral data browser also publishes full web services interfaces for all database queries and analysis functions. The project is developing a web services-based interface to the Open Archive Initiatives (OAI) protocol for metadata harvesting.

8.2 Grid Services (OGSA)
The full specification of Grid Services is moving from the Open Grid Services Infrastructure (OGSI) to the Web Services Resource Framework (WSRF), and NVO projects are waiting for the specification to be clarified and its implementations to be come more robust before making software development commitments. The Grist project at Caltech is building grid services and prototyping on the NSF TeraGrid. A significant barrier in this activity is that the security infrastructure is not well defined or implemented. These services include source extraction (SExtractor), and image processing through scaling and dynamic range reduction.

8.3 Computational resource management
The Caltech Grist group is working with the Triana package for building a dataflow of SOAP web services that is operating on the TeraGrid. These dataflows will be used to build astronomical pipelines for analysis of the Palomar-Quest sky survey. The ISI group continues to improve workflows for the Montage astronomical image mosaicing application (see WBS 3).

8.4 Virtual data
The Atlasmaker image mosaicing application is running on multiple TeraGrid machines and generating resampled images. A Virtual Data system is attached to the application, which examines each new file to see if it can be a source for some target; if all sources are available, the target is generated. In this way several derived data products are automatically built and cached, including scaled images, range-reduced images,

composited images, as well as federated images from Principle Component Analysis. Currently the SDSS and 2MASS surveys are being injected into the system.

8.5 Application and service integration with Grid

The NVO produced a proposal to the NSF TeraGrid for significant resources to support an initial selection of applications, including a mosaicing portal, image federation (hyperatlas), deep spectral analysis, data mining the Sloan catalogs, and cosmological simulation. The application has been approved, and we expect to be running the applications on TeraGrid in 2004Q2.

9 Applications

A major accomplishment in this area was the initiation of an Application Interest Group as a general IVOA forum for users to discuss issues in the use of VO development and use of VO end-applications. The charter for the group was given as:

> This interest group is intended to support developers and users of Virtual Observatory applications and provides a forum to

- announce the release of software tools and libraries for use by other developers and the public
- discuss the use of existing tools and libraries
- suggest enhancements or upgrades to existing tools and
- develop consistent approaches to software release throughout the IVOA.

Also, an NVO proposal for TeraGrid resources (R. Williams, PI) was successful. Application areas include atlas image generation, image mosaicing, star formation in galaxies, quasar spectrum fitting/modeling, the N-point correlation function of galaxies, and the cosmic microwave background. We are currently awaiting access to the TeraGrid resources in order to port these applications.

9.1 Data location services

Work continued on the Data Inventory Service, in particular, to use SIA services more fully. In the current release, only the first image returned by a service is available for use. In the next release scheduled for early in the coming quarter, all of the images returned by an SIA service will be available for browsing and us by the user. This release will also generate quick look images when none are provided by the SIA service. The current SIA protocol provides no standard mechanism for associating quick-look and scientific images, and a variety of ad hoc approaches are used in practice.

The Data Inventory Service is being used by the community at a rate of a few requests each day.

9.2 Cross-correlation services

A number of new SkyNodes have been implemented, including the HST observation catalog and the GALEX mission database. A total of 15 major survey catalogs are now available for cross-correlation.

Related activity: 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.

9.3 Image combination, registration

Beta testing of Montage at ISI continues. This work was supported in part by the NVO and in part by Montage.

9.4 Visualization tools and services

L. Dobos (JHU) has released a filter profile service that allows users to search for and to store spectral response functions. The Filter Profile Service for VO is an integrated database and web service designed specially for the Virtual Observatory. Users can register their own filter profiles and search the existing database, which consists of almost 100 filter profiles of the main astronomical instruments. The database can be accessed via a web interface, as well as via XML web services. It supports different formats for downloading profiles: XML, VOTable, plain ASCII and CSV.

9.5 Theory

P. Teuben presented the globular cluster simulation VO-Theory demo at the AAS in Atlanta (see FY2004 Q1 report). The web page for the demo remains active (http://bima.astro.umd.edu/nemo/tvo/nvodemo2004/), and has been upgraded with new features that have been discussed at the MODEST meeting in Geneva (Jan 10-15), which Teuben attended right after the AAS. One particular feature the observers requested is to add the isochrones, and of course more models.

9.6 Statistical analysis

Work progressed on the complementary project VOStat, which will be demonstrated at the next AAS meeting. VOStat will provide statistical services and is being designed to be fully VO compliant. Demonstrations will make use of the SDSS spectrum server from Budavari et al., and the NVO TeraGrid time allocation.

An N-point clustering Web service developed at the University of Pittsburgh and interfaced to the SDSS data archive is currently being tested. Initial results appear consistent with published correlation functions. A web form interface to the service is also being developed to enable simpler access for the community.

Progress continues at U. Pittsburgh on implementing a parallel/grid algorithm for calculating the correlation function of galaxies. The prototype version was successfully implemented and a new version that builds upon the initial application is underway. This will address the issues of how to work in a high latency environment (i.e. the grid) and how to balance the workload across many hundreds of processors.

9.7 Data mining, outlier identification

Work has started at U. Pittsburgh on an anomaly finder that is interfaced to the Spectrum Services as developed for the NVO. The initial implementation uses a Java application to identify supernovae present within SDSS spectra. This will be extended to identify spectra that do not fit into the low dimensional manifold that exists for most galaxies.

The data mining applications which will use the TeraGrid under the award discussed above are a parallel/grid implementation of an n-point correlation function code (that calculates the 2 point and 3 point redshift space and projected correlation functions) and a frequentist approach to analyzing CMB maps through the use of multiple realizations of the CMB power spectrum.

Related Activity: The NASA AISR funded ClassX project continued to use NVO-developed technologies to help in its efforts to classify all detected ROSAT sources.

9.8 Interfaces to/from legacy software systems
No activity this quarter.

10 Community Engagement

10.1 Documentation
No activities this quarter.

10.2 Web site
A major effort to update the NVO project web site has been started, with two key goals:

1. Make it easier for astronomers/end-users to locate research tools.
2. Allow NVO developers to post and modify descriptions of their projects directly to the web site.

The new focal point of the web site will be astronomical applications, described in jargon-free language. We will also make it easy for more technical audiences to learn about the NVO framework technologies, for data providers to learn about how to publish their data collections, and for education and outreach providers to find tools and services of interest to them.

10.3 Technical training initiatives
We submitted proposals to NSF and NASA for the sponsorship of an NVO Applications Software Summer School, to be held September 13-17, 2004, at the Aspen Center for Physics in Aspen, Colorado. We developed a web site describing the Summer School began accepting applications from prospective students. Information is available at http://www.us-vo.org/summer-school/.

10.4 Advocacy
We have begun work on a long-term plan for NVO operations, updating various white papers and management plans that were written prior to the inception of the NVO framework development project. The new plan will describe a transition from IT research and development to ongoing operations and system enhancement, with new budgetary estimates based on the past 2+ years of experience in national and international VO development. Our goal is to complete this long-term plan in the next quarter and to make it available to the US funding agencies.

11 Education and Public Outreach

The EPO component of the project was basically inactive this quarter as the Executive Committee sought ways to revitalize this work. F. Summers (STScI) stepped down as EPO Coordinator. Following extensive discussion in the Executive Committee, we have decided to replace Summers with Carol Christian. Carol is the former head of the Office of Public Outreach at STScI and has extensive experience in developing innovative projects with external partners. We are looking to Carol to help us define an NVO EPO program of modest scope and high visibility.

11.1 Strategic partnerships

No activity this quarter.

11.2 Formal education

No activity this quarter.

11.3 Informal education

No activity this quarter.

11.4 Outreach and press activities

No activity this quarter.

11.5 Technical development

No activity this quarter.

Activities by Organization

Caltech–Astronomy Department

- Prepared data for the Palomar-Quest OpenSkyNode
- Developed science cases for the Grist data mining project.

Caltech–Center for Advanced Computational Research (CACR)

- Continued work on "Carnivore" registry for NVO resources, with publishing, qrich query interface, and OAI harvesting.
- Built new web portal for NVO software, science, and data services. These projects will be maintained by NVO personnel, rather than a single webmaster.
- Prepared NVO proposal to TeraGrid.
- Wrote "Architecture of the IVO" document.
- Worked on updates to the SIA protocol with correct UCD and with Logical Name attributes.

Caltech–Infrared Processing and Analysis Center (IPAC)

- Began a follow-up to the brown-dwarf pilot project.
- Ported ROME to the JBOSS application server; completed developer testing; designed formal test plan.
- Acted as a test bed for registering services.
- Developed generalized VOTable output for queries to NED.
- Registered 11 SIAP services for IRSA and NED data.

Canadian Astronomy Data Centre/Canadian Virtual Observatory

The transfer of Sloan Digital Sky Survey Data Release 1 tables to DB2 was completed in February. Data engineering has begun to transform SDSS quantities into the CVO data model for querying. The SDSS Data Release 2 is looming on the horizon, and efforts will be undertaken to obtain this metadata as soon as possible.

Data engineering of the CFH12K mosaic camera data has begun in anticipation of ingestion into the CVO database. These high-resolution wide-field images will be valuable new content for CVO.

In light of the collaboration that has been undertaken with the Human-Computer Interaction (HCI) group in the Department of Computer Science at the University of Victoria, the CVO's web interface will soon have a different look. Input from an undergraduate HCI class has provided an independent view of the interface and has provided a number of suggestions that will be incorporated into the tool. Additional image display and graphics functionality will also be added.

ROSAT RASS and pointed-observation catalogues have now been ingested into the CVO catalogues. The collaboration with GAVO, however, will continue with work on a probabilistic catalogue matcher and the development of an object catalogue.

The 2QZ source catalogue will be ingested into the CVO Prototype after modifications are made to accommodate multi-valued entries (in this case the multiple fluxes included in a spectrum. The next major datasets planned for inclusion are the 2dF Galaxy Redshift Survey, the 6dF survey, and 2MASS.

Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt)
CMU and Pittsburgh are reorganizing their NVO efforts and funding. Work progressed on the complementary project VOStat, which will be demonstrated at the next AAS meeting. VOStat will provide statistical services and is being designed to be fully VO compliant. Demonstrations will make use of the SDSS spectrum server from Budavari et al., and the NVO TeraGrid time allocation.

The npt clustering web service (interfaced to the SDSS data archive) is currently being tested. Initial results appear consistent with published correlation functions. The web service is now being interfaced to a web form to enable a simpler access for the community. Work has started on an anomaly finder that is interfaced to the Spectrum Services as developed for the NVO. The initial implementation uses a java application to identify supernovae present within SDSS spectra. This will be extended to identify spectra that do not fit into the low dimensional manifold that exists for most galaxies.

The NVO proposal for TeraGrid resources edited by Connolly and Gardner with Williams (Caltech) was successful. The PITT/CMU applications comprise image and spectral analysis as well as classic grid-based data mining routines. The spectral and image analysis routines involve a spectral energy distribution fitting technique applied to multicolor images of galaxies and a genetic algorithm for fitting spectral features to QSO spectra. The data mining applications are a parallel/grid implementation of an n-point correlation function code (that calculates the 2pt and 3pt redshift space and projected correlation functions) and a frequentist approach to analyzing CMB maps through the use of multiple realizations of the CMB power spectrum. We are currently awaiting access to the TeraGrid resources in order to port these applications.

Progress continues on implementing a parallel/grid algorithm for calculating the correlation function of galaxies. The prototype version was successfully implemented and a new version that builds upon the initial application is underway. This will address the issues of how to work in a high latency environment (i.e. the grid) and how to balance the workload across many hundreds of processors.

Fermi National Accelerator Laboratory (FNAL)
V. Sekhri worked primarily on deploying new SDSS databases and interfaces this quarter. The bulk of this work involved assisting loading and configuring the Catalog Archive Server (CAS), which is the backend MS SQL database of the SDSS SkyServer web service. For Data Release 2 (DR2), which was released to the public in March, he shipped a copy from Fermilab to JHU (via Fedex). He also carried out some robustness tests using queries for the cluster finding algorithm from a previous NVO demo and comparing MS SQL to MySQL databases under heavy load (with MS SQL proving to be

more robust). He also deployed a test version of DR3, which is planned to be released in October. DR3 is about 35% larger (3 terabytes). This work was done in close collaboration with A. Thakar at JHU and other SDSS members.

V. Sekhri is working with Argonne Laboratory and NCSA to provide a gsiftp and a gsirft (reliable file transfer protocol) interface to the DR2 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. (WBS 3).

High Energy Astrophysics Science Archive Research Center (HEASARC)
HEASARC personnel suggested the formation of a group devoted to the discussion of end-user applications in the VO. This has resulted in the creation an IVOA Applications Interest Group of chaired by T. McGlynn. A charter for this group has been published and the TWiki site has begun to develop a set of links to assist in developing and using VO applications.

Development of the DIS focused on properly integrating SIA services. This included dealing with special issues for specific services, e.g., patches for violations of the SIA protocol or for finding the quick-look data associated with the FITS data, and working to handle archive SIA services that return multiple images for a given region of the sky. Users will be able to select one or more images from any of these archives and send them to the DIS supported analysis tools or retrieve them. Generation of quick-look data for services that do not provide it has also be incorporated into the development version of the DIS.

The HEASARC developed and circulated a list of currently registered SIA services along with issues found with them. The ensuing discussion unearthed a lack of coordination among the registries so that several services that had been registered at the NCSA had not been properly harvested at JHU. This was quickly addressed by the parties concerned.

A long discussion between the HEASARC and other institutions regarding the HEASARC's OAI-delivered registries has been underway. Some OAI harvesters have difficulties in understanding the format produced at the HEASARC but this seems to be a problem with lack of specificity in the standards rather than an actual error in the HEASARC produced dataset.

HEASARC personnel attended the weekly NVO metadata telecons.

Johns Hopkins University
T. Budavari continued to work on Simple Spectral Access protocol. Working with L. Dobos, Budavari implemented a draft SSAP specification on top of the existing Spectral Database.

V. Haridas revised ADQL numerous times in line with comments on the mailing list, and kept the translator in synch with all changes. Worked on OpenSkyNode implementation with latest ADQL (0.7.4).

W. O'Mullane has worked on the OpenSkyQuery portal with V. Haridas and N. Li (web pages, testing, serialization problems with XML). Work continued with G. Greene (STScI) on the searchable registry. Worked with A. Szalay and G. Greene to enable Full Text Searches in the registry (very fast keyword searches). Continued to provide feedback on nascent registry specification document.

N. Li worked on an OpenSkyQuery portal. Changed the portal to use newer ADQL. In particular, an XPATH parser for ADQL was developed to do query rewrites for nodes; this makes the portal less dependent on ADQL. Li is also working on a schema browser for nodes.

G. Fekete, A Szalay, and W. O'Mullane have worked on the HTM and region support. A. Szalay and J. Gray have been devoting time to dealing with polygons and complex shapes in SQL Server. A new HTM paper in the form of a Microsoft Technical report has been drafted.

M. Nieto-Santisteban has been working on a SQL version of an algorithm for finding galaxy clusters developed by J. Annis (Fermilab). The original application runs on a 50-node grid cluster and takes about a week to run. We are trying to improve the speed of this application by first using SQL and databases technology instead of files, and second, by partitioning the data so we can run the code on a SQL Server cluster in parallel. Work done during the last few weeks shows that the SQL implementation of the algorithm (A. Szalay with help from J. Annis) running on a single server has the same performance as the 50-node grid implementation. Nieto-Santisteban has been working on the parallel approach using a 3-node SQL server cluster. So far, tests have shown that by using partitioning we can reduce the execution time of the SQL version of the application by half. We intend to use our experience on this application to extend it to other surveys and to implement a cross-matching service among NVO SkyNodes.

J. Raddick has been in touch with his teaching contacts to see what teachers would like from NVO as part of the EPO effort. He is also working on some EPO outreach cases based on this feedback and experience with SkyServer. SkyServer has delivered about 100 pupil-years of teaching at this point.

Contact is also being maintained with S. Parastatis who is heading up an effort to built large Grid Type applications using only web services (http://www.neresc.ac.uk/ws-gaf/). We hope to leverage this project for security in some of our NVO services.

M. Nieto and G. Fekete attended the Global Grid Forum in Berlin.

Microsoft Research
J. Gray worked closely with the Johns Hopkins team on a spatial data library for the NVO. This included extensive revisions of the HTM code and the routines that use them, development of a generic library to handle data regions, Boolean algebra on the regions, and point containment and overlap predicates. This work is embodied and documented in SkyQuery and in the SkyServer We also worked to bring the SDSS Data Release 2 online, helped the Quest group with some database issues, and helped the AstroGrid team develop their OpenSkyQuery portal.

National Optical Astronomy Observatories (NOAO)
NOAO work this quarter was limited to the participation of the Project Scientist in the regular Executive Committee telecons and his other activities. No technical work was carried out.

National Radio Astronomy Observatory (NRAO)
No report received.
In the past quarter D. Tody contributed to NVO in the following areas:

- Lead for the NVO data access layer and chair of the IVOA DAL WG.
- Further development of the DAL services, primarily SIA and SSA.
- Many discussions regarding data representation issues.
- Review of the proposed VOTable 1.1 standard.
- Worked with E. Greisen (also at NRAO) to make FITS spectral WCS compatible with VO and SSA.
- Scalable data analysis framework research and design.
- Participated in IVOA system architecture WG.
- Feedback to the VOQL WG on DAL requirements.
- Fostered development of data models and formats for visibility data.

NRAO hosted R. Hanisch for a Virtual Observatory colloquium on February 27. This was very well attended.

J. Ulvestad, following a meeting with D. DeYoung, has organized an effort within NRAO to develop a pilot project to produce images from the VLA archive. Several meetings have been held to discuss various strategies, but a final plan has not yet been reached.

An "E2E architecture and oversight committee", co-chaired by W. Cotton and D. Tody, has been set up within NRAO to design and implement an end-to-end 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.

Work continues to replicate the NRAO archive to NCSA. NCSA has delivered a several terabyte disk box to be used for bulk data replication between NRAO Socorro and the NCSA. D. Tody, with J. Benson (NRAO), and R. Plante and A. Kemball (NCSA), has drafted a plan for a simple disk or network-based scheme for replication and data delivery to users. This will use the new VO identifiers for global dataset identification and grid-based technology for data transport and secure messaging. A final plan still needs to be generated before implementation can go forward.

Raytheon/ADC

- Developed XML Schema in accordance with DM Quantity working draft.
- Reworked space-time-coordinates (STC) schema so that it uses Quantity.xsd as a basis and can be incorporated into quantities.
- Worked on editing the DM Quantity working draft and Observation WD.
- Developing Instrument Description Schema.
- Provided comments and input on Resource draft and STC coverage terms.
- Studying XForms specification.
- Running tests on Berkeley native XML Database.
- Attended Metadata Working Group telecons.

San Diego Supercomputer Center

SDSC has participated on the grid architecture specification for the IVO, participated in the Astronomy Working Group at the Global Grid Forum, and continues to support the replication of astronomy sky surveys onto TeraGrid storage resources. The sky surveys currently available on the TeraGrid are supported by the Storage Resource Broker data grid federation system. This version of the software supports federation of independent metadata catalogs, each of which manages a local data grid. The SRB installation at ANU for the Macho collection is based on data grid federation technology, allowing researchers in Australia and the US to implement separate catalogs for each replica of the collection. The federation technology supports access constraints between the independent data grids. Surveys currently on the TeraGrid include:

- SDSS DR1 release
- 2MASS DR2 release
- DPOSS

Discussions continue on replicating the USNO-B and Macho collections onto the TeraGrid.

Smithsonian Astrophysical Observatory

SAO continued to lead the Data Model design (WBS 5.0). J. McDowell chaired a DM technical meeting in Garching to discuss the Observation data model with AVO team members. This meeting led to progress on the Observation data model. A draft document on the Observation data model was released to the Working Group in March. B. Thomas and E. Shaya (Raytheon) together with P. Dowler (CVO) and D. Berry (Starlink/UK) held extensive email discussions and telecons this quarter. A document with a draft definition the Quantity model and its serialization was released to the DM Working Group in March. A draft document on Mapping is currently being circulated in the Quantity subgroup and should be sent around the WG mailing list in mid-April. A. Rots has made progress converting the STC schema to using substitution groups, rather than choice groups. Both J. McDowell and A. Rots are working on documents to more fully describe the STC design. J. McDowell has presented a proposal for how UTYPEs might be used in SIAP to express data model structure in the SIAP return VOTABLE.

A. Rots attended the AAS meeting in Atlanta, January 2004. J. McDowell attended the IVOA Exec in Garching, January 2004 and presented a WG report.

Space Telescope Science Institute
STScI developed SIAP (Simple Image Access protocol) services for HST Preview images, GOODS, and UIT data. They were publicly released by Web announcement via the MAST interfaces and also registered in the NVO STScI/JHU Registry. The HST observation catalog and GALEX object catalogs were made available as SkyNodes.

Registry work has continued with progress in support of the Data Inventory Service. Several Web services were written which support the OAI standard adopted by the IVOA for registry harvesting. Work is still in progress to support the complete OAI implementation.

A final revision of Resource Metadata was completed and submitted to the IVOA Executive for review. We expect RM V1.0 to be adopted as an IVOA Recommendation, the highest level of IVOA document.

IVOA document preparation procedures were reviewed and a guidelines document was prepared.

Proposals were prepared and submitted to NSF and NASA for support of the NVO Applications Software Summer School.

The agendas for the spring NVO team meeting and May IVOA Interop Workshop were planned and distributed. R. Hanisch prepared a draft IVOA Roadmap update.

R. Hanisch attended the AVO (European VO) Science Working Group meeting and IVOA Executive Committee meeting in Garching, Germany, 27-29 January 2004. Hanisch also gave a colloquium above NVO at the National Radio Astronomy Observatory Array Operations Center in Socorro, NM (27 February).

United States Naval Observatory
Following last quarter's disk crash, software to convert the binary form of USNO-B1 from Binary to ASCII was reconstructed, and a start made on regenerating an ASCII version of the catalog suitable for ingestion by commercial databases.

University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA)
R. Williamson and R. Plante implemented a SkyNode interface to the ADIL using Java, Apache Axis, and PostgreSQL. As part of this work we developed a set of XSL style sheets that can be used to convert ADQL into standard SQL. On top of these, extensions can be built to adapt to particular a particular database or table implementation. We also developed a general way of plugging in XSL transformations directly into Axis. We intend to create a SkyNode server kit that will package these developments and make it easier for providers to bring up a SkyNode using Axis.

R. Brunner has begun work on a SkyNode application that collects remote catalog data for correlation with large catalogs housed at NCSA. The model employed by this application, in which data is gathered to a central location for correlation allows him to experiment with custom cross-matching algorithms.

In collaboration with C. Cowart of SDSC, R. Plante has been prototyping a Web Service version of the Open Archives Initiative interface. He is also working on a definition of a standard Harvesting interface based on OAI.
R. Plante has also been prototyping new XML modeling techniques intended to address many of the usability issues raised with regard to the VOResource schema.

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

- Continued to create NVO web services for the analysis of lightcurves. In this quarter we developed an algorithm to detect extra solar planets by analyzing light curves. We implemented an early version of the algorithm and we are testing it on the Macho lightcurve DB. For the NVO implementation the planetary transit detection must be done on the fly and thus a serious consideration is given to the performance of the program.
- Created a design prototype for a computational platform in which astronomers can create pipelines that contain existing surveys, VO data centers and Grid computational facilities. Under this framework (Virtual Astronomical Pipelines) one can create a complex workflow. We are working on extending the Penn software PlatformD and already existing frameworks to integrate NVO VOTables and OGSI web services.

University of Southern California (USC/ISI)
The work at ISI has been focused on 1) hardening of Montage and Pegasus, and 2) development of a Metadata Catalog Service.

We have been running large region mosaicing that has not been done before. As a result of the code stressing, 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 addition of another step in the abstract DAG that is yet to be added into mDAG service. The critical load problem on our gridftp server led us to update the transfer client in Pegasus. Also, several feature improvements are currently pending to satisfy large Montage workflow. The condor/ condor-G have been updated due to the problems we have discovered during the stress test, i.e., the periodic release/held problem and the postscript problem in the rescue DAG. We also have highlighted problems in the Replica Location Service from our testing, which has lead to eventual bug fixes.

MCS is a Metadata Catalog Service for data intensive applications. It allows the user to record metadata attributes for data objects and to discover data objects based on these attributes. We have implemented a new prototype version of MCS that uses the OGSA-DAI framework (http://www.ogsadai.org.uk) as the middleware. The earlier version of MCS was implemented as a web service. Using OGSA-DAI, we have developed MCS as a grid data service that implements few extra activities specific to MCS. We have implemented authorization in this version of MCS. Authentication is handled by the OGSA-DAI framework. We have also streamlined the Client Application Programming Interface for MCS and added a few new API calls. Additional information about the new version of MCS can be found at http://www.isi.edu/~deelman/MCS.

University of Wisconsin
No report received.

Publications

"There Goes the Neighborhood: Relational Algebra for Spatial Data Search", Jim Gray, Alexander S. Szalay, Gyorgy Fekete, Wil O'Mullane, Arnold H. Rots, Aniruddha R. Thakar, MSR-TR-2004-22, March 2004

"Extending the SDSS Batch Query System to the National Virtual Observatory Grid", Maria A. Nieto-Santisteban, William O'Mullane, Jim Gray, Nolan Li, Tamas Budavari, Alexander S. Szalay, Aniruddha R. Thakar, MSR-TR-2004-12, February 2004

Virtual Observatory Articles in the Popular and Technical Press

None this quarter.

Acronymns

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

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