

Quarterly Report
October 2003—December 2003

Building the Framework for the
National Virtual Observatory

NSF Cooperative Agreement
AST0122449



INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE



| | |
|--|-----------|
| Executive Summary | 1 |
| 1 Management | 2 |
| 1.1 General (planning, reporting, communications, team meetings, etc.) | 2 |
| 1.2 Science | 4 |
| 1.3 Technical (including standards, configuration management) | 4 |
| 1.4 Financial | 5 |
| 1.5 International coordination/collaboration | 5 |
| 2 Science Requirements | 5 |
| 2.1 Usage scenarios for all areas of astronomy research, including theoretical simulations | 5 |
| 2.2 Requirements analysis | 5 |
| 2.3 Demonstration definition and review | 6 |
| 3 System Architecture | 6 |
| 3.1 System design, components; relationships to Grid components | 6 |
| 3.2 Computational facilities (processing, bulk data storage, network access, security, authentication) | 8 |
| 3.3 Digital library integration | 8 |
| 4 Registries | 8 |
| 4.1 Resource metadata | 8 |
| 4.2 Resource metadata schema | 9 |
| 4.3 Publishing and harvesting protocols | 9 |
| 4.4 Query protocols | 10 |
| 4.5 Replication, synchronization, maintenance, revision control, and curation | 10 |
| 5 Data Models | 10 |
| 5.1 High-level (image, spectrum, time series, event lists, visibilities, catalogs, simulations, data quality) | 10 |
| 5.2 Low-level (measurement, quantity, uncertainty, relationship) | 11 |
| 5.3 Descriptors and ontologies (UCDs) | 12 |
| 5.4 Space-time and regions | 12 |
| 5.5 Standard schema | 13 |
| 6 Data Access Layer | 13 |
| 6.1 Data access services (catalog, image, spectrum, time series, visibilities, ...) .. | 13 |
| 6.2 Data representation (VOTable, etc.) | 14 |
| 6.3 Framework (mediators, components) | 15 |
| 6.4 Data provider/consumer implementations and end-to-end testing | 16 |
| 7 Query Language | 16 |
| 7.1 Low-level: Astronomical Data Query Language | 16 |
| 7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode | 16 |
| 7.3 High-level: Complex queries | 17 |
| 8 Web and Grid Services | 17 |
| 8.1 Web Services (SOAP, WSDL, etc.) | 17 |
| 8.2 Grid Services (OGSA) | 18 |
| 8.3 Computational resource management | 18 |
| 8.4 Virtual data | 19 |
| 8.5 Application and service integration with Grid | 19 |

| | | |
|-----------|--|-----------|
| 9 | Applications | 20 |
| 9.1 | Data location services | 20 |
| 9.2 | Cross-correlation services..... | 20 |
| 9.3 | Image combination, registration | 21 |
| 9.4 | Visualization tools and services..... | 24 |
| 9.5 | Theory | 24 |
| 9.6 | Statistical analysis..... | 25 |
| 9.7 | Datamining, outlier identification..... | 25 |
| 9.8 | Interfaces to/from legacy software systems | 26 |
| 10 | Community Engagement | 26 |
| 10.1 | Documentation | 26 |
| 10.2 | Web site | 26 |
| 10.3 | Technical training initiatives..... | 27 |
| 10.4 | Advocacy | 28 |
| 11 | Education and Public Outreach | 28 |
| 11.1 | Strategic partnerships..... | 28 |
| 11.2 | Formal education | 29 |
| 11.3 | Informal education | 29 |
| 11.4 | Outreach and press activities | 29 |
| 11.5 | Technical development | 29 |
| | Activities by Organization | 30 |
| | Caltech–Astronomy Department | 30 |
| | Caltech–Center for Advanced Computational Research (CACR)..... | 31 |
| | Caltech–Infrared Processing and Analysis Center (IPAC)..... | 31 |
| | Canadian Astronomy Data Centre/Canadian Virtual Observatory | 32 |
| | Carnegie-Mellon University/University of Pittsburgh (CMU/UPitt) | 33 |
| | Fermi National Accelerator Laboratory (FNAL)..... | 33 |
| | High Energy Astrophysics Science Archive Research Center (HEASARC) | 34 |
| | Johns Hopkins University | 34 |
| | Microsoft Research | 35 |
| | National Optical Astronomy Observatories (NOAO)..... | 35 |
| | National Radio Astronomy Observatory (NRAO) | 35 |
| | Raytheon/ADC..... | 36 |
| | San Diego Supercomputer Center..... | 37 |
| | Smithsonian Astrophysical Observatory..... | 37 |
| | Space Telescope Science Institute | 38 |
| | United States Naval Observatory..... | 39 |
| | University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications (UIUC/NCSA)..... | 39 |
| | University of Pennsylvania..... | 39 |
| | University of Southern California (USC/ISI) | 40 |
| | University of Wisconsin | 41 |
| | Publications | 43 |
| | Virtual Observatory Articles in the Popular and Technical Press | 46 |
| | Acronyms | 47 |

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

Period covered by this report:
Submitted by:

1 October 2003 - 31 December 2003
Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

The primary activities in the first quarter of the NVO project's third year were oriented toward a new set of science demonstration projects that were successfully shown at the Atlanta (January 2004) American Astronomical Society meeting, and toward several key technical development areas: resource registries, data models, a spectral data access protocol, and increasing use of grid-based computational resources. A new collaboration on public outreach in which a next-generation web browser will provide easy access to NVO-based content is taking shape, and initial results were also shown at the Atlanta AAS. Following a successful VO Software Tutorial held just prior to the annual Astronomical Data Analysis Software and Systems (ADASS) Conference in Strasbourg, France (October, co-organized with IVOA partners), planning has begun on an NVO Summer School in which we will provide detailed instruction to astronomers outside the NVO project in how to use NVO toolkits and build NVO-enabled science applications. A redesign of the NVO web site has also been started, changing the orientation to the astronomer end-user rather than the NVO development team. We held our annual review meeting with the NVO Advisory Committee, with initial informal feedback being very positive. (A formal report is expected in late January.) NVO team members continue working with the International Virtual Observatory Alliance, leading several IVOA Working Groups, actively participating in all others, and helping to organize the important Interoperability Workshops in which technical discussions covering all aspects of VO development take place.

At the start of Project Year 3 we undertook a major revision of the project work breakdown structure, resolving overlaps and ambiguities in work areas and aligning NVO development activities more closely with the IVOA. Lines of responsibility have been clarified and reporting has been simplified.

The project is financially stable and operating at a full-staff level.

Activities by WBS

1 Management

1.1 General (planning, reporting, communications, team meetings, etc.)

The NVO project held a team meeting in Baltimore on 9-10 Dec 2003, hosted by The Johns Hopkins University. The focal points of this meeting were

- Final planning for the science demonstrations to be shown at the January 2004 meeting of the American Astronomical Society
- Recap of progress on major development areas (registry, data models, data access layer, UCDs)
- Preparation for the subsequent Advisory Committee meeting
- Review and agreement on new work breakdown structure

All matters were discussed and addressed successfully.

We met with the NVO Advisory Committee on 11-12 Dec 2003, also in Baltimore at The Johns Hopkins University. Committee members in attendance included Dr. Sidney Wolff, NOAO (chair), Dr. Gerry Gilmore (Institute of Astronomy, Cambridge, UK), Dr. John Huchra (Harvard University), Dr. Sid Karin (San Diego Supercomputer Center), Dr. Rob Kennicutt (University of Arizona), Mr. Carl Lagoze (Cornell University), Dr. Paul Messina (Caltech), and Dr. Eve Ostriker (University of Maryland). Dr. Tony Hey (University of Southampton, UK) was unable to attend this year. Dr. Eileen Friel attended from NSF. NASA HQ representatives were invited but were unavailable.

The NVO Executive Committee and other senior members of the project team presented a progress report to the Advisory Committee, and the Committee asked many follow-up questions. Following a two-hour executive session, the team was informally debriefed. Comments were very favorable, especially in the areas of standards development, international cooperation, and in our technical approach. Further attention needs to be paid to our EPO program, especially in establishing stronger partnerships and setting priorities. A full written report is expected by the end of January.

The new work breakdown structure can be seen in the outline of this report. After two years of development we found that our original WBS was too complex and had too many overlapping work areas to be tractable. WBS leaders and team members were not clear on which WBS to use when reporting work, and lines of responsibility were confused. We believe we have rectified this in the new WBS, in which the major work elements also parallel development efforts at the international level.

In compact form, the new WBS is as follows. Each primary element has a responsible person or group, shown in brackets. It is possible to draw a correspondence map between the old WBS and the new one, though it would have many intersecting lines and would be unlikely to provide much insight. We have taken care, however, to assure that all activities encompassed previously are represented in the new structure. Areas of

overlap remain, but they are now primarily matters for coordination rather than duplication.

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

9. Applications [T. McGlynn]
 - 9.1 Data location services
 - 9.2 Cross-correlation services
 - 9.3 Image combination, registration
 - 9.4 Visualization tools and services
 - 9.5 Theory
 - 9.6 Statistical analysis
 - 9.7 Datamining, outlier identification
 - 9.8 Interfaces to/from legacy software systems

10. Community Engagement [R. Hanisch]
 - 10.1 Documentation
 - 10.2 Web site
 - 10.3 Technical training initiatives
 - 10.4 Advocacy

11. Education and Public Outreach [F. Summers]
 - 11.1 Strategic partnerships
 - 11.2 Formal education
 - 11.3 Informal education
 - 11.4 Outreach and press activities
 - 11.5 Technical development

The Metadata Working Group continues to hold weekly telecons, in the past quarter focusing most on implementation of the first internationally interoperable resource registries. The NVO Executive Committee also meets weekly by telecon, reviewing management issues and priorities as matters arise. R. Cutri has been standing in for G. Helou as a senior IPAC representative on the Executive Committee while Helou attends to commissioning matters for the newly launched Spitzer Space Telescope (formerly SIRTf).

1.2 Science

The Executive Committee took an active role in the definition and oversight of the theory science demonstration, in which globular cluster simulations were compared to observed color-magnitude diagrams. D. De Young helped to set realistic requirements, and R. Hanisch worked closely with P. Teuben (U. Md.) on the implementation and preparation of the demonstrations to be shown at the Atlanta AAS.

1.3 Technical (including standards, configuration management)

Several NVO-led standards made their way through the IVOA standards process and have been promoted to higher levels of maturity. *IVOA Document Standards V1.0*, which describes the standards process itself, was promoted to the level of Recommendation, the highest standards level of the IVOA. *IVOA Identifiers V1.0* is now a Proposed Recommendation. The NVO project also discussed identifiers with the NASA Astrophysics Data Centers Executive Council (ADEC) and reached agreement that the IVOA identifiers can serve as bibliographic references to underlying datasets. *Resource Metadata for the Virtual Observatory V1.0* is a Working Draft, but should be advanced to the higher levels of standard in the coming quarter.

1.4 Financial

Overall project spending remains on track, at the planned level, with on-hand funding sufficient to sustain current levels of effort through the first half of FY05/Project Year 4. The Financial Supplement contains additional details.

As of 1 Oct 2003 NASA terminated the operations of the Astronomical Data Center (ADC) and its support contract with Raytheon ITC. NVO personnel formerly affiliated with Raytheon at the ADC have relocated: K. Borne to George Mason University and E. Shaya and B. Thomas to the University of Maryland. Raytheon has agreed to remain our contractual contact for this work, however, and will flow through funding to GMU and UMd. To establish new subawards directly with GMU and UMd would have entailed JHU overhead costs in excess of the available funds for labor.

1.5 International coordination/collaboration

NVO participation in the International Virtual Observatory Alliance remains strong. NVO personnel lead several IVOA Working Groups (D. Tody: Data Access Layer, J. McDowell: Data Models, R. Williams: UCDS, R. Hanisch: Standards and Processes) and are active in all others. R. Hanisch led the organization of the IVOA Interoperability Workshop that followed the October ADASS Conference in Strasbourg, France. The two-day workshop was attended by 120 people. Specific development activities with international involvement are described in the appropriate WBS below. The VO development effort is truly an international collaboration.

2 Science Requirements

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

In this quarter the NVO project focused on the area of theoretical simulation data. We selected a science demonstration project that would utilize NVO core services to enable a direct comparison of globular cluster simulations (including both dynamics and spectral evolution) with observational color-magnitude diagrams. This topic in stellar evolution and dynamics has appeal to a broad cross section of astronomers, and it provides a nice counterpoint to the usual cosmology based theory presentations. The globular cluster calculations follow both the dynamical and evolutionary history of every star in a globular cluster, and hence a direct comparison of the calculated and observed color-magnitude diagrams is possible. This in turn permits evaluation of both the initial conditions and the evolutionary path of model globular clusters to see which ones most closely reproduce the observed current C-M diagrams.

2.2 Requirements analysis

Current leading edge computing capability now permits detailed theoretical calculations of astronomical processes that incorporate enough physics to permit significant testing of models against observational data. The datasets produced by these calculations are comparable in size to the large-scale observational datasets, and one of the goals of the NVO from its inception has been to create a structure that will facilitate direct comparison of theory with observation.

The second effort in the science requirements area involved the implementation of simple and useful science applications to be disseminated to the US astronomical community by early 2004. In particular, attention was paid to the NVO Data Inventory Service, which is now in usable form and is available from the us-vo.org web site. Efforts were also initiated to make this and other NVO services more readily accessible to the astronomical community through improvement of the NVO web site and through the use of focus groups. These efforts gained valuable input from the NVO Advisory Committee at its December 2003 meeting, and it is anticipated that the coming months will see significant improvements in making the NVO more visible and more useful to US astronomers.

2.3 Demonstration definition and review

The demonstration was developed in large part by P. Teuben (U. Md.) and his collaborators, in consultation with D. De Young and R. Hanisch. The goal was to have this demonstration ready for presentation at the January 2004 AAS meeting, and this goal was in fact met. The next steps that follow from this demonstration are to develop more general tools and protocols for easy incorporation of theory datasets into the overall NVO structure.

3 System Architecture

3.1 System design, components; relationships to Grid components

The system architecture has been described in prior status reports. A layered software infrastructure is being developed, that separates access interfaces, from service registries, from data management and computation management. The NVO approach towards choosing grid components is strongly driven by risk mitigation, with alternate implementations being evaluated at each level of the software hierarchy. The implementations come either from other NSF ITR projects, or from standards efforts.

To track standards, the IVOA has initiated an *Astronomy Applications Research Group* in the Global Grid Forum. The intent is to promote interactions between the IVOA and GGF. This includes providing input to the GGF on the requirements of the IVOA community for grid and web services infrastructure, and providing evaluations of Grid performance and robustness. At the Global Grid Forum 9 meeting, a mapping was proposed between the IVOA assessments that are underway, and the provision of NVO requirements to the above working groups. The co-chairs of the working group are Nic Walton and Reagan Moore.

The alternate implementations being evaluated for the NVO grid components for each level of the software hierarchy include:

- Portals—web service interfaces to analysis procedures (OASIS, French Aladin, SkyQuery, the JPL YourSky, and the new Data Inventory Service from NASA Goddard.) In the long run, the grid community will provide portlet technology that simplifies the construction of portals. The challenge is to implement the same features provided by current portals such as SkyQuery with grid portlet technology to

test the functionality and robustness.

- Process management systems—data processing pipelines to create derived data products (Chimera, Montage, Matrix). The two approaches are application based workflow environments, and web services-based dataflow environments. This is discussed in section 8.3.
- Web services—uniform capabilities provided across NVO catalogs and image archives (cone search, VOTable catalog query, simple image access). We currently use WSDL services, but are evaluating OGSA based services. This is discussed in sections 8.1 and 8.2.
- Data access layer—management of methods on data encoding formats for access based on physical quantities (UCDs for semantics, VOTable transfer format, FITS files). The approach taken by NVO is very similar to that of the Planetary Data System. This includes the use of registries to define services, characterization of the data format and semantics through a resource service, and development of name spaces to define semantics, allowed ranges on attributes, and types of permitted operations on the data. The NVO should track the data access layer developments within the PDS. The Grid system that should be tracked is the Data Format Description Language Working Group of the Global Grid Forum, which is looking at the same characterizations of semantics, structure, and behaviors.
- Data grid—management of distributed collections, provision of logical name space for global persistent identifiers, and support for remote proxies (SRB, RLS/MCS/CAS/GridFTP). The Global Grid Forum is debating the mechanisms that should be used to support organization of distinguished name spaces for distributed data. Either service handles will be assigned, or Globally Unique Identifiers. Until this debate is resolved, the current data grid technology provided by the SRB is being used to replicate image archives onto the TeraGrid (NVO testbed). To minimize the impact of the GGF choice, a GUID has been implemented in the SRB to provide equivalent functionality.
- Computational grid—access to distributed compute resources (Globus toolkit, Condor). The Globus toolkit is undergoing a major evolution from version 2 to version 3. The NVO testbed is built on version 2 technology, while the OGSi architecture becomes more robust.
- Persistent archives—management of technology evolution (SRB). A Global Grid Forum information draft has been prepared on use of data grid technology to implement persistent archives. The draft is being reviewed by GGF. In the meantime, a second version of the draft is being reviewed by the NHPRC InterPARES 2 project, and a third version of the paper has been accepted for publication in the IEEE Symposium on Mass Storage. From these three efforts we expect a standard to start emerging.
- Astrophysics catalogs and image archives (SDSS, 2MASS, DPOSS)—efforts continue to port additional sky image archives onto TeraGrid resources. We have installed a node of the NVO data grid in Australia, using the latest version of the software (SRB 3.0.1), and will start replicating the MACHO data set onto the TeraGrid.

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

The major change in the data management environment has been the installation of SRB version 3.0.1 on the NVO testbed. This version supports federation of independent data grids, making it possible for multiple data grids to interact. Access and consistency constraints are defined for four distinguished name spaces: users, resources, files, and metadata. The federation environment has a wide range of possible approaches. The federation can be set up as peer-to-peer, with user controlled sharing of data between data grids, or as a hierarchical organization with data and metadata replicated into an archival grid. The DPOSS image archive has been ported onto SRB 3.0.1, and the SDSS R1 and 2MASS R3 releases are being installed on SRB 3.0.1. The Macho collection will be replicated onto SRB 3.0.1 to provide additional control mechanisms. The 2MASS R2 release will remain on SRB 2 data grid technology.

A second major development for the data grid community is the design and implementation of authority delegation systems. In collaboration with the UK Data Grid, a library is being designed to support the creation of a delegation of authority that can be used by the SRB data grid to access data within a user's account at a remote site. Since the NVO data grid stores data under NVO collection ownership, individuals can only register data into the system from the same site where the service request was issued. To support registration from a remote site, a mechanism is needed to delegate authority to the data grid to run under the user's ID at the remote site. This development will make it easier to manage the import of data in federated environments.

3.3 Digital library integration

The use of digital library technology to improve collection management is being tracked through additional projects with the digital library community. The technologies of interest are document life-cycle management systems such as DSpace, Fedora, and the OCLC document management systems. These systems impose curation and archival processing on data as it is being registered into a collection. Collaborations have been initiated with the University of California and MIT to evaluate these mechanisms.

A second effort is related to the organization of metadata attributes for transport or storage, through use of the METS Metadata Encoding and Transmission Standard. METS provides extension schema support for defining domain dependent attributes associated with administrative, descriptive, structural, or behavioral metadata. A reasonable goal is to use the extension schema to define the NVO attributes that are needed for characterizing records in NVO catalogs and images in NVO image archives.

4 Registries

4.1 Resource metadata

Following extensive discussions, both within the NVO project and with IVOA colleagues in the IVOA Registry working group, we were able to reach agreement on the core resource metadata concepts and their definitions. R. Hanisch led the development of the document *Resource Metadata for the Virtual Observatory V1.0*, and at the IVOA interop

workshop following the ADASS Conference in Strasbourg it was agreed to promote the document to the level of Proposed Recommendation. Based on this, the implementation of XML schemas derived from the basic concepts could proceed.

4.2 Resource metadata schema

At the IVOA Interoperability meeting in October, we agreed to “freeze” the Resource schemas in preparation for the second year demonstrations, and the accepted XML Schemas were posted to the IVOA web site. We set about updating all of our prototype registries to use these latest schemas. During this period, a number of minor changes of a “bug fix” nature were made. We successfully exchanged records using the new schema not only between US registries but also with the AstroGrid registry and CDS in Strasbourg.

While we were successful in demonstrating the use of Registries by the DIS, the schema upgrading process was not without difficulty. We are now assembling our lessons learned about using XML Schemas based on our experience. An overview of the difficulties include:

- Variation in tool support: most XML tools are not perfect in verifying schemas; thus, a number of “bugs” were undetected before the schemas were frozen. Furthermore, not all tools had good support for all of the features we were using.
- Many of the “ground-floor” developers found the schemas complex and thus difficult to produce valid instances.
- Our use of multiple namespaces is error-prone

We are now considering whether to address these concerns in another version in the next quarter.

We also found that even mildly complex XML schemas can be difficult to map to a relational model using traditional semi-normalized tables. On the other hand, XML Databases are not as mature as RDBMSs, and performance is a concern. A number of minor differences between the various registry still exist that can affect interoperability; these will need to be addressed in the coming months.

4.3 Publishing and harvesting protocols

For our DIS demo, we used the OAI harvesting interface to gather resource descriptions into the searchable registry. J. Lee and T. McGlynn created a new publishing registry at HEASARC. Through the OAI interface they published hundreds of catalog descriptions; this serves as a good example of supporting a large publisher. W. O'Mullane (JHU) and G. Greene (STScI) further developed the searchable registry, working closely with Lee and McGlynn on both harvesting from the HEASARC publishing registry and supporting searches by DIS.

At NCSA, R. Williamson and R. Plante concentrated on upgrading “version 1” of their VORegistry-in-a-Box, leaving version 2 until after the demos. Plante also worked with developers to identify compatibility issues between registry systems. At Caltech, M. Graham and R. Williams are now experimenting with implementing a searchable registry using an XML database.

As part of our prototyping, we were able to integrate with registries in Europe. As a result from a visit to STScI from K. Benson of the UK AstroGrid project, we were able to exchange records with the AstroGrid registry. We also were able to retrieve records from an OAI interface to the CDS holdings.

C. Lagoze, lead for the OAI project and member of the NVO Advisory Committee, has encouraged us to propose a Web Service version of the OAI standard. We expect to develop this in the next quarter.

4.4 Query protocols

For the purposes of the last round of demonstrations, we agreed that it would not be necessary to standardize on a registry query language or interface. However, this will need to be addressed in the next year.

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

Over the last two prototyping iterations, we have found that curation of registry information will likely be an important issue for the future. At the searchable registry at STScI, we found that it was necessary to hand-correct descriptions due to errors or critical omissions. Mechanisms for replication and synchronization of registries will need to preserve these corrections in conjunction with automated harvesting that could over-write them.

5 Data Models

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

Spectrum Model

D. Tody and J. McDowell are collaborating on a simple Spectrum data model to be used in the Simple Spectral Data Access protocol. We tentatively propose that:

- For spectral energy distributions, we represent photometry points as single-point spectra and SEDs as a collection of spectra, thus mixing continuous spectra and photometry in a single paradigm.
- Observation details are shared by each point in a single spectrum.
- An SED therefore consists of a number of Spectrum entries, each of which contains observation metadata and a set of one or more data points (and optionally a corresponding number of error values and flags)
- The error model to be used should support at least two-sided errors and upper limits, since these are important in SEDs.
- The model can be serialized in FITS as a binary table with one row per spectrum, using the variable-length array mechanism to accommodate spectra of different lengths.

The formal document has not yet been completed or circulated. A preliminary paper setting out ideas for the spectrum data model was presented by J. McDowell at ADASS and will be published in those proceedings.

Observation model

A technical meeting to reach consensus on an Observation data model was held in October in Strasbourg during the ADASS conference. Significant progress was made in reconciling the models proposed by different groups at the high level. A second technical meeting to prepare a document on the Observation model is planned for January 2004.

We identified several common areas in our different views of the Observation model.

- For array data such as images, the data itself will be represented by a Quantity object.
- A CoordinateSystem object is required (this is now to be incorporated into the Frame object of the Quantity model)
- Curation object will hold traceability information and be compatible with the curation data required by the Registry
- Coverage will summarize the bounds of the observation in at least the space, time and frequency domains.
- Provenance will combine information on the instrument and observation process, and the data acquisition process (the boundary between these being fuzzy for modern instruments). This model will include Observatory, Instrument, and Processing sub-models.
- Source, Target, Field will need to be modeled to describe the target of the observation.

The next step is to elaborate each of these models, with the goal of agreeing on a simple representation that will be suitably extensible in future.

XML Schema for astronomical objects

In a poster at the ADASS meeting, E. Shaya presented an initial concept for an all-inclusive astronomical schema that begins with VO: AstroObject/Universes/Large-Scale Structures/Galaxies/Clusters etc. down to meteorites. Each node contains as children the appropriate parameters and attributes (luminosity, mass, position, velocity etc.). This could be used to support XMLQuery in a distributed astronomical data system.

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

Work on a Quantity model is now well advanced with a subgroup consisting of J. McDowell (CfA), B. Thomas (Raytheon/U. Md.), M. Dittmar (CfA), E. Shaya Raytheon/U. Md.) and Canadian (P. Dowler) and UK (D. Berry) representatives iterating on a definition document. A Quantity model would define a standard method to connect numerical information with its errors, units, and physical semantics in a way that could be reused in the high level image and spectrum models. M. Dittmar and B. Thomas presented preliminary Quantity models at the ADASS meeting.

Following a technical meeting in December in Baltimore, we identified broad areas of agreement:

- A quantity model representing both single numerical values and arrays of values should provide the common interface to data items which may have units, errors and quality information as well as semantics provided by a UCD or equivalent; the units, coordinate system and UCD form a context which we will call a Frame and may be used separately from the values.
- When data values are provided in arrays, they may be associated with coordinate axes which themselves have Frames giving UCD semantics and units, and whose coordinate values may be defined algorithmically (by a WCS type function) or explicitly by enumerated values.
- The interface to the quantity model should support alternate Frames for both the Quantity values and (if present) the coordinate axes. For example, a Q instance may know how to present itself in several different coordinate systems.

However there were several remaining issues:

- Whether a coordinate axis on an image can be represented by a Quantity model (the axes and the Q itself have very similar properties)
- Whether a SimpleQuantity without errors, coordinate system etc. should be defined as a separate object which Quantity inherits from, or whether it is defined as a restricted interface to the Quantity object
- Whether the support for explicit values on coordinate axes can be provided without undesirable software overhead.

These issues are fairly minor and we hope to issue a draft IVOA Recommendation soon.

5.3 Descriptors and ontologies (UCDs)

The UCD2 proposal was presented at Strasbourg in October 2003, and got a mixed reception. The proposal (in the WG chair's opinion) is much more capable of describing the semantics of astronomical data; but some participants felt that it was too complicated and it wasn't clear how to apply it in practice. An enhanced UCD1 proposal is now being worked on as an interim solution. R. Williams (Caltech) has led NVO work on the UCDs. E. Shaya and B. Thomas (Raytheon/U. Md.) have presented a candidate OWL (Web Ontology Language) ontology for physical units.

5.4 Space-time and regions

The Space-Time Coordinate metadata schema's transformation from choice groups to substitution groups made substantial progress, although not as much as hoped because of other obligations. A proper UML diagram, independent of the XML schema is in an advanced state of completion. We are still planning to generate two or three simplified schemas that are upward compatible with the full schema. That upward compatibility is extremely important since it provides a mechanism to make defaults and implied assumptions explicit, which is a crucial requirement. The design and plans were presented to the NVO Advisory Committee.

General agreement was reached on shorthand notations with the people working on the VOQL. This issue is similar to the issue of the simplified STC schemas mentioned above. We have promised a separate design document for the Regions specification for the first quarter of 2004.

5.5 Standard schema

R. Plante (UIUC) has proposed style standards for XML Schemas to be used in the VO.

6 Data Access Layer

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

Catalog access. W. O'Mullane and others at JHU are developing an astronomical data query language (ADQL), which is SQL-like with astronomy-specific extensions for region-based queries and the like. A family of catalog query and cross-matching services is currently in the prototyping stage. The plan is to develop a new catalog access service for DAL to replace cone search, based on the SkyNode prototype being developed at JHU. ADQL will later be integrated into all the DAL services to provide an optional advanced query capability.

Image access. Version 1.0 of the Simple Image Access interface (SIA) has been in use for some time now. Improvements to the general VO infrastructure are required before a new version of SIA is warranted. Priorities for enhancement to SIA include the following:

- Image provenance and identification
- Characterization of the physical attributes of an image
- Improved support for image formatting options, e.g., templating
- Registry integration
- Service verification
- Error handling

R. Plante (NCSA) and others within the Registry Working Group have produced a draft standard for VO resource identifiers, which will serve as the basis for image identification metadata in SIA and other DAL interfaces. D. Tody (NRAO, DAL WG Chair) is working with the data models working group on the problem of dataset characterization, a key issue for DAL dataset queries and image characterization within SIA. F. Bonnarel (CDS) has identified requirements for improved image formatting options for SIA, and proposed an access reference URL templating mechanism as a possible means for addressing this problem.

Spectrum access. Following the ADASS Conference and follow-on Interoperability Workshop, draft specifications of a query interface, simple spectral data model, and FITS format for spectra were prepared as part of the work on the Simple Spectral Access interface (SSA). SSA will provide uniform access to both 1D spectra and SEDs. M. Dolensky (ESO) has written an internal draft specification for the SSA query interface. J. McDowell (SAO) has written an internal draft specification for the SSA data model. T. Budavari (JHU) has developed a spectral data service for JHU, including an experimental SOAP-based interface for spectral access. D. Tody has been working with the FITS WCS authors to ensure that the proposed FITS spectral WCS standard will be consistent

with SSA. At this point these documents are still internal to the SSA working group, and more work is required before they can be released for public comment or used for the first trial SSA interfaces.

Time series. Our thinking at this point is that time series can be handled as a specialization the general SED data model. A SED consists of a series of photometry points with both spectral and time attributes. A 1D spectrum is a projection of the SED model along the spectral axis at constant time. A time series is a projection of the SED model along the time axis at a constant spectral value. Further work is required to evaluate the characteristics of actual time series data before we can be sure this approach will work, but if so basing time series data access on SSA would allow time series data to be handled at little additional cost.

Event and visibility data. Little has been done yet about VO access for event data. Within the radio community, efforts are currently focused on data models and data formats for visibility data. ALMA (led by A. Wicenek of ESO) is attempting to use a combination of VOTable and FITS BINTABLE to represent the ALMA science data model. This is one of the earliest attempts we are aware of to attempt to use VOTable to handle large quantities of data.

The priorities for near-term development of DAL services include:

- Concluding the current work on spectral and time series data access and releasing the first version of SSA.
- Initial implementation of SSA services and end-to-end testing and tuning of SSA.
- Continuing work on the DAL infrastructure followed by an update to SIA once this new infrastructure is available.
- Specification of a new catalog access service to supplant cone search.
- Work on multi-protocol (Web Service) versions of the DAL services.

6.2 Data representation (VOTable, etc.)

The problem of how to represent astronomical data within the VO is a major issue that is just now starting to be addressed. We have made a good start with catalog data, which the current VOTable handles fairly well. Further work is needed on catalog metadata but this is more of a data-modeling problem than a data representation problem.

For astronomical datasets such as images, spectra, time series, etc., little has yet been done. SIA avoids the problem by using VOTable to convey standard metadata for images. The image data itself is returned using FITS, with no attempt to standardize the contents of the FITS file other than what is already provided by FITS. A hybrid scheme such as this is probably what is needed in the long run as well, using XML for complex structured metadata, and FITS for bulk binary data.

For SSA the data representation issue is something we can no longer avoid. There is no standard FITS representation for spectra, so to provide uniform access to spectra via SSA we need to define a standard data format for spectra. More precisely, we need to define a

data model for 1D spectra and SEDs, and a standard mapping of this model into several different storage formats, e.g., VOTable (or some other XML-based format), FITS, and simple text. J. McDowell (SAO) and D. Tody have defined a preliminary data model for 1D spectra and SEDs, and are currently working on mappings of this data model into XML (VOTable) and FITS.

Recent discussions in the IVOA VOTable Working Group have raised a number of issues that need further investigation.

- How should we model complex astronomical datasets such as a spectrum? Do we produce one model for the entire spectrum, or model the components of astronomical datasets and merely associate these components to model actual datasets?
- How to represent a data model in XML or in VOTable is an issue. For example, if we map the attributes of a data model to the fields of a VOTable how do we associate this storage representation back to the abstract data model? How do we do schema validation in such a case? Two alternate mechanisms have thus far been proposed. The UTYPE tag allows the fields of a table to be associated to the attributes of an externally defined data model. The VOTable "group" mechanism provides a mechanism for describing structure and relationships among subsets of the fields of a flat table.
- Whether to use generic containers like table to represent astronomical data objects is an issue. To merely pass an object across a wire, e.g., via a Web service, the simplest approach is to serialize the object as an XML entity, with a custom schema defined for the object. Mapping such an object into a table complicates a simple "object to wire and back to object" mapping, but allows generic table tools and libraries to be used to manipulate the data component of a dataset.
- How to handle bulk binary data is an issue. FITS can do this, but FITS alone does not go far enough, and FITS is not a good solution for handling complex structured information such as metadata. XML is much better for this purpose. The challenge will be to determine how to combine the two to provide both the flexibility and standard-based strengths of XML, plus the efficiency of binary format standards such as FITS.

The DAL, data model, VOTable, and other working groups within VO are actively researching all of these issues. The SSA interface will provide the first real-world test case for working out how to represent astronomical data natively within the VO.

6.3 Framework (mediators, components)

The initial DAL protocols are simple enough to be implemented "from the ground up" by service providers and client applications. As data access in the VO becomes more sophisticated this will no longer be practical; it will become necessary to provide reference-grade software implementing the data access protocols, both on the server and the client side. Integration and reuse of legacy data analysis software with the VO is also

a goal, for example to enable application of server-side data analysis components during data access. Further work on the DAL framework is not scheduled until Q2 2004.

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

Numerous implementations of the cone search and SIA services currently exist, along with registry integration and client applications to make use of these services. This software was most recently demonstrated by the NVO at the AAS in January 2004. An early test version of the SSA query interface has been implemented by ESO and ISO and will be demonstrated at the AVO demo in January 2004.

The current plans for SSA are to provide at least two initial test implementations of the service (at JHU and ESO), with at least one client implementation (probably the SpecVIEW Java client from STScI).

7 Query Language

The NVO project is concentrating on VOQL level 1 and 2. But some though is still being given to level 3.

7.1 Low-level: Astronomical Data Query Language

There has been a lot of activity on the VOQL mailing list. A new version of the ADQL specification (0.7) is now available and the SQL to ADQL translation service is up (<http://skyservice.pha.jhu.edu/devel/AdqlTranslator/ADQLTrans.aspx>). A web page also provides some sample queries at

<http://skyservice.pha.jhu.edu/devel/AdqlTranslator/Convertor.aspx> and <http://skyservice.pha.jhu.edu/devel/AdqlTranslator/AdqlToSql.aspx> going the other way. The changes were mainly syntactical and bring us in line with JDBC function naming conventions. Many of these issues were raised by IVOA colleagues from Astrogrid, C. Page and T. Linde.

7.2 Mid-level: VOQL and OpenSkyQuery/OpenSkyNode

There has been much discussion as to what the layers entail exactly. In general it is accepted that Virtual Observatory Query Nodes should at least support the Circle Region construct so they are at least as powerful as cone search. This will probably be answered best by having some prototype systems up and running and seeing what is useful. Some people worry about having a spectrum of nodes of varying ability; this, however is inevitable and most should be of some use even if they may not participate in any automated distributed application

The SOAP Question. There was a long mailing list discussion, prompted by J. Good, on whether we should be using SOAP or simple HTTP POST/GET. This was raised at the team meeting and a consensus was reached. SOAP is definitely desirable and what we should use for this task. A Simple HTTP POST/GET interface should be provided where appropriate. The .NET platform for example does this for free.

Open SkyQuery Portal. Work continues on OpenSkyQuery portal including specifications and a reference implementation. This will help to define what Full

SkyNodes actually need to implement to become useful in a federated sense. For basic SkyNodes the existing specification is adequate, i.e., the ability to select particular rows from a result and to project columns is already immensely useful. Reference implementations of a SkyPortal and SkyNode are being developed at Johns Hopkins. These nodes communicate using VOTable and ADQL. Many individuals (e.g., T. McGLynn, R. Plante, R. Williams, B. Berriman) at the team meeting expressed interest in making SkyNodes and being involved in the OpenSkyQuery portal.

International Sky Query Effort. A plan is underway to develop a Grid service version of SkyQuery. This is being driven by J. Gray and B. Mann (AstroGrid), together with members of the Edinburgh-based edikt and OGSA-DAI teams, which have developed the two reference implementations of the GGF Data Access and Integration Services (DAIS) specification. This project will port the existing SkyNode code and develop that into Grid services using the two OGSI-based DAIS implementations, as well as the non-OGSI Web Services Grid Application Framework (WS-GAF) devised by S. Parastatidis (UK e-science program). This project will remove the Microsoft specific features of the original SkyQuery system, and help drive the development of OpenSkyQuery, as well as producing valuable expertise in using Grid services. The JHU group will be closely involved in this work, to ensure the continued collaborative development of the OpenSkyQuery specification.

7.3 High-level: Complex queries

In a poster at the ADASS meeting E. Shaya presented an initial concept for an all-inclusive astronomical schema that begins with VO:AstroObject/Universes/Large-Scale Structures/Galaxies/Clusters/ etc. down to meteorites. Each node contains, as children, the appropriate parameter and attributes (luminosity, mass, position, velocity, etc). Since XMLQuery works on a specific schema at a time, this could be used to support XMLQuery in a distributed astronomical data system.

E. Shaya presented, in a poster, examples of a Problem Statement Language (PSL) form that would make it easier for scientists to submit complex queries by using a Pseudo Natural Language (PNL) and possibly XForms.

8 Web and Grid Services

8.1 Web Services (SOAP, WSDL, etc.)

Web services provide an appropriate level of access abstraction for interacting with NVO testbed resources. A graphic example of this is the migration of the NVO sky surveys onto version 3 of the Storage Resource Broker. The standard NVO service for Simple Image Access Protocol works on top of the new data grid technology without any changes. Web services have been implemented that support both data and metadata access, updating, and transport for the SRB data grid.

L. Dobos and T. Budavari at The Johns Hopkins University have been extending the browsing capabilities the VO Spectrum Services. The production version of the web site and services have been made public at <http://voservices.net/wave/> and demoed at the

AAS in January at the STScI, SAO, and IPAC booths. The enhancements include an advanced interface for overplotting more spectra in one figure as well as utilities for building quasar and galaxy composites on the fly. The site publishes half a million spectra (SDSS DR1 and 2dFGRS) that is accessed not only by the astronomer community but recently high school teachers also started to use it in education.

Web Services form an increasing part of the Virtual Observatory, and it is felt there are some basic interfaces which all of these services should provide. For service description, an extension of the WSDL protocol is proposed for increased semantic information. Services should also implement a “heartbeat” interface, with uptime, contact information, etc; another proposal is for a standard way to ask about various kinds of usage information of the service. For more information, see:

<http://www.ivoa.net/internal/IVOA/IvoaGridAndWebServices/StandardInterfaces-0.1.pdf>

8.2 Grid Services (OGSA)

The NVO, GEON, SCEC, and GriPhyN NSF ITR projects are evaluating implementations of OGSA-based services. At SDSC, the results from the multiple NSF ITR projects are being assessed for impact on the NVO testbed. An OGSA service for query and data retrieval has been implemented on top of the Storage Resource Broker data grid. The service is able to issue a query to the SRB metadata catalog, wrap the result in XML, and deliver the result back to the user. In this case the system encapsulates the return data within a SOAP message. A second project has evaluated the ability to port the WSDL services for manipulating data and metadata within the SRB data grid, to OGSA. The goal was to send data as attachments, independently of the SOAP messages. This is needed for large sky images. This project was unable to implement a successful OGSA service for sending attachments. The implication for NVO is that the OGSA environment is still in an initial development phase. The experiences learned from these other ITR projects will help control expectations about the OGSA environment and its suitability for NVO services.

8.3 Computational resource management

We have developed large parts of a test bed needed to support ROME testing. We have developed a multi-threaded job manager, written in Java, which automates submission of large numbers of jobs to ROME. This is a contributed effort, supported by the NASA/IPAC Infrared Science Archive. We have also built the JBoss application server at IRSA, and have begun to port ROME to it.

Computational resources can be managed either through Grid Toolkit application execution environments such as Chimera, or through web service dataflow environments such as Matrix. The two approaches focus on different aspects of the resource management challenge. One can apply controls to the execution of applications, or one can apply controls to the flow of data between processes. The Matrix environment is being developed as part of the GriPhyN and SCEC NSF ITR projects. The Matrix web service approach supports a subset of the W3C XQuery language. The current system provides a service to issue XQuery/XPath commands to the SRB, which dynamically

generates the SQL required to process the command against the SRB metadata catalog. The XPath commands can also be issued against an in-memory copy of the state information that is generated by the execution of the web services. The implications for the NVO are that the separation of Query formulation from SQL generation is feasible. A generalization of this is used in Matrix. A request can be issued to discover the attributes specified in a collection. An XQuery request can then be formed on the named attributes and issued against the collection. For federation of multiple metadata catalogs, additional knowledge is needed to understand the logical relationships between the collection attributes.

8.4 Virtual data

The ability to describe the derivation or provenance of a derived data product makes the concept of “virtual” data feasible. The derivation steps can be applied to create the desired product, or the product can be retrieved from storage if already created.

The concept of virtual data is being used in the Atlasmaker project, which is creating standard digital reference sets from the image archives that have been ported to the TeraGrid (see “A Virtual Observatory Vision based on Publishing and Virtual Data,” <http://bill.cacr.caltech.edu/usvo-pubs/files/VO-vision.pdf>). The data objects are derived by computation from sky survey images, each dependent on others through transformations such as resampling, dynamic range reduction, scaling, principle component analysis, and multi-channel source extraction. These objects and their computational provenance will be published through the NVO registry infrastructure, so that they can be produced either from cache or by on-demand computation.

The Montage project is also building Virtual Data product through the Pegasus software, a product of the NSF GriPhyN collaboration. We have developed a prototype architecture for on-demand production of image mosaics on the TeraGrid. A web service located at JPL creates an abstract workflow description of a mosaic request made to Montage by calling the DAG-building script. This workflow description, written in XML, is submitted to Pegasus, which can map and schedule application workflows expressed in an abstract fashion, i.e., without reference to particular data and resources.

8.5 Application and service integration with Grid

A proposal for resource allocation has been submitted to the TeraGrid for 230,000 service units. The request includes six applications that will analyze image archives and object catalogs through NVO services. The expectation is that the lessons learned from these projects can be used to implement better systems to support astronomer applications on the NVO testbed. In this initial component of a multiyear effort we proposed to use the TeraGrid to (1) expose massive data to massive computing through NVO protocols, (2) run representative applications to explore that data, (3) foster new projects in astronomy that use NVO services and TeraGrid resources, and (4) encourage new ways to use supercomputing facilities for science.

The applications include:

- Image processing (two applications, Atlasmaker and Montage) that use NVO Image Access Protocols (SIAP) to bring data from NVO into TeraGrid.
- Resolving star formation in galaxies, which analyzes cutout images brought through web services from the SDSS archive.
- Fitting quasar spectra, where the spectra are obtained through NVO web services (see above).
- N-point correlations of galaxies, where input data (galaxy positions) are obtained from the SDSS archive through NVO web services.
- Cosmic microwave background grid; computes early-universe conditions through simulation and comparison with WMAP observations.

The NVO TeraGrid proposal is located at

<http://bill.cacr.caltech.edu/usvo-pubs/files/teragrid-nvo-final.pdf>.

9 Applications

Activities in this area were largely driven by the preparations for demonstrations at the January AAS meeting. Three demonstrations were prepared: an updated version of the Data Inventory Service, a database of Sloan spectra, and for users to conduct and display the results of simulations of globular clusters.

We note that a standardized approach for releasing and publicizing VO applications has yet to be developed. This should be one focus of efforts in the current quarter.

9.1 Data location services

Substantial work was performed in upgrading the Data Inventory Service (DIS) to use registries that conformed to the VO Resource Metadata guidelines and were able to access OAI compliant publication registries. Small bug fixes were also made to this released service. Personnel primarily at the STScI, HEASARC, and NCSA worked to ensure that all metadata needed for the DIS were present in the registries and properly used by the DIS. The DIS was demonstrated by various NVO personnel at the January AAS meeting.

The Data Inventory Service is being used by the community at a limited rate. No significant advertising of this site has been done other than demonstrations at the AAS.

9.2 Cross-correlation services

No work scheduled. The Brown-Dwarf Demonstration cross-correlation capability is still available.

Related activity: The German Astrophysical Virtual Observatory (GAVO) project released a service to correlate information from multiple VO compliant databases.

9.3 Image combination, registration

Atlasmaker. The Atlasmaker project is funded in part by NVO and the NSF-GRIST project (Grid Services for astronomy). It uses Grid technology, in combination with NVO interoperability, to create new knowledge resources in astronomy. The product is a multi-faceted, multi-dimensional, scientifically trusted image atlas of the sky, made by federating many different surveys at different wavelengths, times, resolutions, polarizations, etc. The Atlasmaker software does resampling and mosaicing of image collections, and is well suited to operate with the Hyperatlas standard (Williams 2003a). Requests can be satisfied via on-demand computations or by accessing a data cache. Computed data is stored in a distributed virtual file system, such as the Storage Resource Broker (SRB).

We expect these atlases to be a new paradigm for knowledge extraction in astronomy, as well as a magnificent way to build educational resources. The system is being incorporated into the data analysis pipeline of the Palomar-Quest synoptic survey, and is being used to generate all-sky atlases from the 2MASS, SDSS, and DPOSS surveys for joint object detection.

Atlasmaker is a suite of software (Williams 2003) that builds mosaics from NVO-compliant image sets—i.e. images that are exposed through the VO Simple Image Access Protocol. Atlasmaker scavenges a grid for enough computational resources to build terabyte-size atlases, which are then available for federation and datamining. Atlasmaker is available on the open-source model, and is based on a Grid architecture of relocatable programs and web services. There is a choice of two trusted codes for the resampling kernel: Montage from NASA IPAC/JPL/Caltech and SWarp from the French Terapix project. Each offers different advantages in terms of quality and speed.

When a user requests a given data product, the manager will check for its existence in the cache that stores already computed products and, if it is found, return it the user. If not found the data product will be computed. The names of surveys are resolved by the NVO registry to get required metadata, including current URLs for the NVO services that can provide external image data.

Atlasmaker is designed to actively find and exploit Grid resources that it can use. It is a self-contained package that can be wrapped into a tar file that self-deploys in a robust way. This package is combined with a collection of service requests (which survey to render on which atlas pages) and thrown to a machine for computation. We hope to use TeraGrid hotpage services (when available) to decide the destination automatically. The Atlasmaker code is written in Python, a highly portable and powerful scripting language, with a central core of pixel manipulation written in C (Montage and SWarp), with MPI parallelism as well as multi-threading. The applications obtain flexibility through the use of services that may reside anywhere on the net; the services are replicated for fault tolerance. Data is stored and retrieved from a distributed file system based on SRB (Storage Resource Broker), a central pillar of the TeraGrid data system. Currently we have implementations for the TeraGrid queue manager, PBS, and also for the Condor

environment. The code has run at Caltech, SDSC, and NCSA, using data services at SDSC, Caltech, and Johns Hopkins.

Montage. IPAC has successfully collaborated with ISI to develop a prototype distributed architecture for ordering image mosaics that are accepted by a web browser and run on the TeraGrid. This contribution describes that architecture, and the ISI contribution to Systems Architecture describes their development efforts on Pegasus needed to support this architecture.

This work was supported in part by the NVO and in part by Montage.

Design Of Montage TeraGrid Portal. The Montage TeraGrid service accepts requests from two portals, one at JPL and one at ISI, underpinned by a common, distributed architecture. Figures 1 and 2 show this common architecture. First we describe the JPL portal, shown in Figure 1. This portal is a prototype of one we will ultimately deploy for astronomers, who will submit mosaic requests through a simple web form that inputs parameters describing the mosaic (location on the sky, size, coordinate system, projection, etc). A service at JPL/Caltech is contacted to generate an abstract workflow, which specifies the processing jobs to be executed, input, output, and intermediate files to be read or written during the processing, and dependencies between the jobs. A 2MASS image list service at IPAC/Caltech is contacted to generate a list of the 2MASS images required to fulfill the mosaic request. The abstract workflow is passed to a service at the Information Sciences Institute (ISI), University of Southern California, which runs software called Pegasus (<http://www.isi.edu/~deelman/pegasus.htm>) to schedule the workflow on the TeraGrid. The resulting “concrete workflow” includes information about specific file locations on the grid and specific grid computers to be used for the processing. The workflow is then executed on the remote TeraGrid clusters using Condor DAGMan. DAGMan is a scheduler that submits jobs to Condor in an order specified by the concrete workflow. Condor queues the jobs for execution on the TeraGrid. More information on Condor and DAGMan can be found on the Condor web site at <http://www.cs.wisc.edu/condor/>. The last step in the mosaic processing is to contact a user notification service at IPAC/Caltech, which currently simply sends an email to the user with the URL of the Montage output.

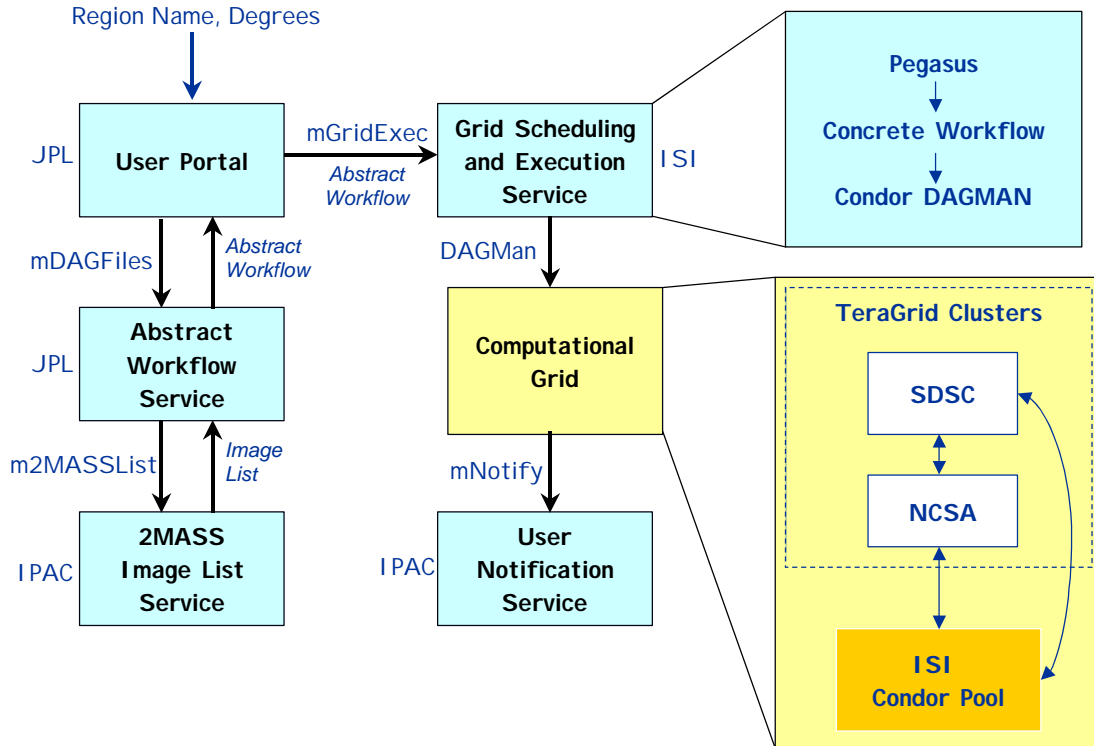


Figure 1. The distributed architecture of the Montage TeraGrid Portal.

The Montage grid portal is comprised of the following five main components, each having a client and server code:

1. User Portal
2. Abstract Workflow Service
3. 2MASS Image List Service
4. Grid Scheduling and Execution Service
5. User Notification Service

The second portal is the Pegasus portal at ISI, which simply takes the place of the User Portal and the User Notification Service. This portal provides complete diagnostic and status information on the processing, and returns all intermediate products. Astronomers simply wishing to receive a mosaic would find the JPL portal more useful.

TeraGrid Performance and Issues. We have run the Pegasus-enabled Montage on a variety of resources: Condor pools, LSF- and PBS-managed clusters, and the TeraGrid (through PBS).

The following table summarizes the results of running a 2-degree mosaic of M16 on the NCSA TeraGrid cluster. The total runtime of the workflow was 107 minutes. The workflow contained 1515 individual jobs. Below is a table that summarizes the runtimes of the individual workflow components.

| number of jobs | job name | average runtime |
|----------------|------------------------|-----------------------------------|
| 1 | mAdd | 94.00 seconds |
| 180 | mBackground | 2.64 seconds |
| 1 | mBgModel | 11 seconds |
| 1 | mConcatFit | 9 seconds |
| 482 | mDiff | 2.89 seconds |
| 483 | mFitplane | 2.55 seconds |
| 180 | mProject | 130.52 seconds |
| 183 | transfer of data in | In the range of 5-30 seconds each |
| 1 | transfer of mosaic out | 18: 03 minutes |

To this point, our main goal was to demonstrate feasibility of running the Montage workflow in an automated fashion on the TeraGrid with some amount of performance improvement over the sequential version. Currently, Pegasus schedules the workflow as a set of small jobs (as seen in the table above, some of these jobs run only a few seconds.) Obviously scheduling too many little jobs suffers from large overheads. In fact, if this processing was run on a single TeraGrid processor, it would have taken 445 minutes, so we are not taking very much advantage of the TeraGrid's parallelism. However, initially structuring the workflow that way allows us to expose the highest degree of parallelism.

9.4 Visualization tools and services

A tool for the display and analysis of Sloan (and other) spectra was developed and demonstrated at the NVO team meeting in Baltimore by JHU personnel. This tool enables the rapid and convenient comparison and summing of spectra. It also provides for a standardized database of response functions. While the current implementation is largely restricted to optical and near IR spectra, the software appears to be usable in all wavebands. The tool was also demonstrated at the January AAS meeting.

This system allows users to upload their own spectra into the database. It is clear that the ability to support user uploads into 'archive' sites is becoming a basic requirement for NVO tools (another example is SkyQuery). This is in addition to the more traditional need for download capabilities.

There was no other NVO funded activity in this area. The JHU developed extensions to the Mirage Java application have been incorporated into its standard distribution.

Related Activity: The Indian Virtual Observatory released their latest version of the VOPlot package for plotting data stored in VOTables.

9.5 Theory

At the January AAS meeting the NVO project presented a demonstration of "theory in the virtual observatory." The demo was largely created at the University of Maryland—College Park through the work of P. Teuben. Much of the demonstration used hardwired

connections to known tools but the user could control the system using CGI scripts and get data back as VOTables.

The user would select a color-magnitude diagram (CMD) from 3 globular clusters and 1 (old) open cluster and then select a model and a simulation snapshot between 0 and 12 Gyr. In addition to a number of runs with models ranging from $N=10,000$ to $N=100,000$, a 3 parameter survey of 270 models was also generated, from which one model could be chosen. Models were placed at the known distance for the user specified cluster, or the user could select another distance.

A comparison was then made and presented to the user of the CMDs of the observations and the model. A run-directory was available to the user from which the VOTables could be downloaded and compared in local tools like VOPlot or Mirage.

P. Teuben also showed the demo at the MODEST meeting at Geneva Observatory.

Issues:

- Limitations of the VOPlot applet hampered its use in the demo. Overplotting is not fully supported. A CGI service was used instead.
- There were a number of limits on the user interface that would need to be rectified in an operational version, including how to pick out the best models, upload user data, select passbands and filters, control fitting, download data, or compare models.

9.6 Statistical analysis

At CMU/UPitt substantial progress has been made on implementing a parallel/grid algorithm for calculating the correlation function of galaxies (currently implemented for 2-point and 3-point correlation functions but extensible to higher order statistics). The code has been tested on single processor and small, tightly couple supercomputers (e.g. the Terascale facility at the PSC). It has been shown to run for systems with 128 nodes and is now being extended to over 1000 nodes for use on the TeraGrid. The results have been tested and verified with existing analysis software and will be interfaced to the clustering web service in the near future.

9.7 Datamining, outlier identification

Staff at the Caltech Astronomy Department are working on the establishment of an NVO-oriented astrostatistics service. This is a part of a collaborative project that also involves groups at PSU (PI = Prof. J. Babu) and at CMU. The activities include:

- Adding more functionality to the previously developed astrostatistics service web pages. This includes additional multivariate techniques (e.g. PCA, clustering, correlation matrices). We are also experimenting with adding external custom written code (e.g., K-density from the CMU group) in addition to the mostly R-based statistical techniques.
- Now that the SOAP-based web-services are being constructed to mirror the capabilities of the CGI based service (see below), we are redesigning the CGI pages to help users understand the VOStatistics web service through tailor-made examples including datasets. This will allow the user to investigate related statistical

techniques before unleashing the best fitting one on geographically distributed large datasets. See: <<http://www.astro.caltech.edu/~aam/science/astrostat/index.html>>

- VOStatistics is a stateful web services-based implementation of the Astrostatistics project and both mirrors and extends the functionality offered by the current CGI-based web pages. We are constructing both a distributed server-side architecture to which the web service is the interface and a client application for the web service novice user (experienced web service users will, of course, be able to implement their own client applications). Data transfer utilizes the VOTable format in both directions (client <-> web service). A prototype of VOStatistics was used in the Web Services tutorial at ADASS XIII in October.

Related Activity: The NASA AISRP-funded *ClassX* project continued to use NVO developed technologies to help in its efforts to classify all detected ROSAT sources. Papers describing the overall classification approach and initial science results on X-ray binaries are in preparation for publication.

9.8 Interfaces to/from legacy software systems

The HEASARC downloaded the VOPlot package and has begun to explore the use of VOPlot as the standard visualization/plotting tool for use within its Browse catalog query service.

10 Community Engagement

10.1 Documentation

In support of the ADASS VO Tutorial, held in Strasbourg, France on 12 Oct 2003, NVO team members developed presentation materials and documentation for many VO core services. See, for example,

<http://www.ivoa.net/twiki/bin/view/IVOA/ADASSDALTutorial>
<http://www.ivoa.net/twiki/bin/view/IVOA/WebgridTutorial>

10.2 Web site

Following discussions at our December team meeting, and subsequent review with the Advisory Committee, we will be undertaking a major redesign of the NVO team web site. The site was originally created to facilitate communication within the team, and only peripherally for the general community. As we increase the suite of VO-enabled applications and core services we wish to reach the user community more directly. The web site redesign will feature VO science applications and software services, with associated reference manuals, sample programs, and other technical information. Links will be provided to all relevant IVOA standards.

Most technical discussions take place in the context of the IVOA (<http://www.ivoa.net>) and the e-mail distribution lists (link “Forum”) and TWiki web pages (link “Community”) located there.

10.3 Technical training initiatives

The ADASS VO Tutorial mentioned in 10.1 was our first effort to provide a structured introduction to VO technology and tools to the astronomical software development community. R. Hanisch organized the overall program, which featured talks and demonstration by representatives of several IVOA project groups. D. Tody, T. McGlynn, G. Greene, W. O'Mullane, and M. Nieto-Santisteban of the NVO project gave presentations. The program is shown below.

VOTutorial: Virtual Observatory Software & Standards

Sunday 12 October

Main Lecture Hall

13:00 to 17:00

Introduction (Bob Hanisch)

VOTable (Sonali Kale)

- The VOTable Format (Francois Ochsenbein)
- VOTable Parsers (Pallavi Kulkarni)
- Developing VOTable-Based Applications
 - VOPlot (Sonali Kale)
 - Mirage (Sonali Kale)
 - Topcat (Mark Taylor)
 - Treeview (Mark Taylor)

Data Access Layer (Doug Tody)

- DAL Overview (Doug Tody)
- How to Implement the DAL Cone Search and SIA services (Gretchen Greene)
- How to Navigate VO Datasets Using VO Protocols (Tom McGlynn)
- How to Adapt Existing Archives to SIAP
 - The XMM-Newton and ISO Cases (Pedro Osuna)
 - Interfacing the CDS Aladin Image Service and Client Viewer (Francois Bonnarel)

Web and Grid Services (Andre Schaaff)

- How to Consume Web Services in C# and Java (Tamas Budavari and Andre Schaaff)
- How to Implement a Client Program in Java or Other Scripting Languages to Use the VO Statistics Web Service (Matthew Graham)
- How to Build a Java or C# Client for a Registry Web Service (Wil O'Mullane)
- How to Build C# and Java Clients for the SDSS Image Cutout (Maria Nieto-Santisteban)
- How to Build a Java Client for a Grid Service Using Globus Toolkit 3 (Guy Rixon)
- Inside the Server (Andre Schaaff)

Over 200 people attended the Tutorial, and feedback was very positive. Our goal was to show a broad audience that VO development was open to all, and we believe this was accomplished.

Planning has begun for an NVO Summer School, to be held 13-17 Sept 2004 at the Aspen Center for Physics in Aspen, Colorado. The Summer School will extend the ADASS VO Tutorial concept, providing more in-depth lecture demonstrations followed by hands-on development. This approach constrains the number of participants to about 40 people, but we believe that this is an effective way to begin to disseminate information about applications software development for the VO into the user community. Proposals for financial support for the Summer School will be submitted to NSF and NASA in the first quarter of CY2004.

10.4 Advocacy

The Astronomical Data Analysis Software and Systems (ADASS) Conference, held in Strasbourg, France on 12-15 October 2003, was replete with presentations and discussions of the Virtual Observatory. The large attendance at the VO Software Tutorial organized by R. Hanisch and presence of VO language in perhaps 2/3 or more of the posters and talks is clear evidence of the growing interest and acceptance of the VO among professional astronomical software developers.

We continue to advocate the NVO to the broader astronomical community, mainly through our demonstrations at AAS meetings and through our web site, which will be redesigned as noted in 10.2. We have had informal conversations with members of the AAAC and NASA Science Archives Working Group, and are prepared to make more formal presentations to such advisory committees at any time.

The NVO Advisory Committee has suggested that we update our long-term strategy for the NVO and share with the agencies our expectations for how NVO can transition from a development phase to an operational phase, and what resources might be required to carry on NVO in a steady state. We plan to update various white papers written several years ago, based on several years of real experience, so that the agencies have more substantial information on which to base future spending plans.

11 Education and Public Outreach

11.1 Strategic partnerships

The partnership with the Berkeley-based SegNVO project continued and developed a prototype demonstration project. This collaboration involves ManyOne Networks, a company that is working to bring a free and encyclopedic collection of knowledge to the public over the Internet. The major vehicle for their work is a new Internet browser, the M1 browser, which includes advanced streaming and compression techniques for presenting multimedia and virtual reality to the public over low bandwidth connections.

The demonstration project was a virtual reality exploration of the stars in the Sun's neighborhood. NVO provided links to automatically retrieving astronomy survey

images, as well as selected links to more processed (i.e., public level) imagery.

11.2 Formal education

We began discussions, along with the SDSS, of providing data to the Northwestern University based Collaboratory project. This project enables groups of students at widely separated venues to work together on projects, some of which will involve analysis of astronomy data sets.

Somewhat related to the above project, we participated in informal discussions with a nascent online student astronomy publication that would be published through the Collaboratory. The value of making the journal articles and the data sets used in the investigations searchable and available through NVO was agreed upon. No formal project has yet been started.

11.3 Informal education

No activities this quarter.

11.4 Outreach and press activities

The SegNVO demonstration project was presented by ManyOne Networks at the NVO booth during the AAS meeting in Atlanta.

11.5 Technical development

Initial work began on developing an SIAP interface to the Hubble press release archive. The STScI Office of Public Outreach agreed to a pilot project to evaluate the workload necessary to make the archive images VO-compatible, both in future press releases and in a re-processing of the archive.

Work was begun on reviving the previously stalled project of a VO-compliant archive of amateur astronomy imagery. The initial model of the data content and structure was developed for future translation into an XML DTD or Schema.

Activities by Organization

Caltech-Astronomy Department

The work at Caltech Astronomy Dept. has been done mainly by Drs. Ashish Mahabal and Matthew Graham, under supervision by Prof. S. George Djorgovski. There are several areas of current effort:

1. Establishment of an NVO-oriented astrostatistics service. This is a part of a collaborative project that also involves groups at PSU (PI = Prof. J. Babu) and CMU.

The activities include:

- Adding more functionality to the previously developed astrostatistics service web pages. This includes additional multivariate techniques (e.g. PCA, clustering, correlation matrices). We are also experimenting with adding external custom written code (e.g., K-density from the CMU group) in addition to the mostly R-based statistical techniques.
- Now that the SOAP-based web-services are being constructed to mirror the capabilities of the CGI based service (see below), we are redesigning the CGI pages to help users understand the Vostat web service through tailor-made examples including datasets. This will allow the user to investigate related statistical techniques before unleashing the best fitting one on geographically distributed large datasets. See: <<http://www.astro.caltech.edu/~aam/science/astrostat/index.html>>
- VOStatistics is a stateful web services-based implementation of the Astrostatistics project and both mirrors and extends the functionality offered by the current CGI-based web pages. We are constructing both a distributed server-side architecture to which the web service is the interface and a client application for the web service novice user (experienced web service users will, of course, be able to implement their own client applications). Data transfer utilizes the VOTable format in both directions (client <-> web service). A prototype of VOStatistics was used in the Web Services tutorial at ADASS XIII in October.

2. Work on a NVO registry service. The Caltech NVO Registry (Carnivore) is currently the only registry employing a native XML database (eXist). It is harvestable via the OAI CGI-based interface, queryable via XPath and data can be entered via a web form interface or a client application. The latter simplifies inputting data for the user by generating a structured interface based on the schema describing the information that the user wants to input. This will work on any valid XML schema and not just the VO resource metadata ones. M. Graham is a regular participant in the NVO Resource metadata telecons.

3. Design of the databases and pipelines for the Palomar-QUEST survey. This will be a fully VO-integrated, major synoptic sky survey. We have been prototyping database designs for the various levels of data products with particular views toward easy federation and a real-time transient discovery system. Many of the functionalities needed for the processing and analysis of data from this survey can be cast in the context of Grid-based computing and Grid services. We started exploring these possibilities for the

newly funded ITR project, GriSt. Some of this was presented at the January 2004 AAS meeting in Atlanta.

4. Conceptual development of the Hyperatlas framework for the NVO (effort lead by Dr. Roy Williams), and the Grid-based implementation of it (Atlasmaker). This work was presented at ADASS XIII.
5. Upkeep of the DPOSS datasets used for a variety of NVO efforts, as a part of the data collections available at SDSC, among others.

Caltech-Center for Advanced Computational Research (CACR)

The Caltech NVO registry project (Carnivore) has been developed to provide a user interface to the VO registry infrastructure, allowing publication and query of an XML database of registry records, with OAI harvesting.

We have been working extensively with the NVO TeraGrid project, bringing up the Atlasmaker code for image federation. This code can use any NVO-SIAP service to build mosaics for multi-wavelength image federation. We have been building connections to the Palomar-Quest synoptic survey, so that Atlasmaker will be used for deep source detection in multi-temporal datasets.

The Hyperatlas standard for interoperable image atlases has been submitted to the arXiv.org preprint archive. The Hyperatlas defines some collections of WCS projections that cover the sphere uniformly (“atlases”) with the intention that image reprojection projects can interoperate by means of this standard.

NVO team members, led by the University of Pittsburgh and Caltech, have generated a proposal to the NSF TeraGrid for use of 257,000 CPU-hours on this massive distributed computing system. The proposal includes support for the NVO testbed at SDSC, as well as support for six science applications that use NVO resources.

The NVO web site is undergoing renovation, converting it in outlook from an internal site for the NVO team members, to a portal by which astronomers can use NVO facilities, tools, and archives.

Caltech-Infrared Processing and Analysis Center (IPAC)

Please note that the scope of the work in the past quarter was limited owing to administrative delays in making FY04 funding available (internal to Caltech).

1. The following SIAP-compliant data sets have been registered with the NVO by IPAC:
 - 2MASS 6x (high sensitivity) Lockman Hole images
 - IRAS Infrared Galaxy Atlas
 - IRAS Extended Infrared Galaxy Atlas
 - IRAS Mid Infrared Galaxy Atlas

- IRAS Infrared Sky Survey Atlas
- 2MASS Large Galaxy Atlas
- 2MASS All Sky Quicklook Images
- NED Image Metadata
- Midcourse Space Experiment (MSX) Atlas
- Spitzer First Look VLA image
- Infrared Telescope in Space Atlas

As part of this effort, we have developed tools for converting column delimited ASCII tables to VO Table format.

2. We have developed large parts of a test bed needed to support ROME testing. We have developed a multi-threaded job manager, written in Java, which automates submission of large numbers of jobs to ROME. This is a contributed effort, supported in part by the NASA/IPAC Infrared Science Archive and in part by the NVO. We have also built the JBoss application server at IRSA, and have begun to port ROME to it.
3. IPAC has successfully collaborated with ISI to develop a prototype distributed architecture for ordering image mosaics that are accepted by a web browser and run on the TeraGrid. This work was supported in part by the NVO and in part by Montage.

Canadian Astronomy Data Centre/Canadian Virtual Observatory

The Sloan Digital Sky Survey Data Release 1 was delivered by sneaker net to CVO in November and the data were transferred to SQL server running on a local machine. The process of transferring the SDSS tables to DB2 has proven to be challenging and informative in terms of database functionality. The transfer to DB2 is expected to be completed in February 2004. At that point data engineering will begin with input from SDSS with the goal of transforming SDSS quantities into the CVO generic data model for querying. An investigation is being done to determine the feasibility of transforming between data models on the fly as opposed to storing transformed values. This work is important in the context of the general problem of data access when non-standard data models are in use.

A project has been undertaken between CVO and the Human-Computer Interaction group in the Department of Computer Science at the University of Victoria. The purpose of the project is to do work flow analysis and task analysis and to use the results to improve ways of accessing VO databases. During the spring semester a class of 60 undergraduate students are using the CVO Prototype as a case study in developing improved user interfaces.

CADC and ST-ECF staff visited STScI to continue planning the production of advanced data products which will be important components of VO. These include a second generation of WFPC2 Associations Stacks and products from the Advanced Camera for Surveys.

Ingestion of ROSAT RASS and pointed-observation catalogues continues in collaboration with the German VO (GAVO). Final work on ingestion of 2QZ source catalogues into the CVO Prototype is nearing completion. Other data content that is under consideration or in the planning stages includes the 2dF Galaxy Redshift Survey, the 6dF survey and 2MASS.

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

Work has continued on the development of a web service interface to the n-point clustering techniques described below. The current implementation queries the SDSS database returning data and random sample masked for the presence of bright stars. Work is on going concerning the scaling of the algorithm with sample size and formatting the outputs of the web service as a VOTable.

A. Connolly and J. Gardner with R. Williams (Caltech) coordinated the NVO proposal for TeraGrid resources. This proposal comprises a data grid component and six grid applications that are designed to work through the NVO standards and protocols. Four of the six analysis applications have been developed at Pitt/CMU and integrated into a grid environment. In the coming months they will be ported to the TeraGrid (once the award of resources is approved). 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.

Substantial progress has been made on implementing a parallel/grid algorithm for calculating the correlation function of galaxies (currently implemented for 2-point and 3-point correlation functions but extensible to higher order statistics). The code has been tested on single processor and small, tightly couple supercomputers (e.g. the Terascale facility at the PSC). It has been shown to run for systems with 128 nodes and is now being extended to over 1000 nodes for use on the TeraGrid. The results have been tested and verified with existing analysis software and will be interfaced to the clustering web service in the near future.

Fermi National Accelerator Laboratory (FNAL)

No significant NVO activity occurred this quarter due to personnel being diverted to work on grid projects and preparation of SDSS data release 2.

In related work not directly funded by NVO, V. Sekhri continued working on the Virtual Organization (VO) project to provide a simple interface for users to authenticate themselves to gain access to grid computing resources. Such an interface is needed by a wide range of distributed computing projects (iVDGL, EDG) and will be useful for integrating NVO with grid computing resources. This project (which is largely for use

by iVDGL) continued through December and is now expected to last through January.

High Energy Astrophysics Science Archive Research Center (HEASARC)

The primary foci of HEASARC activities during the past quarter were in the continued deployment and development of the Data Inventory Service, and the development of standardized VO resource metadata for HEASARC catalogs and the publication of that metadata in an OAI compliant publication registry. The HEASARC and STScI worked together to fully understand the mechanisms for OAI transport and to work the kinks out of the system. Parallel versions of the DIS service were made operational. Only a few tiny internal differences were made between these two versions, however they talked to different registries at JHU. One of these registries was dynamically updated using OAI harvesting from HEASARC-provided resources.

Resource metadata for the approximately 300 HEASARC catalogs was developed and internally reviewed. This metadata was then published using public OAI software tools. Developing VOResource schema-compliant metadata was exacting and also seemed to be a moving target as the validation tools used by the target registries were changed.

The HEASARC contacted the VO-India and downloaded the VOPlot software for evaluation as part of the HEASARC's standard catalog query environment.

HEASARC personnel attended NVO team meetings in Baltimore and participated extensively in the weekly NVO metadata telecons.

HEASARC personnel suggested a major revision in the proposed UCD2 protocol. This suggestion attempted to tie together developments in VOTable structures, data models and UCDs. The HEASARC participated in the extensive discussion that ensued.

Johns Hopkins University

T. Budavari has been involved in the definition of the spectral model for NVO. He continued to support the Spectral website at JHU for the AAS demonstration. Working with A. Szalay he made a SkyQuery template for setting up SkyQuery for 3rd party datasets.

W. O'Mullane (with G. Greene, STScI) has done much work on the Registry prototype. This now works with the latest VOResource schema, located at <http://sdssdbs1.stsci.edu/nvo/voregistry/>. This was also demonstrated at the AAS. ServiceInterfaces0.1 was also produced and posted on the IVOA TWiki, completing an action from the Strasbourg IVOA meeting. We worked in collaboration with K. Benson (AstroGrid) on registry services and harvesting. Part of the Healpix Library was converted to TSQL to make position-to-HealpixID conversion available in SQL Server.

V. Haridas, W. O'Mullane, and N. Li have been defining ADQL, producing a new IVOA ADQL proposal, and are working on the OpenSkyNode portal. ADQL0.7 was

posted on the IVO TWiki. V. Haridas has developed a SkyQL<->ADQL translation service and web pages links given in WBS 7.1 above.

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.

M. Nieto-Santisteban has incorporated the SDSS DR1 image cutout and finding chart service into the recently announced public Data Release 1 (DR1) of the SDSS catalog data (<http://skyserver.sdss.org/dr1/en/tools/chart/>).

G. Fekete is deploying VDT (virtual data toolkit) at JHU making us part of the Condor/Grid configuration at Fermilab.

Microsoft Research

J. Gray assisted with the SDSS DR1/DR2 data loading and schema, and co-authored a paper on scientific data federation for *The Grid 2* (see Publications).

National Optical Astronomy Observatories (NOAO)

D. De Young participated in Executive Committee telecons, team meetings, and the Advisory Committee meeting. De Young led the science requirements definition for the globular cluster theory demonstration project and helped guide the demonstration development. He also arranged for NVO to use the facilities of the Aspen Center for Physics for the planned September 2004 NVO Summer School.

Several NOAO/DPP staff (R. Shaw, M. Fitzpatrick, F. Valdes, and N. Zarate) attended the IVOA Interoperability Workshop in Strasbourg in mid-October. Shaw, Fitzpatrick, and G. Chisholm participated in the NVO team meeting at JHU in early December. A new SOW for NOAO work on NVO was drafted (and is currently undergoing internal review).

National Radio Astronomy Observatory (NRAO)

D. Tody organized the DAL session for the VO tutorial at the ADASS in Strasbourg, and chaired the IVOA DAL working group meeting following the ADASS. A small working group meeting was organized at the NVO team meeting in Baltimore to work out the details of the simple spectral data model.

Development of the simple spectral access interface (SSA) for DAL was the main NVO-related activity at the NRAO this quarter. D. Tody organized this activity, working with M. Dolensky (ESO) to draft the query interface, and with J. McDowell (SAO) and others to define the data model, and with E. Greisen (NRAO) and others to ensure that the proposed FITS spectral WCS standard would be consistent with SSA. As part of the FITS standards effort, D. Tody critiqued the FITS MIME type proposal and reviewed the FITS spectral WCS proposal overall.

The NRAO archive (J. Benson, NRAO) was commissioned in October 2003 and has been a big success in the first few months of operation. The archive provides, on spinning disk, the entire historical and new VLA archive, as well as portions of the VLBA archive. Ingest of the remaining 10 TB of historical VLBA data into the archive will take another 1-2 years at the current rate. Efforts are underway to replicate the entire NRAO archive at NCSA to provide both backup and high bandwidth access for the community to the bulk data. Work continues on development of an automated pipeline processing system for ALMA data. We are currently looking into extending this to include on-demand pipeline processing of archival VLA data, which would potentially allow mining of the VLA data archive in the image and spectral domains using NVO tools and interfaces.

Raytheon/ADC

NVO project members formerly affiliated with Raytheon (under a support contract at the Astronomical Data Center at NASA/GSFC) have relocated to other organizations following NASA's decision to close the ADC.

K. Borne (George Mason University): K. Borne attended the December team meeting. Complications in rerouting Raytheon funding to GMU have delayed work this quarter.

E. Shaya, B. Thomas (University of Maryland): Raytheon/UMD staff intensively participated on the development of the VO data-model, specifically for the general Quantity object. This included attending two days of meeting prior to the Baltimore general NVO meeting to hash out consensus on the model, writing and submitting a draft Quantity data model paper submitted to the ADASS proceedings, a presentation on it at the ADASS meeting, and we, along with J. McDowell, are in the process of preparing an official Note to the IVOA executive committee to propose an IVOA standard. This object will be the base object for exchanging all physical quantities and includes values, or value lists, errors, units, and coordinate information. We also developed an OWL (Web Ontology Language) Ontology for physical units.

In a poster at the ADASS meeting E. Shaya presented an initial concept for an all-inclusive astronomical schema that begins with VO: AstroObject/Universes/Large-Scale Structures/Galaxies/Clusters/ etc. down to meteorites. Each node contains, as children, the appropriate parameter and attributes (luminosity, mass, position, velocity, etc.). Since XMLQuery works on a specific schema at a time, this could be used to support XMLQuery in a distributed astronomical data system.

E. Shaya presented, in a poster, examples of a Problem Statement Language (PSL) form that would make it easier for scientists to submit complex queries by using a Pseudo Natural Language (PNL) and possibly XForms.

E. Shaya wrote a report on the Astronomy Instrument Markup Language. B. Thomas and E. Shaya attended the NVO team meeting in Baltimore. B. Thomas attended the ADASS meeting in Strasbourg, France. E. Shaya took part in weekly Metadata Work Group teleconferences.

San Diego Supercomputer Center

The major activity at SDSC has been the release of versions 3.0 and 3.0.1 of the Storage Resource Broker data grid for use by the NVO testbed. SDSC has ported the current releases of the SDSS (R1), 2MASS (R3), and DPOSS to the new data grid federation technology. The MACHO collection will be replicated onto the NVO testbed using the new data grid federation technology.

A second activity has been support for the porting of NVO applications onto the TeraGrid. This has helped quantify the expected performance of NVO services. SDSC contributed to the TeraGrid resource allocation request that will execute six NVO applications on the NVO testbed.

A third activity has been support for standards efforts. SDSC helped form the Astrophysics Application Research Group of the Global Grid Forum. The GGF views applications as imposing end-to-end performance and consistency requirements on grid services, and is seeking feedback from the NVO community on the robustness and performance of grid software.

SDSC staff involved in the NVO project include R. Moore (Executive Committee, WBS 3—System Architecture lead), G. Kremenek (archive replication), D. Archbell (catalog optimization), V. Nandigam (SDSS catalog implementation), and L. Brieger (2MASS reprojection).

Smithsonian Astrophysical Observatory

SAO continued to lead the Data Model design (WBS 5). J. McDowell coordinated several technical meetings within the NVO and IVOA community. These meetings led to progress on several areas of data model. A proposal for the spectrum data model was outlined after discussions between J. McDowell and D. Tody. Significant progress was made in reconciling the observation models proposed by different groups after technical meetings in Strasbourg in October and documented. A technical meeting in December on a quantity model led to broad areas of agreement among the participants. A. Rots made substantial progress on Space-Time Coordinate metadata schema's transformation from choice groups to substitution groups. He presented the Space-Time Coordinate design and plans to the team and to the NVO Advisory Committee in December. The SAO team presented 2 NVO Data model related posters at ADASS 03.

Meetings:

- A. Rots and J. McDowell attended the IVOA Interoperability meeting, Strasbourg, October 16-17, J. McDowell led the Data Model Working Group session.
- A. Rots and J. McDowell attended the NVO Team meeting, Baltimore, December 9-10. J. McDowell led a Data Model Technical Meeting in Baltimore in conjunction to that meeting.
- A. Rots and J. McDowell attended the NVO Advisory Committee Meeting, Baltimore, December 11-12.

Space Telescope Science Institute

NVO Services. STScI has published and maintains several standard NVO cone search and SIAP services. These include Guide Star Catalog -I and II, Digitized Sky Survey -I and II, Tycho, Hipparcos, HST observation pointing catalog, and MAST (Multimission Archive at Space Telescope) observation catalog holdings including GALEX. MAST has also integrated its observation catalog queries with the VOPlot analysis tool for data visualization. In addition, STScI staff are working with NVO Data Access Layer designers for the Simple Spectral Access Protocol development. The STScI Office of Public Outreach is currently evaluating the publication of press release images as an NVO SIAP service.

STScI staff are working to develop a plan for coordinating NVO work between multiple divisions and identifying areas where NVO services may facilitate or improve existing HST archive access.

NVO Infrastructure. An NVO searchable registry was successfully developed and publicly released. The registry is built using .NET technology with an SQL Server backend. It has been integrated and tested with external client packages (DIS, Mirage) and also harvested by our IVOA collaborator, AstroGrid, in its registry. The registry also has direct web form access that supports simple SQL query statements and in addition allows users the ability to perform several administrative tasks with service authentication via password. The registry now contains approximately 6000 resource entries. There is ongoing work in developing methods for managing data integrity and quality of these resources.

As part of the AAS science demonstrations, the STScI registry was fully integrated with the GSFC DIS (Data Inventory Service) client through the registry suite of web services. These services enable client search and retrieval of astronomical resource and service metadata providing analysis tools the required information for data visualization.

STScI has successfully demonstrated harvesting from other NVO registries that publish resources using the OAI (Open Archives Interface) standard protocols. These include NSCA, Caltech, VizieR, Astrogrid, and HEASARC publishing registries. The resource metadata harvested is accessible either directly from the registry searchable forms or through DIS and Mirage client applications. A VOResource schema parser was developed and integrated into the OAI registry harvesting.

Technical Exchange. The collaborative work between STScI technical and scientific staff and the JHU SDSS Science Archive team continues in specific areas of NVO development and also biweekly technology exchange meetings. These meetings have progressed to include NVO-related project status reporting, demonstration of new technology, and coordination of activities both internal to each group and being undertaken jointly.

Education and Public Outreach. Initial work began on developing an SIAP interface to the Hubble press release archive. The STScI Office of Public Outreach agreed to a pilot

project to evaluate the workload necessary to make the archive images VO-compatible, both in future press releases and in a re-processing of the archive.

Project Management. Project management activities at STScI included organization of the team and Advisory Committee meetings, organization of the VO Tutorial at the ADASS Conference, organization of the two-day IVOA interoperability workshop following ADASS, and preparation of the Year 2 Annual Report. R. Hanisch also helped coordinate the development of the globular cluster theory demonstration with P. Teuben (U. Md.), arranging for access to observed color-magnitude diagram data from M. Shara (AMNH).

United States Naval Observatory

An ASCII version of USNO-B1 suitable for ingestion by commercial databases was constructed at USNO Flagstaff. However, before it was delivered to SDSC for use in their tests of merging large datasets into databases, the disk on which this copy lived died. Work is on-going to regenerate the catalog.

A USNO NVO team member attend the December working team meeting.

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

R. Plante primarily worked with the various registry developers to support their efforts to upgrade to the latest schemas and identifying compatibility issues. He also developed a schema documentation style sheet for NVO schemas by adapting an open-source, general-purpose product called XS3P. This allows one to create a metadata dictionary meant as an aid to users of NVO schemas.

R. Williamson and R. Plante upgraded the VORegistry-in-a-Box and are now working on a reference implementation of the SkyNode service using Java and PostgreSQL.

University of Pennsylvania

P. Protopapas has worked on the following:

- *MACHO DB and NVO:* Completed the creation of the Macho database. A new dedicated machine has been purchased and configured. All light curves, template files and statistical tables are stored in a relational database (postgresql), which has been installed on the same machine. Wrote C++ library for accessing the database and Java based WEB services that can be used to retrieve light curves remotely. We are working in extending beyond the light curves through a SIAP for request of the original images by pointing to the store images in Australia or SDSC. Need to coordinate that with the people at ANU and SDSC.
- *Combining NVO web service applications with federated data for solving a science problem:* Completed the first phase of porting the orbit calculator to a NVO service and demonstrate how it can be used in combination with other NVO databases to better determine orbits or schedule follow-up observations.

- *Developed an algorithm for finding clusters and outliers based on a similarity matrix.* So far we have explored two different definitions of similarity; one that is based on straight comparisons of the light curves (including phase differences) and one based on the wavelet decomposition.
- Extended the relational database functionality to allow for queries that will compare light curves by comparing the actual shape of the curve in addition to the standard SQL-type queries on the metadata.

University of Southern California (USC/ISI)

In the last three months, the work at ISI has been on focused in two major areas.

1) Interfacing Pegasus with external services such as the Montage mDAG service, which provides abstract workflow descriptions.

The Montage workflow can be described as an abstract directed acyclic graph (DAG). The abstract DAG is defined in an XML format. A web service was developed at IPAC which when provided with the Montage metadata information such as object name and the mosaic degree returned the abstract DAG and other related files. The IPAC service was named mDAG. ISI has been integrating Pegasus with mDAG.

The integration process was iterative and included changes of the mDAG software as well as the syntax of the DAG. For example, the DAG syntax needed to be modified to accommodate execution across multiple execution pools. As a result we were able to run the mDAG-generated abstract workflows over multiple condor pools as well as the TeraGrid.

2) Developing the Pegasus Grid portal for Montage and other applications.

Initially, the Montage job submission was being done using Pegasus and Condor-G command line tools. ISI developed a web-based portal that provided a user-friendly interface for submitting and monitoring Montage jobs as well as other applications.

Login. The user needs to logon the portal using a MyProxy username and password. This username and password is used to retrieve the user's credentials from a MyProxy server. These credentials are used to authenticate with and submit jobs to various grid resources selected by the user.

Parameter Selection. The user can enter the Montage object name (the sky region) and the degrees for the required Montage mosaic. Currently we require only these two parameters but they can be easily extended to include more parameters. Once the user provides these parameters, the portal shows a list of previous Montage mosaics computed over the same sky region. The user can still go ahead and start a new run.

Job Submission. The user needs to specify the execution pools and the output pool. The execution pools are the resources that can be used for computing the Montage mosaic. The output pool indicates the resources where the final computed mosaic should be transferred. The portal authenticates with the resources selected as the execution and the

output pools using the user's credentials. This ensures that the resources are alive and available at the time of submission.

Job Monitoring. When the Montage computation is running, it is periodically polled to ascertain its progress. The user can view the progress of the job on the portal. The portal shows the total number of jobs in the concrete DAG and the number of jobs completed till now. It also shows a break up of individual jobs with their types and start and finish times. The portal creates a visualization of the concrete DAG in the form of a jpeg image.

Notification. When the mosaic is created, the portal transfers it to an http URL and sends an email to the user. The email contains the details of the job and the http URL where the generated mosaic is stored.

The user can also view the Montage mosaics computed by other users using the portal. Reusing earlier computed mosaics can save time and reduce redundant computations. The portal is available for preview at <http://pegasus.isi.edu/portal>.

University of Wisconsin
No activities this quarter.

Publications

Alexov, A., & Good, J., 2003, "ATLAS – a general search service for complex collections of data," ADASS XIII, P2.27, in press.

Berriman, B., Bergou, A., Deelman, E., Good, J., Jacob, J., Katz, D., Kesselman, C., Laity, A., Singh, G., Su, M.-H., & Williams, R., 2003, "Montage: a Grid-enabled image mosaic service for the National Virtual Observatory," ADASS XIII, O7.3, in press.

Borne, K., Huang, Z., & Luna, J. C., 2003, "Data mining large astronomical databases for interacting galaxies," BAAS 35 (5), 20.02.

Budavari, T., Szalay, A., Malik, T., Thakar, A., O'Mullane, W., Williams, R., Gray, J., Mann, R., & Yasuda, N., 2003, "OpenSkyQuery – VO compliant dynamic federation of astronomy archives," ADASS XIII, P2.18, in press.

Carliles, S., Ho, T. K., & O'Mullane, W., 2003, "VO enabled Mirage, the IVOA client package, and using JSky and JAI to view FITS images in Java applications," ADASS XIII, P3.12, in press.

Colberg, J. M., Connolly, A., & Scranton, R., 2003, "Web services for the Sloan Digital Sky Survey," ADASS XIII, P1.20, in press.

Derriere, S., Gray, N., McDowell, J., Mann, R., Ochsenbein, F., Osuna, P., Martinez, A. P., Rixon, G., & Williams, R., 2003, "UCDs in the IVOA context," ADASS XIII, P3.17, in press.

Dobos, L., Szalay, A., Csabai, I., & Budavari, T., 2003, "Spectrum and bandpass services for the Virtual Observatory," ADASS XIII, P2.21, in press.

Dolensky, M., & Tody, D., 2003, "Survey among spectral data providers and consumers," ADASS XIII, P3.23, in press.

Dowler, P., Shaya, E., Thomas, B., Borne, K., & Huang, S., 2003, "Virtual Observatory object query language," ADASS XIII, P3.27, in press.

Durand, D., Micol, A., Schade, D., Simard, L., Koekemoer, A., Corbin, M., & Donahue, M., 2003, "The WFPC2 associations version 2: toward VO," ADASS XIII, P2.29, in press.

Fekete, G., Szalay, A., & Gray, J., 2003, "HTM2: Spatial toolkit for the Virtual Observatory," ADASS XIII, P3.9, in press.

Graham, M. J., Williams, R., Djorgovski, G., Mahabal, A., Baltay, C., Rabinowitz, D., Bauer, A., Snyder, J., Morgan, N., Andrews, P., Szalay, A., Brunner, R., & Musser, J.,

2003, "QUEST: A case study in designing sky surveys in the VO era," ADASS XIII, O1.3, in press.

Gray, J., and Szalay, A., 2003, "Scientific Data Federation," in *The Grid 2: Blueprint for a New Computing Infrastructure*, I. Foster, C. Kesselman, eds., Morgan Kaufman, pp. 95-108.

Greene, G., O'Mullane, W., Hanisch, R., & Gaffney, N., 2003, "Searchable registry for the National Virtual Observatory," ADASS XIII, P3.8, in press.

Hanisch, R., Linde, T., Plante, R., Richards, A., Auden, E., Noddle, K., Greene, G., & O'Mullane, W., 2003, "Resource metadata for the Virtual Observatory," ADASS XIII, P3.5, in press.

Haridas, V., Budavari, T., O'Mullane, W., Szalay, A., Thakar, A., Pence, W., Conti, A., & Volpicelli, A., 2003, "Making FITS available in .NET and its applications," ADASS XIII, P4.16, in press.

Lemson, G., & Dowler, P., 2003, "A science data model for the Virtual Observatory," ADASS XIII, P4.28, in press.

Lowe, S., Cresitello-Dittmar, M., De Ponte-Evans, J., Evans, I., Harris, M., McDowell, J., & Rots, A., 2003, "Data model design for the Virtual Observatory," ADASS XIII, P3.6, in press.

McDowell, J. C., 2003, "Spectral data models for the VO," ADASS XIII, P3.4, in press.

McGlynn, T., Lee, J., Hanisch, R., O'Mullane, W., & Greene, G., 2003, "A Virtual Observatory Data Inventory Service," ADASS XIII, P3.18, in press.

Monkewitz, S., 2003, "Compare: a scalable and portable catalog cross-comparison engine for the NVO," ADASS XIII, O7.2, in press.

Nieto-Santisteban, M. A., 2003, "ImgCutout, an engine of instantaneous astronomical discovery," ADASS XIII, O8.4, in press.

O'Mullane, W., Gray, J., Li, N., Budavari, T., Nieto-Santisteban, M. A., & Szalay, A., 2003, "Batch query system with interactive local storage for SDSS and the VO," ADASS XIII, O4.4, in press.

Plante, R., Greene, G., Hanisch, R., McGlynn, T., O'Mullane, W., Williams, R., & Williamson, R., 2003, "Resource registries for the Virtual Observatory," ADASS XIII, O7.1, in press.

Purger, N., Szalay, A., Budavari, T., Thakar, A., & Csabai, I., 2003, "Build your own SkyNode!" ADASS XIII, P2.26, in press.

Suchkov, A., & Hanisch, R., 2003, "ClassX detection of hard x-ray binaries in ROSAT PSPC data from pointed observations," BAAS 35 (5), 84.03.

Szalay, A., Gray, J., Fekete, G., Pope, A., Thakar, A., O'Mullane, W., & Kunszt, P., 2003, "Spatial queries with VO regions in SQL," ADASS XIII, P2.17, in press.

Thomas, B., Shaya, E., & Huang, P., 2003, "Proposal for a query-capable quantity based data model in the VO," ADASS XIII, P3.19, in press.

Williams, R., Brieger, L., Feldmann, M., Jacob, J., Kremenek, G., & Moore, R., 2003, "Atlasmaker: Grid-enabled multiwavelength imaging," ADASS XIII, O4.1, in press.

Williams, R., Connolly, A., Gardner, J., et al., 2003, "The NSF National Virtual Observatory —TeraGrid Utilization Proposal to NRAC,"

<http://bill.cacr.caltech.edu/usvo-pubs/files/teragrid-nvo-final.pdf>

Williams, R. D., Djorgovski, S. G., Feldmann, M. T., and Jacob, J. C., 2003, "Hyperatlas: A New Framework for Image Federation," <http://arxiv.org/abs/astro-ph/0312195> .

Williams, R. D., Djorgovski, S. G., Feldmann, M. T., and Jacob, J. C., 2003, "Atlasmaker: A Grid-based Implementation of the Hyperatlas," <http://arxiv.org/abs/astro-ph/0312196>

Williamson, R., & Plante, R., 2003, "A prototype publishing registry for the Virtual Observatory," ADASS XIII, P3.22, in press.

Yasuda, N., O'Mullane, W., Haridas, V., Szalay, A., Ohishi, M., & Mizumoto, Y., 2003, "Astronomical Data Query Language: simple query protocol for the Virtual Observatory," ADASS XIII, P3.10, in press.

Note: ADASS XIII abstracts are available on-line at http://www.adass.org:8080/meetings/adass2003/events/final_program/presentationTOC.html

Virtual Observatory Articles in the Popular and Technical Press

Nothing this quarter.

Acronyms

| | |
|---------|---|
| AAS | American Astronomical Society |
| ADASS | Astronomical Data Analysis Software and Systems (Conference Series) |
| 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 |
| WSDL | Web Services Description Language |
| XML | Extensible Mark-up Language |
| 2MASS | Two-Micron All Sky Survey |