Quarterly Report
April-June 2003

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

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

Period covered by this report: 1 April—30 June 2003
Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

Highlights:

- *Scientific.* Work was completed on V1.0 of the Data Inventory Service, the project's first public/supported data access tool. The DIS will be demonstrated and released for community use in conjunction with the IAU General Assembly in Sydney, Australia (14-25 July). The DIS makes use of a dynamic resource registry containing metadata about catalog and image archives available to the NVO.

- *Technical.* International technical collaboration has continued at a high level in the six focus areas identified in January: registries, data models, UCDs (Uniform Content Descriptors), VO query language, data access layer, and VOTable. An International Virtual Observatory Alliance interoperability workshop was held in Cambridge, UK, 12-16 May, and was attended by 60 people from the 12 worldwide VO initiatives. Substantial progress was made in all areas.

  We successfully deployed a prototype registry service, which can be populated through a web interface and through harvesting metadata from OAI (Open Archives Initiative) compliant servers. The NVO project has been leading the development of Resource and Service Metadata, bringing forward two new iterations of a proposed international standard. We expect to reach closure on V1 of this standard by the end of August 2003.

  The first full specification for space-time metadata was completed.

  We developed an alternative naming convention for UCDs (Uniform Content Descriptors) and presented it at the interoperability workshop in Cambridge. There was general agreement on the new convention.

  We also agreed in Cambridge to extend the Simple Image Access Protocol, making only minor revisions at this stage, and to focus efforts next on a Simple Spectral Access Protocol that will accommodate one-dimensional spectral data. We are leading the interface definition activities.

- *Programmatic.* Preparations for the IAU General Assembly were completed. We participated in a joint display/demonstration with our IVOA partners, and co-organized Joint Discussion 8 on Future Large Telescopes and the Virtual Observatory. A meeting of the IVOA Executive was planned in conjunction with the IAU General Assembly.

Issues and Concerns: None at this time.

Activities by WBS

1 Management

1.1 Science Oversight (Executive Committee)

Status: Our primary focus in this period was on the Data Inventory Service demonstration, planned for the IAU General Assembly in July. The DIS supersedes the gamma-ray burst follow-up service, being a general utility for finding information about an object or position. The DIS locates image and catalog resources through a prototype registry. The contents of the registry (its metadata elements) were developed with direct EC participation.

We continued planning for a science prototype based on a theoretical simulation data set, namely, a globular cluster simulation. This will be targeted for the January 2004 AAS meeting.

1.2 Technical Oversight (Executive Committee)

Highlights: In the International Virtual Observatory Alliance (IVOA) we reached agreement on the major technical focus areas for 2003. These include registries, data models, UCDs, data access layer, VO Query Language, and VOTable. Major progress was reached in all areas during an IVOA interoperability workshop, held in Cambridge, UK.

In this context, the NVO project developed a prototype resource registry, and led many IVOA-level discussions concerning metadata content and structure.

1.3 Project and Budget Oversight (Executive Committee)

Highlights: The project's Education and Outreach Coordinator, Mark Voit, will be leaving STScI this summer. We are transitioning the EPO responsibilities to Dr. Frank Summers, a scientist in the Office of Public Outreach at STScI with extensive experience in museum exhibits, planetariums, and data visualization.

Status: Project expenditures for this quarter were $777,765, a figure that includes previously delayed invoices from several groups. See the financial supplement for additional information.

2 Data Models

2.1 Data Models / Data Model Architecture (McDowell, SAO)

Highlights: J. McDowell led the DM Working Group meeting at the IVOA inter-operability workshop in Cambridge (May 2003). We achieved consensus on an international adoption process for data models, pending acceptance of a general IVOA standardization process. Established DM work packages and leaders for the IVOA working group. Progress is now being made on the small scale (Quantity model) and large scale (Observation model).

Status: We are working towards consensus on initial models for the October ADASS meeting. R. Plante (NCSA) is developing a style guide for rendering data models in XML. An initial version appears as Section 2 of the VOResource Overview document.
(http://www.ivoa.net/internal/IVOA/IVOARegWp03/MDinXML-Summary. html).

2.2 Data Models / Data Types (McDowell, SAO)

Highlights: A proposal for a Spectral data model was released for comments. The Quantity data model is under discussion in IVOA working group forum.

C. Alcock and P. Protopapas (U. Penn.) worked on refinement of one of two competing designs of new standards for the incorporation of time-series data into a federated database system: insertion into modified VOTable format. This approach is being implemented in an SQL database containing the MACHO lightcurve data. Another approach, which they have been exploring for a longer period in time, is a FITS extension with NVO compliant metadata. Much of this effort is on hold pending the outcome of discussions regarding the incorporation of WCS coordinates into FITS, which has profound consequences for the lightcurve implementation (WBS 3).

Status: An initial burst of discussion on work packages has died down but will restart after the IAU meeting. We are monitoring progress of Canadian, French, and ESO groups, and are working to agree on an observation model. R. Plante (NCSA) is moderating an effort to model general scientific quantities.

2.3 Data Models / Data Associations (McDowell, SAO)

Status: S. Lowe (SAO) is working on a description of image mappings and coordinate systems. A. Rots (SAO) continues work on region descriptions (see WBS 3.1).

3 Metadata Standards

3.1 Metadata Standards / Basic Profile Elements (Rots, SAO)

Highlights: The Space-Time Coordinate metadata specification was completed this quarter in the form of an XML schema and presented to the community at the IVOA meeting in Cambridge. An agreement was reached on the specification of regions and an XML schema implementation was included in the STC specification. The next step will be to integrate this metadata standard in the existing interfaces and formats.

Status: The STC and region specification is complete, although there may be small improvements as we gain experience with its use. What is still lacking is the projection definition; we hope to collaborate with StarLink people on this aspect.

3.2 Specific Profile Implementations (McGlynn, HEASARC)

Status: C. Alcock and P. Protopapas (U. Penn.) implemented a test system comprising an SQL database of (a subset of) MACHO lightcurves. Further exploration is required of models for data provenance in situations where moving objects are detected, tracked, and retroactively recovered in multiple datasets.

At the IVOA meeting in May, much discussion of the controlled vocabulary (UCD) produced several ways to significantly enhance the utility of the UCD system. A UCD will now be written as a combination of a base element and modifiers, so that, for example, there would not be a separate UCD for the error estimate of a quantity, but rather a modifier "error" would be attached to the UCD for the base element. The UCD tree will has been pruned, modifiers defined, and usage of the vocabulary defined more closely.

A new feature of the UCD system is namespaces, included so that new UCDs can be created and used without confusion with existing UCDs. The creation of a namespace also carries the responsibility to build and maintain a "resolver" web service that provides descriptive information about the UCD set.

There is a new, international, UCD Steering Committee, chaired by R. Williams, to provide the balance between flexibility and interoperability. This committee provides the means for new UCDs to be added to the existing set.

3.3 Metadata Representations and Encoding (Plante, UIUC/NCSA)

Highlights:

Schema Definition Framework. The definition of the VOResource XML Schema (see Sect. 3.4) was used to develop general techniques for defining metadata in XML. The lessons learned were recorded in a section of the VOResource Overview document (http://www.ivoa.net/internal/IVOA/IVOARegWp03/MDinXML-Summary.html). These lessons will be expanded into a general style guide and submitted for review by the IVOA Registry and Data Model working groups. Most recently, we have successfully identified the techniques that work well with off-the-shelf tools that convert Schemas into software classes.

Naming Standards, DOIs. Building on the requirements development by the Metadata Working Group last quarter and in collaboration with the IVOA Registry Working Group, we developed an outline specification for IVOA identifiers. (Work during the IVOA Interoperability Workshop in May 2003 proved instrumental in solidifying the proposal.) Plante (NCSA) is currently converting the outline to an IVOA Working Draft, the first step in the IVOA standardization process. Publication of the first version is planned for July 7.

Issues and Concerns:

Schema Definition Framework. International review of the proposed resource metadata has brought up some fundamental issues about how metadata are defined. One is whether definitions should be fuzzy, to enable broader interoperability across diverse resources, or precise, to ensure finer processing control over the resources. Another concerns the general approach to our ontology: should we attempt to design the metadata model comprehensively up front or piecemeal, as required by applications? For this issue, NVO prefers the latter approach as it is consistent with our overall program plan; however, this is not necessarily so for other international VO projects.

Naming Standards, DOIs: Two issues are not addressed in the current proposed Identifier specification. First is the issue of URNs—persistent, location-independent names for resources; this is expected to be addressed in a separate specification that builds on the standards for Identifiers and Registries. The second issue is referring to components of an identified resource—say, an image in a data collection or a record from a catalog; it was thought that a single solution might not be appropriate for all applications.

We also recognize that our resource metadata should allow the use of ADS bibcodes to refer to relevant items in the published literature. An additional metadata element, Source (from the Dublin Core), will be proposed to accommodate this information.

Status:

- In general, metadata definition discussions now take place in the international forum of the IVOA.
- A precursor to a metadata definition style guide has been published; a more complete guide is still in development.
- A Working Draft of the Identifier specification is to be released in early July.

3.4 Profile Applications (Plante, UIUC/NCSA)

Highlights:

Query Profile. Development of general query languages has moved into the IVOA VO Query Language (VOQL) Working Group. Two languages are planned. The first is a high-level, science-oriented language that will allow users to form queries that will be intelligently dissected and distributed by a VO portal. E. Shaya and B. Thomas (Raytheon TSC) will concentrate on this level. A lower-level, database-oriented language will capture queries (pulled out of high-level queries by the portal) that can be answered by individual data and service providers. W. O'Mullane (JHU) will collaborate with the Japanese VO project on this level.

Service Directory Profile. The document "Resource and Service Metadata" (RSM), edited by R. Hanisch (STScI) was updated to version 0.7 for review at the IVOA Interoperability Workshop in May. This new version included a refinement of the resource model, a number of additions to support EPO resources, and additions specific to services. An XML form of this schema, VOResource, was also developed by R. Plante and R. Williamson (NCSA) and released for review (http://www.ivoa.net/twiki/bin/view/IVOA/IVAORegWp03). Included was an extension of the basic RSM data to describe Simple Image Access Services, demonstrating how schemas from different namespaces can be used together in a single resource description.

Issues and Concerns:

Query Profile. A number of fundamental issues concerning the VOQL were addressed at the IVOA Interoperability Meeting. Splitting VOQL into two layers allows us to separate the users' needs for complexity from the data providers' needs for simplicity. While the low-level language will be SQL-based, an XML-tagged parse tree version will be available for handling by data providers.

Service Directory Profile. The fundamental resource metadata model is currently under review by the IVOA Registry Working group and may result in substantial changes to the RSM in the next quarter. The general issues being considered are discussed in Section 3.3. More specific to resources, there is concern among some of our international partners that some of the general resource metadata may not be appropriate for all types of resources. We need to decide if such items should be relegated to extensions to the RSM definition or retained and simply ignored by those using the schema when the metadata is not applicable.

Status:

- E. Shaya and B. Thomas are continuing work on their query framework in coordination with the IVOA VOQL Working Group.
- R. Hanisch is preparing version 0.8 of the RSM document for release in early July.
- R. Plante is preparing version 0.9 of the VOResource XML Schema (based on RSM). He is also involved in the general review of RSM within the IVOA Registry Working Group.

3.5 Metadata Standards / Relationships (Rots, SAO)

Status: No work planned at this time.

3.6 Metadata APIs (Plante, UIUC/NCSA)

Highlights: The Registry "Tiger Team," established last quarter, completed a prototype Data Inventory Service (DIS) based on the Gamma-Ray Burst Science Demonstration. Components include:

- Publishing registries at NCSA and Caltech, where data providers can publish descriptions of resources.
- A centralized, searchable registry at STScI that can harvest resource descriptions from the publishing registries.
- The Data Inventory Service portal at HEASARC: a user-oriented, web-based interface for locating data related to a position in the sky. It searches the central registry via a web service interface.

The DIS will be released as the NVO's first end-user service in July at the IAU Assembly in Sydney.

Plante and Williamson have refined their prototype of a deployable, harvestable registry and have packaged it up for external use. It combines a harvesting interface tool from the OAI community with support for VO resource metadata. With this package, data providers will be able to easily describe their resources and expose the descriptions to the VO. Release for experimentation is expected in early July; however, further development will be needed as the VO resource metadata evolves.

Caltech (R. Williams) has built a registry node that has both OAI and publication interfaces. The OAI interface means that records can be harvested by other registry nodes, thereby creating a single virtual system from the distributed collection. The publication interface presents web forms that allow input of resource metadata. The interface supports international users through proper treatment of Unicode text, meaning that, for example, Russian or Japanese text can be used in the NVO registry. This system will be tested, debugged, and deployed to the community in the next quarters, with encouragement to publish to the registry.

Issues and Concerns: The creation of the IVOA working groups will necessitate some rescheduling of our development activities to synchronize with the international efforts. Exchange of metadata between registries is a critical area of interoperability that we need to achieve on an international level. Although our use of OAI to collect metadata has been quite successful, other VO projects have not looked at this existing technology too closely as of yet.

Status:

- The Data Inventory Service will be released to users in July.
- A deployable publishing registry will be released (as alpha) in July.

4 Systems Architecture

4.1 System Design (Moore, SDSC)

Highlights: We have a multi-layer architecture of web services, portals, data processing systems, grids, and catalog/image collections. The current activities are focused on understanding where the data model affects interfaces between these levels, and how these levels can be integrated through process management technology such as Montage. An analysis of the interaction mechanisms that need to be added to Montage has been done (access to Grid computing resources, publication of ROME processor resources and ROME applications, use of ROME to drive a processing pipeline).

Issues and Concerns: A major issue is the appropriate mechanism for collections to be referenced from Grid technology. The NVO is currently supporting two implementations:

- File based access to images within sky surveys
- Collection based access to images within sky surveys

The NVO needs to reach a consensus on the best way to access existing collections. The SRB Data Grid provides collection-based access to 2MASS and DPOSS, and the USNO-B, SDSS, and MACHO image collections are being registered into the SRB Data Grid.

A second issue is the distribution of analysis tasks between data management systems and Grid technology. Data Grids support remote proxies, the execution of data subsetting commands directly at the remote storage system. This effectively is the movement of the application to the data. Grid systems such as Chimera move data to a computation platform for analysis. The choice for the most efficient approach is determined by the complexity of the application (number of operations per byte of data moved).

The NVO needs a plan for integrating high complexity operations through Chimera with low complexity operations through the data handling system? This will require extensions to Chimera to make an appropriate choice for where computation is done.

For semantic based access to data, we need technology similar to that provided by OpenDAP, namely the description of digital entity structure, and the description of the semantic labels applied to the structures. The GGF Data Format Description Language Research Group is examining XSIL and other data structure description mechanisms to propose a standard. NVO will need to track the emerging proposal and see how it can be integrated into the NVO data model.

Status: The hardware and software systems installed within the NVO remain substantially the same. The replication of additional sky surveys onto the TeraGrid infrastructure is expected in 3rd and 4th quarters of FY03. USNO-B will be accessible in 3QFY03. MACHO is under review, and SDSS is being replicated onto TeraGrid infrastructure. The TeraGrid infrastructure is designed to support bulk data access, such that the entire SDSS collection could be accessed in an hour, versus an access rate of 35 days through the present mechanisms.

4.1.1 System Design

The system design of the NVO architecture has the following components:

1. Portals - web service interfaces to analysis procedures (OASIS, French Aladin, the JPL YourSky, and the new Data Inventory Service from NASA Goddard.)
2. Process management systems - data processing pipelines to create derived data products (Chimera, Montage)

3. Web services – uniform capabilities provided across NVO catalogs and image archives (cone search, VOTable catalog query, simple image access)
4. Data access layer - management of methods on data encoding formats for access based on physical quantities (UCDs)
5. Data Grid - management of distributed collections, provision of logical name space for global persistent identifiers, and support for remote proxies (SRB)
6. Computational Grid - access to distributed compute resources (Globus toolkit)
7. Persistent archives - management of technology evolution (SRB)
8. Astrophysics catalogs and Image archives (SDSS, 2MASS, DPOSS)
9. Persistent disk systems - interactive access to sky survey image collections (Grid Bricks)
10. High performance disk caches - high-speed access for bulk data analysis (SAN)
11. Compute platforms – NSF TeraGrid

The implementation of persistent disk caches is being extended at SDSC through the acquisition of a Sun Honeycomb disk cache. This system is being provided by Sun to SDSC to support large scientific collections (30 TB). SDSC plans to provide access to replicas of 2MASS, DPOSS, and USNO-B images on the platform.

4.1.2 System-Level Requirements Definition, and 4.1.3 Interaction with Grid Components and Tools

A provisional charter has been created for an Astronomy Research Group within the Global Grid Forum. N. Walton (AstroGrid project, UK) and R. Moore have volunteered to lead the Research Group. The goal 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.

An important issue with respect to the Grid community is the use of the Open Grid Services Architecture, and the underlying Open Grid Service Infrastructure for managing the life cycle of Grid services. A release was made in June of the OGSA infrastructure. However debates between the OGSA and Web Services Description Language (Semantic Web) communities is still in progress. OGSA manages state information about each service and the WSDL community is debating the best way manage state. NVO needs to track the discussions, and the proposed Astronomy Research Group provides a good way to do this.

For executing Montage on the TeraGrid, there are three implementations: a port by L. Brieger, a version by R. Williams called Atlasmaker, and the original code by J. Good. A port to Globus is being done by E. Deelman. The ports by Brieger and Williams use collection-based access to the surveys. The port by Deelman is intended to be file-based access independently of the collection. A collection-based version has the advantage that greater automation of the interaction can be done.

The immediate Grid integration effort is driven by an assessment of the ROME process management system. The assessment raised the following questions with respect to cost and schedule for implementation:

(1) Target for availability of ROME as a launch pad for Grid applications that is open source and free, including the required subcomponents.

(2) Target for an ODBC/JDBC database access mechanism to replace the EJB framework.

(3) A certificate handling component to ROME, so that authentication can be taken at the portal and forwarded to TeraGrid resources that require authentication.

(4) Integration of ROME with Grid computing to control jobs on remote Grid resources, support jobs that can "roam", i.e., run anywhere. This includes a strong argument about how ROME interacts with Condor-G, including opportunity for Condor team comment and/or collaboration.

(5) A development path to the control of networks of OGSA (Grid and web) services, including DAI data services.

(6) Creation of a diverse committee to decide on messaging protocol and API, syntax, semantics, with report on how it is combined, processed, and archived.

(7) Addition of an OAI interface to get service definitions from NVO registry.

(8) Direct, authenticated access to Storage Resource Broker, to allow use of SRB as a distributed virtual file system.

(9) Demonstration of two distinct applications that use the ROME environment:

(a) image mosaicing on Grid with security and choice of host machines and

(b) big cross-compare of remote archives, without caching, using NVO protocols.

The demonstrations should show off the new features in the above list.

Action items for the TeraGrid community are to decide what additional sky survey collections will be accessible through the TeraGrid and to demonstrate the porting of NVO services onto the TeraGrid Data Grid. The idea is that once a collection is registered into the NVO Data Grid, then NVO services will become automatically available.

4.1.4 Logical Name Space

SDSC has been creating logical name spaces for NVO collections through their registration into the TeraGrid Data Grid. We need to take advantage of these logical name spaces within the NVO services. At the same time, the TeraGrid needs to provide feedback on the choice of logical name spaces for whether they meet NVO requirements. The Metadata group recognizes four types of identifiers:

1. Unique identifier, based on an OID or handle
2. Logical name, used to organize a digital entity within a collection
3. Descriptive metadata, used to support discovery independently of the unique identifier or logical name
4. Physical file name

The SRB supports all four form of digital identify. A key requirement for NVO is consistency between these naming conventions. This can be cast as a decision to apply hard state management technologies to the mapping between these identifiers.

4.2 Interface Definition (Williams, CACR)

See WBS 5.4 Data Access Portals.

4.3 Network Requirements (Williams, CACR)

Status: Work not scheduled until late in CY2003.

4.4 Computational Requirements (Williams, CACR)

Status: Work not scheduled until late in CY2003.

4.5 Security Requirements (Deelman, USC)

Status: No work done in this quarter. The AstroGrid project in the UK is actively investigating authentication and resource allocation issues in the Grid framework, and we are likely to follow their lead in this area.

5 Data Access/Resource Layer

5.1 Resource and Information Discovery (Szalay, JHU)

Highlights: We have implemented a prototype of a resource registry for the NVO, which is distributed, yet unified. A resource description can be published at any participating registry node, and after a short time, that resource will appear in all the other nodes. The registry nodes communicate through the harvesting protocol of OAI (Open Archives Interface), an international standard for exchanging library metadata. The registry is being used to store metadata that is relevant to services, projects, data collections, and organizations.

We have used the registry prototype to implement a Data Inventory Service (DIS), which gathers and federates data about a given point in the sky. DIS uses services that have been registered with the distributed registry, meaning that it can use services as soon as they are published to the NVO.

5.2 Data Access Mechanisms (Deelman, USC)

See WBS 6.1 Computational Services for a description of the progress on data access protocols in the context of implementing the Montage application on the Grid.

5.3 Data Access Protocols (Williams, CACR)

See WBS 6.1 Computational Services for a description of the progress on data access protocols in the context of implementing the Montage application on the Grid.

5.4 Data Access Portals (Tody, NRAO)

Highlights: The first meeting of the IVOA Data Access Layer (DAL) working group was held in Cambridge, UK May 12-16, 2003. Agreement was reached on the concept of the data access portal for client access to VO resources, on the scope and high level architecture of the data access layer, and on the roadmap and priorities for IVOA DAL standards development over the next year. The highest priorities for the remainder of 2003 are a second version of the simple image access (SIA) protocol, and a first version of a simple spectral access (SSA) protocol, to be used for 1D spectra and SEDs. Agreement was reached on a prioritized list of enhancements for SIA V1.1. Work continues on design of a scalable data analysis framework to integrate conventional data analysis and VO.

Issues and Concerns: The SIA prototype defined in late 2002 simplified much of the technology involved in order to provide a functional image access interface for the first-year science demonstrations. For the second phase of DAL development we face some new challenges, e.g., development of underlying technology such as data models and a service registry, upon which the DAL services are based, and specification of standards for data access via an international collaboration.

Some key technologies required to define the new DAL services include the following:

- A working registry is essential to be able to effectively develop, deploy, publish, and use DAL services.

- One of the goals of SIA V1.1 (the second version of SIA, now in preparation) is to better characterize image data. This is being done by defining component data models, used to characterize dataset properties such as time of observation, spectral and spatial bandpass, spectral and spatial resolution, limiting flux, and so forth. Development of these component data models requires coordination of the IVOA DAL and data model working groups.

- To use component data models in DAL interfaces we need a way to represent data models in XML (VOTable). The current UCD mechanism was designed for a different purpose and does not provide the means to specify a "pointer into a data model." Several alternatives have been proposed for solving this problem, including extending the concept of UCD by adding a namespace, or by defining a new tag UTYPE in VOTable.

- An issue for simple spectral access is that there is no widely used standard format (e.g., in FITS) for storing spectra. We will have to invent such a format to provide a uniform interface to spectra. Multiple representations are possible, e.g., both XML (VOTable) and FITS. A similar issue is how to represent image data in VOTable, e.g., for transmission of small image cutouts back to the client.

Status: Development of international standards for VO data access is now being done by the IVOA DAL working group. All international VO partners, including the NVO, participate in standards definition. Implementations, such as data access services or client analysis applications, are more specific to each international partner. Given standard DAL protocols, multiple implementations are possible while still achieving interoperability.

Initiation of the DAL working group was the major activity for NVO data access this past quarter. The kick-off meeting of the IVOA DAL working group was held in Cambridge, UK, on May 12 2003, with follow-on meetings the rest of the week. The goal of this first meeting was to agree on what the VO data access layer is, what we would ultimately like to produce, and what we would like to accomplish within the DAL working group over the next year.

Specific working group agreements were achieved in the following areas:

- Concept of DAL portal
- DAL scope and high level architecture
- Principal data types within the scope of the DAL
- Mapping of data types to access services
- Priorities for implementing the data access services
- Roadmap and priorities for the next year
- Enhancements to SIA V1.1

The DAL portal (the so-called generic client interface) provides unified client access to VO data. This portal is the primary interface between client data analysis applications and the VO. Developers use the DAL portal to build distributed multiwavelength data access and analysis applications. DAL client applications see mainly the portal interface and are largely isolated from the underlying VO architecture.

The principal classes of data to be supported by the VO data access layer include the following:

- Source catalog (object, astrometric, photometric)
- Image (2D sky projection, spectral data cube, etc.)
- 1D spectrum and SEDs
- Time series
- Event and visibility data
- Generic dataset

Each principal type of data handled by the DAL has a corresponding data access service, which is specific to and optimized for that particular class of data. Each type of data has a corresponding data model, which is implemented by the service. Often the same data can be viewed via multiple services, e.g., synoptic or multi-band imagery could be viewed as an image, as a spectrum or SED, or as a time series. Event and visibility data could be viewed as a table or as an image, spectrum, time series, and so forth, depending upon the capabilities provided by the service provider and the type of analysis being performed by the client.

The highest priority services are image access, in particular 2D sky projections and spectral data cubes, 1D spectra and SEDs, and catalog access. Time series and the general NDImage come next, followed by access to calibrated visibility and event data.

In general the first year of DAL development emphasizes specification of the access protocols rather than reference implementations, which necessarily come later (although we have the science demos even in the first year or two). The priorities identified for the next 12-18 months are as follows:

- Simple Image Access (SIA) V1.1 (target: summer 2003).
- Simple Spectra Access (SSA) V1.0 (target: fall 2003).
- SIA V2.0 (target: summer-fall 2004). General image model.
- Better integration with VO standards, e.g., UCDs, VOTable.
- First steps for event and visibility data.
- Web Services versions of DAL services.

Simple image access (SIA) is currently at V1.0, the initial version released in the fall of 2002 and used in various demos in early 2003. Version 1.1 will be the first IVOA sponsored version of the SIA protocol. Due to the schedule, SIA V1.1 will retain the same form as V1.0, with evolutionary enhancements. More extensive changes will be deferred to V2.0. A key question for these first IVOA DAL protocols is whether we continue to prototype the underlying technology (with possibly delay in delivery), or press for needed technology development before releasing the next version of a standard.

Enhancements were discussed in the Cambridge meeting as well as via the mail exploders and telecons, and in small meetings, e.g., at CDS (with F. Bonnarel, F. Ochsenbein, and others) and at ESO (M. Dolensky and others). Planned enhancements include the first real registry support and integration, improved image characterization, e.g., to better define image provenance and identification, time of observation, and spatial resolution, further evolution towards formal data models, and normalization of UCDs including support for pointers into data models.

The top priorities for the remainder of 2003 are to release SIA V1.1 (for this to be worthwhile we need to evolve the underlying technology sufficiently first, e.g., the component data models), and to produce the initial version of simple spectral access. In addition we plan to continue research on scalable computational frameworks. A scalable data analysis framework is needed to integrate astronomical data analysis with VO, as well as to provide reference-grade framework software to implement scalable DAL services.

6 NVO Services

6.1 Computational Services (Berriman, IRSA)

The Caltech group has been working closely with the NSF-TeraGrid project, doing large-scale image mosaicing with the Atlasmaker software. Atlasmaker uses Montage (described further below), a new and rigorous code for mosaicing images, as well as other, faster ways to mosaic images. Atlasmaker is one of the TeraGrid "flagship" applications. Under NPACI funds, the code has been parallelized and scripted for high-performance, wide-area computation on TeraGrid, using SRB for (some) input images, but essentially for the distributed storage of the resulting atlases. There is new code for connecting to arbitrary image archives that are using the NVO publishing protocol (SIAP). The protocol allows for multiple retrieval mechanisms: if the input data is on an SRB system, it can be retrieved that way, or else through HTTP. Code has also been built for the creation of atlases — coherent collections of mosaiced images that lead directly to multi-wavelength imagery. We expect these atlases to be a new and powerful paradigm for knowledge extraction in astronomy, as well as a magnificent way to build educational resources.

As the TeraGrid matures, we expect to be computing large numbers of mosaics, each a reprocessing of a particular image survey to a particular page from an atlas. The results will be stored back in a single virtual file system managed by SRB, but physically located at SDSC, JPL, and CACR.

At IPAC, testing of the first public release of Montage has been completed. This release is designated version 1.7 and runs under Linux on single 32-bit processors. The test results and documentation have been submitted to the sponsor for their review and approval. We have already begun work on the next release of Montage, a Grid enabled version that will ultimately run in operations on the TeraGrid. Our aim is to develop an operational quality version of Montage that takes full advantage of the parallelization inherent in the Montage design and of software that runs jobs submitted to computing grids. Initially, the work is being performed on a 64-bit Linux machine at NCSA and 32-bit Linux machines at ISI, and once we have TeraGrid credentials and accounts, we will migrate the work to the TeraGrid following upgrades to its processors.

Earlier prototyping work (reported previously) revealed the following as important issues in developing an operational, Grid-enabled version of Montage:

- Generation of an abstract Directed Acyclical Graph (DAG) to describe the workflow.
- Re-projection of the images in parallel
- Overlap between images should be determined before re-projecting the images.
- Image files must have unique names.

We have generated an abstract DAG, in XML format, to describe the workflow. The overlap between the images is analyzed before generating the DAG since we cannot dynamically modify the DAG once it is submitted to Pegasus. The DAG describes the operations (e.g. Projection, background correction, co-addition) to be performed to get the final mosaic and the control dependencies between them. It includes the names of the image files to be used by the projection jobs. These images files are registered in the Replica Location Service (RLS) before the DAG is submitted to Pegasus.

Pegasus generates a concrete DAG to be executed on a condor pool. Currently we are running montage over a single condor pool only. This concrete DAG identifies the path of the montage executables on the particular pool and adds nodes for transferring the images files from the GridFTP URLs to the execution directory. Once the final mosaic is generated it is also registered in the RLS. Currently we are not registering the intermediate data products. They can also be made permanent and registered in the RLS by modifying their persistence attribute in the abstract DAG.

The projection jobs were made parallel in the DAG. There were 47 projection jobs in one of the test run. The scheduling of jobs on various machines was done by the Condor matchmaker. All the projection jobs were completed in about 20 minutes whereas it takes about 2 to 2.5 minutes for a single projection on a standard Linux machine having a Pentium 4 processor running at 2 GHz. This shows the excellent speedup achieved by making the projection jobs parallel. Calculating difference images etc can also be made parallel using a similar mechanism. It took about half an hour in all to execute the concrete DAG and generate the final mosaic.

Unique names are provided for the image files. These images can be used across different computations since they are registered in the RLS. However we would need a unique naming scheme for the intermediate data products and the final mosaic also in order to make them available for future runs and discriminate products ducts generated by different runs.

Once the compute infrastructure is in place, we will begin formal regression testing that will compare results on the 64-bit grid machines with those on single processor 32-bit machines.

6.2 Computational Resource Management (Moore/SDSC)

IPAC deployed a demonstration of the Request Object Management Environment through a server at Caltech. The demonstration can be found at http://irsatest.ipac.caltech.edu/applications/Rome/tutorial/intro.html, and includes a step-by-step tutorial. Astronomers place an order for a mosaic of IRAS data that will be generated by Montage (http://montage.ipac.caltech.edu). ROME provides the middleware for managing the requests.

The demonstration allows users to create an account (used by ROME as an identifier to allow them to find their own jobs), log in to the system, order multiple custom IRAS mosaic according to the user's specifications of position, mosaic size, coordinate system and equinox. An unlimited number of jobs can be submitted. ROME responds with confirmation messages and a job status web page that reports whether jobs are pending (accepted by ROME but not yet running on the server), processing, aborted and completed, and has links to messaging information and the final mosaic image. Because all job information lives on the server side, users can log out of ROME and return later on to monitor their jobs. There is a mechanism for filtering status requests, which will give update information on jobs meeting specific criteria: those that have taken more than four hours to complete, those that have aborted and so on. The demonstration also provides a request monitor deployed as a Java applet, which eliminates the need to reload web pages to update status message. A sample status page is shown in Figure 1 below. Finally, users can configure ROME according to their needs; for example, they can choose to be notified by email whenever each job is finished.

The demonstration successfully processed over 150 requests from NVO team members, including requests for very large mosaics, 45 to 60 degrees on a side, which took several hours to complete.

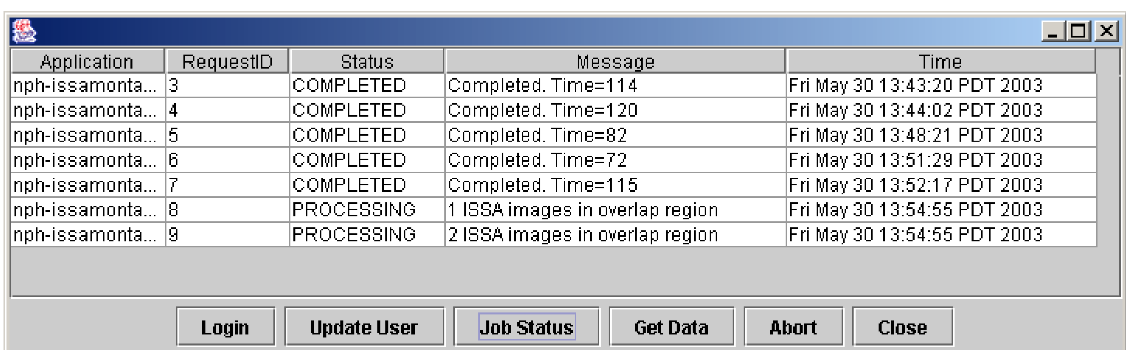

Figure 1: Sample Status Applet Showing the Status of Jobs Submitted to ROME

We have begun to plan future development of ROME, so that it interacts with all NVO data access and environments and the TeraGrid. Broadly speaking, the goals are as follows:

- Submit jobs to the TeraGrid via Condor-G (the next major milestone)
- Forward digital certificates as part of the data package passed from the original user to the ROME Processor. A ROME processor requires an authenticated user identity for a task to be executed. This information will have to be part of a registry that

describes ROME Processor access requirements. ROME will need to process these access constraints to match requests to the correct ROME Processors.

- Support interactions with NVO compliant data collections through the NVO registry. Support the shipping of data (images and catalogs) to Grid services that require these data as input.

7 Service/Data Provider Implementation and Integration

7.1 Service/Data Provider Implementation (Hanisch/STScI)

Status: The existing ConeSearch (catalog) and Simple Image Access Protocol services were entered into the prototype registry. Substantial work was required on the resource metadata, as the initial ConeSearch service registry contained only a subset of the metadata required by the Resource and Service Metadata document. We also discovered inconsistencies in metadata usage that had to be corrected.

7.2 Service/Data Provider Integration (Hanisch/STScI)

To support the Galaxy Morphology Demo at the AAS (Jan 2003) meeting, IPAC/NED created a "cone-search" program for searching through the NED archives and produced in XML formatted VOTable output with the resulting NED objects (Galaxies, Groups of Galaxies, Clusters, Quasars etc.), their best known positions, redshifts, and references.

While designing Simple Image Access Protocol (SIAP) to NED's Image Archive, we had to start by building layers between the NED archive and NVO services needs (due to the extreme heterogeneity of the imaging datasets in NED). This includes new DB tables and software, supporting these particular types of data and access to them.

The "FourCorners Table" (a metadata table with extracted World Coordinate System (WCS) keywords from the FITS headers and similar information for non-WCS compliant FITS and JPG images, provided to us by dozens of observatories and space missions) is completed. In the case of the 2MASS All-Sky Survey we have links to the data in the IRSA archive and also table entries for their sky coverage. This is very valuable to have in NED because it is not only just simply accesses the images, but all the objects are also cross-correlated and most of them have latest photometry, SEDs, and astrometry (which are constantly being updated in NED). After setting the groundwork of flagging each image with a WCS quality flag, we have constructed the first version of the metadata required to support SIAP queries.

Progress on the SIA (celestial coordinate based) search capabilities for the NED image archive: The search software, which makes use of, the WCS metadata to support SIAP queries is now in development:

a) a search program for extracting from the metadata tables a listing of images we have in the archive for a particular query, filter it by various conditions (e.g., being fully WCS compliant, FITS, but not WCS compliant, also might be by wavelength (color), date, etc.). For JPG formatted images one is going to be given a list of URLS with abbreviated information about color, context of images (radio maps or contour diagrams, rare scanned images from old atlases, and so on).

b) Providing XML_VOTable output for results, returned by the NED Image Archive. This step is now done.

c) Combining steps a) and b) in order to fully satisfy SIAP queries. Expect to have this ready for testing in September 2003.

We plan to start converting all other NED services into XML-for-NVO and for NED-specific formats, which will include complete "Cone Search" capabilities by object names and by positions.

IPAC/Infrared Science Archive:

The Infrared Science Archive has been evolving its architecture to support NVO protocols. Given that deployment of distinct NVO services would impose an unacceptable maintenance burden, we adopted the approach of evolving our services to respond according to the type of request made to them. For instance, if the request to the service is for HTML output, it will generate an HTML that will be displayed in the client's browser. If the request is for an inventory of a region, then the service will write information on the available data to a staging area. If the request is from an NVO-complaint program, then it returns catalog information in VOTable format, or image information that is compliant with the SIAP. This approach enables IRSA to support multiple modes with only minor modifications to extant services.

In more detail, the NVO web services (catalog cone search and an image metadata search) both work via an HTTP Get, and can be supported by minor modifications to IRSA's CGI services. The cone search is set up so the only free parameters are the CGI keywords `RA`, `DEC` and `SR` (sky location in J2000 decimal degrees and search radius in degrees). The rest of the information (the "base" URL) must be fixed. For example, while we have only one service, which searches any of our INFORMIX catalogs, each one must be "registered" with a base URL like:

```
http://irsa.ipac.caltech.edu/cgi-bin/Oasis/CatSearch/nph-
catsearch?CAT=ntmass:ext_src_cat_01&
```

to which three parameters can be attached:

```
RA=12.8&DEC=-33.4&SR=0.5
```

Similar remarks apply to the SIAP services.

For any IRSA service, which can return a table subset or image metadata list, there will be a special "raw data" mode that returns appropriate VOTables. This is quite easy to do. For example, the Cone Search "services" we provide are actually a slight reworking of the catalog access code we developed for OASIS. Two things were added: An "NVO" flag was set if the `CATALOG` parameter was detected, in which case the RA, `DEC`, and `SR` parameters were looked for instead of the normal locstr, *etc*.; and if that flag was set the output table was run through `tbl2votable` and the XML/VOTable results were echoed straight back to stdout (as mime type `text/xml`).

Finally, we have developed a prototype service*, Tbl2votable,* a C module used by the above services to convert column-delimited ASCII tables to VOTable format.

Our aim is to make all IRSA services NVO-compliant by October 2003.

8 Portals and Workbenches

8.1 Data Location Services (McGlynn, USRA/HEASARC)

Highlights: Intense activity took place in this area in preparation for the release of the Data Inventory Service at the IAU meeting in July 2003.

The data inventory service was completely revamped to use information from a data registry to determine the data of interest. This was done in two stages. In the initial stage the registry included basically the same information as had been used in the local static registry in the earlier GRB follow-up demo. This successfully demonstrated the query connectivity to the registry enabled dynamic modification of the registry to affect the services queried.

Once the program was able to establish connections to the registry, a thorough review of the registry contents, both the fields included for each service and the services available, was conducted. Many inconsistencies and short-term kludges that had been used in the initial registry were eliminated. The registry was made fully compliant with the Resource and Service Metadata document. The data inventory service was updated to use only information available from the registry. Modifications to the Aladin and OASIS services made them able to directly use the VOTables produced by cone search services. The final result was a much improved system with many more resources queried than the previous GRB follow-up service, and with a far more consistent and coherent registry.

Issues and Concerns:

Further development of the DIS service will be needed to appropriately filter services that are to be queried. The number of services returned to a user can already be substantial and may soon overwhelm the user if presented naively.

Status:

The service is ready for public announcement. Minor modifications are anticipated for the next month. More substantial revisions may be made after receipt of community feedback.

8.2 Cross-Correlation Services (Djorgovski, Caltech)

Status: No scheduled activities prior to CY2002 Q3.

8.3 Visualization Services (Williams, CACR)

Status: No scheduled activities prior to CY2003 Q3.

8.4 Theoretical Models (De Young, NOAO)

Status: STScI and JHU team members met with P. Teuben (U. Maryland) to explore the feasibility of integrating globular cluster simulations into the NVO framework. See WBS 10.2 for additional information.

9 Test-Bed

The TeraGrid is now available for friendly use. The components running on the TeraGrid have been characterized by L. Brieger:

- Globus GSI - This seems to be pretty reliable on the TeraGrid.
- GridFTP - available on the host/login nodes of the TeraGrid but not on the compute nodes. This means that parallel data transfers cannot be done from compute nodes to compute nodes at different sites, and trying to use the login nodes for massive data transfers will make for bottlenecks. Otherwise, it is reliable on TG.
- GRAM - the GRAM services allow Globus to run remote jobs. These are up and running reliably on TeraGrid.
- MDS services - in development. MDS is running on TG now, but not in final form. For now, NCSA is running an MDS GIIS serve (mds-TeraGrid.ncsa.uiuc.edu); each TG site (except possibly ANL) has a GRIS that reports to it. These are just Globus default MDS service reports, for now reporting test data.
- Condor - Condor-G has been a part of the systems all along
- MPI-IO/MPICH-G - For now, the MPI implementations on TG are MPICH, MPICH-GM
- (for using the Myrinet switch), and MPICH with VMI (for cross-site MPI). MPICH-G2 is in development and a delivery date is not known. Once GPFS is really stable on TG, MPICH2 may be installed in order to have MPI-IO functionality.
- Java - This is installed on TG.
- Perl - This has been used extensively on the TG test systems and is OK there. It's been less tested on the production systems, but is there.
- SRB servers are installed at SDSC, Caltech, and PSC. SRB clients are available at all sites, with ANL coming up this month.

10 Science Prototypes

10.1 Definition of Essential Astronomical Services (Szalay, JHU)

Status: It is clear that SOAP services could and should play an important role in the Virtual Observatory. The registry prototype service has shown the advantage of SOAP. There was a surprising amount of interest in the Web Services section of the IVOA meeting in Cambridge. Many people are implementing or trying to implement SOAP WebServices.

One of the major Services that are currently being defined is the Astronomical Data Query Language (ADQL) Service. Work on ADQL is being performed jointly with VO Japan. At Cambridge there was broad agreement to an ADQL based on slightly extended SQL, namely with some extensions for definition of Shapes. This together with other WebServices(e.g. VOTable Joins and CrossMatch) could form the basis for the more advanced Virtual Observatory Query Language (VOQL), or Problem Statement Language as it is called in some documents. It was generally agreed the transfer of ADQL between services should be done in the form on an XML document, which would be a parse tree of the ADQL statement. A proposal for this format has been produced and will be made publicly available after some iteration with VO Japan.

The notion put forward was that a Resource, which would support ADQL, would be required to implement a certain set of (SOAP) WebServices. A proposal for the definition these WebServices is underway.

The Hyperatlas project has formed under NVO influence, a collaboration of Caltech, Jet Propulsion Lab, and San Diego Supercomputer Center, defining standard WCS projections for astronomical images. A collection of these projections form the “pages” of an “atlas” that covers the sky in a uniform way, at a uniform scale. The project has created web services that give the standard projection corresponding to a given page number or point of the sky.

We are using Atlasmaker as part of a thrust to build digital “reference atlases” of the sky, based on these standard projections. Images that are projected to the same standard can be directly compared, pixel for pixel. We will be creating research projects to do advanced data mining in these reference atlases.

10.2 Definition of Representative Query Cases (De Young, NOAO)

Status: Discussions were initiated with NVO partners interested in theoretical models and simulations to see what science prototypes might be implemented in the coming six to nine months. P. Teuben (U. Maryland) visited STScI/JHU to discuss such possibilities in detail, and we are currently planning a demonstration project based on a globular cluster simulation data set. The prototype will address the question of mass segregation in globular clusters as ascertained from making simulated observations of a suite of models, and then comparing their "observed" properties to actual observations from HST, Chandra, and other observatories. Attention was diverted from this work in order to prepare the Data Inventory Service demonstration for the IAU, but will resume in July and August. The plan remains to show a demonstration based on the globular cluster simulations at the January 2004 AAS meeting.

10.3 Design, Definition, and Demonstration of Science Capabilities (De Young, NOAO)

Highlights: The planned work on converting the January 2003 gamma-ray burst follow-up service into a generic data discovery tool, the Data Inventory Service, was completed. The demonstrations at the July IAU meeting will incorporate a dynamic queryable registry rather than the internal static list of services used in Seattle.

Status: The original prototype remains fully operational. A distinct and simpler interface to the same basic service has been developed where the user gives only the position and size of the region of interest. See WBS 8.1 for more details. An enhanced version of the prototype that uses dynamic registries will be presented at the July IAU meeting.

A powerful new camera is installed on the 48" Ochsin telescope of the Palomar observatory in California, and is being used for the new synoptic sky survey called QUEST, a collaboration of Yale and Caltech. The QUEST data warehouse is being designed in the light of NVO protocols and methods, with data delivered through the SkyQuery system designed and implemented under NVO funding at JHU, with help from Microsoft.

11 Outreach and Education

M. Voit is leaving STScI and the NVO project on 1 August 2003. He will be replaced by F. Summers (STScI).

11.1 Strategic Partnerships (Voit, STScI)

A partnership has been formed with UC Berkeley, the American Museum of Natural History (Hayden Planetarium), and ManyOne Networks (http://www.manyone.net) to provide NVO content via the newly developed ManyOne web browser. UCB and ManyOne are taking an aggressive approach, focusing on an initial audience of the general public. Within the NVO project we have some concerns that, as a first EPO initiative, this is fairly risky: the general public is the least tolerant audience for a new product. However, we do wish to make NVO-enabled content widely available, and will work with these organizations to help make the initiative a success. (ManyOne is owned by a non-profit foundation and is committed to providing NVO content at no cost.)

11.2 Education Initiatives (Voit, STScI)

The requirements for resource metadata necessary to describe educational and outreach-related resources were defined and integrated into the Resource and Service Metadata document. Using these metadata definitions, a set of example resource profiles were developed to guide content providers in populating the resource registry.

11.3 Outreach and Press Activities (Voit, STScI)

The New York Times ran an article about the NVO in May:

> "Telescopes of the World Unite! A Cosmic Database Emerges," 20 May 2003, B. Schechter, *The New York Times*.

Activities by Organization

Caltech–Astronomy Department
A. Mahabal has continued working on the Topic Map applications in the VO context (see the previous report).

Jointly with the PSU and CMU groups, we have been developing a web-based astrostatistics service for the NVO. Much of this work has been supported separately through another grant. However, Mahabal has added some NVO-specific functionality at:

http://www.astro.caltech.edu/~aam/science/astrostat/index.html

The service currently provides several univariate and bivariate functions, and some basic statistical plotting functions. The available set of statistics will be greatly expanded in the near future. The VO-oriented enhancements are that the data input can be in the form of VOTables or ASCII, and columns from URLs or local files can be chosen directly using the information in header files.

An interesting and immediately useful addition has been the interface with some of the VO cone search services. It is now possible to output the results from cone searches directly into the various astrostatistics routines. These pages can be found at:

http://www.astro.caltech.edu/~aam/science/astrostat/cones.html

Not all cone search services have currently been incorporated, but we plan to do so in the future.

We have also started exploring the use of Genetic Algorithms (GA) for data explorations in the image domain. Currently we are trying out a proof-of-concept galaxy morphology GA package.

We are developing several possible scientific demo cases for the VO, including outlier searches in parameter spaces, and exploration of the time domain.

Finally, we have started a preparatory design work for the VO inclusion of the new Palomar-Quest sky survey. A number of VO-relevant issues, including database design, interfaces, interoperability, etc., are being explored. One novel aspect is the suitable database design for synoptic sky surveys, which will facilitate exploration of the time domain.

We are also starting an investigation of Grid services in the VO context, which we now expect will be funded separately through another grant.

Caltech–Center for Advanced Computational Research
At Caltech, a postdoc (M. Graham) began work on July 7, half time on NVO and half time on the Quest synoptic sky survey. The data processing and dissemination of Quest results will be based on NVO protocols.

R. Williams leads the international discussion group on UCDs (Unified Content Descriptors), an emerging shared semantic vocabulary for VO. Major revisions to the structure of UCDs were discussed at the Cambridge (UK) interoperability meeting in May.

Caltech worked closely with other NVO organizations—specifically NCSA and STScI—to implement the prototype NVO resource registry. Caltech created a local registry using the OAI (Open Archives Initiative) metadata harvesting protocol, the contents of which were ingested into the central NVO registry at STScI/JHU.

Caltech continued work on the Hyperatlas project, defining standard WCS projections for astronomical images. A collection of these projections form the "pages" of an "atlas" that covers the sky in a uniform way, at a uniform scale. The project has created web services that give the standard projection corresponding to a given page number or point of the sky.

Caltech–Infrared Processing and Analysis Center
During this period IPAC
- Began a collaboration with ISI to develop the infrastructure to support a Grid enabled version of Montage.
- Deployed a fully documented demonstration of ROME.
- Began planning for future development of ROME.
- Developed NVO compliant services that support the richness and diversity of data served by NED or accessed through NED.
- Matured an architecture that ensures NVO compliance of IRSA services with minimal coding and modest maintenance costs.

Canadian Astronomy Data Centre/Canadian Virtual Observatory Project
The CVO Linux/DB2 parallel database machine was brought into operation during May-June. This 16-processor parallel machine has 7-Terabyte capacity. A DBA was hired in April to work solely on the CVO system.

The CVO database exploration prototype (released to the public in February 2003) was migrated to the DB2 machine in preparation for public release in July prior to the IAU demonstration and was functional in June. Performance testing and tuning continues.

Vigorous collaborations with the German VO (GAVO) and the Australian VO projects continued with the goal of ingesting metadata for the ROSAT All Sky Survey and the 2QZ spectroscopic survey from the AAO in time for the IAU demo. The multi-wavelength query-level data model developed for WFPC2 was extended to accommodate these new instruments. The archival data for 2QZ and ROSAT will reside in Australia and Germany respectively and will be retrieved remotely. The CVO Prototype now contains 100,000 datasets and 25 million sources.

A substantial revision of the CVO Prototype software was completed during April-June to accommodate new content and enable new functionality.

Alberto Micol (ST-ECF) and Richard Hook (STScI/ST-ECF) visited CADC to continue work on the WFPC2 Associations Phase II project and to prepare for a similar project with the Advanced Camera for Surveys, focusing on the multi-drizzle algorithm.

The CVO project will be part of three demos at the IAU: the CVO Prototype, the US-NVO Galaxy Morphology demo, and the AVO demo.

Carnegie-Mellon University/University of Pittsburgh
As our contribution to both the NSF ITR NVO proposal and the NSF FRG proposal (through Penn State), we continue to develop web-services and web interfaces to our fast and efficient algorithms. This quarter we have developed prototype systems at:

http://skyservice.pha.jhu.edu/colberg/NPoint_WF/WebForm1.aspx
http://www.astro.caltech.edu/~aam/science/astrostat/npt.html

We plan to continue these efforts through extensive testing before making them available to a wider public.

Fermi National Accelerator Laboratory
V. Sekhri developed both an SIAP and a cone search interface to the imaging data in the SDSS DR1 data release. The SIAP interface returns URLS to image cutouts of objects in the object catalog — it does not make cutouts exactly at a user specified position. A second service provides access to the FITS files of the cutouts themselves. The DR1 catalog covers 2099 square degrees and has 53 million objects. An interface to spectroscopy awaits clarification of the SIAP protocol (WBS 7.2.1). The data are already online; just the SIAP interface needs to be implemented.

V. Sekhri and J. Annis worked on developing an improved version of the galaxy morphology demo. This work involved creating a front-end galaxy morphology web service in addition to the SIAP service and integrating them with a Grid service that runs the GriPhyN virtual data toolkit. The galaxy morphology web service accepts a query much like an SIAP query and eventually returns a VOTable containing URLS to galaxy cutouts and galaxy morphology parameters. The cutouts are fetched from the SIAP service. The morphology parameters are generated dynamically using Grid services and the Chimera virtual data application. The Grid computing component currently creates one job per galaxy, which is extremely inefficient. A goal for next quarter is to have one job run per cluster (WBS 10.3.1).

In related work, V. Sekhri continued work on a 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 (WBS 5.2.2). This project (which is largely for use by IVDGL) is expected to continue through December.

High Energy Astrophysics Science Archive Research Center
HEASARC personnel attended the Cambridge interoperability workshop, demonstrating the data inventory service and particularly discussing the development of VO registry services. They participated in numerous telecom and mailing list discussion of VO topics.

The development of the DIS service was the main focus during this quarter. A high degree of collaboration with personnel at JHU and STScI was involved as the HEASARC built the actual DIS service and linked it to the Aladin and OASIS system, while JHU and STScI built and refined the registry and registry access services.

Work on defining metadata for observation tables, especially as they relate to high-energy observations was begun. This involves looking at the SRM descriptions and existing UCDs and seeing how well they may be used to describe the HEASARC's primary observation tables and archives. A preliminary report on this activity will be made in the third quarter.

Related Activities: The NASA ITWG effort continued to work to develop a distributed service for the verification of dataset identifiers. Services were deployed at the HEASARC and several other sites and communications with the astronomy journals on using this service are ongoing.

The ClassX classifier continued to successfully use the VOTable format to access information from a number of multiwavelength catalogs.

Johns Hopkins University
T. Budavari has made a CASService available, which returns VOTable or FITS files (and other formats) on submission of a query to the SDSS database. An information page, simple java client, and a simple Python client are available at: http://skyservice.pha.jhu.edu/develop/vo/casclient.html. In conjunction with W. O'Mullane the list of JHU services have been annotated with examples at: http://skyservice.pha.jhu.edu/develop/vo/index.html

W. O'Mullane has done much work on the Registry prototype to make it ready for the IAU demos. Extra web pages were added and specific changes to match the RSM and facilitate DIS were made. This work is carried out in close collaboration with STScI and HEASARC. (http://sdssdbs1.stsci.edu/nvo/registry/index.aspx)

S. Carliles has integrated the registry prototype service and JAVOT and SAVOT with Mirage (a data analysis tool), A form in mirage now allows the user to submit cone requests to a Cone Service. A list of Cone services is created by querying the registry. (http://skyservice.pha.jhu.edu/develop/vo/mirage/mirage.html)

V. Haridas is working on the definition of the ADQL XSD for exchanging ADQL queries. The grammar (based on SQL with CIRCLE as an extra function) has been defined and a draft of the XSD will be available soon. Haridas has also made the FITSIO library available through C#.

M. Nieto 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.pha.jhu.edu/dr1/en/tools/chart/). This is essentially a suite of ASP clients that connect to the .NET web service that provides the image cutout functionality. A description of this service was included in the previous quarter's report.

G. Fekete visited Fermilab to discuss with J. Annis the deployment of VDT (virtual data toolkit) at JHU. This will be deployed on a new cluster that we are in the process of ordering as a Condor/Grid testbed.

Microsoft Research
J. Gray helped with SDSS DR1 SkyServer in cooperation with the JHU team. He built a QSO catalog for it in cooperation with R. Lupton, A. Szalay, and G. Richards. Considerable time was spent designing the site, doing performance analysis of it, and bringing it online.

Gray spent a week with the AstroGrid team in Edinburgh talking about the US efforts and helping them get the SSS and WFCAM archives online. There is also interest from both Edinburgh and Cal Tech to federate more datasets with the SkyQuery service.

National Optical Astronomy Observatories
M. Fitzpatrick attended the Cambridge IVOA meeting where he participated in the Data Access Layer and Data Model Working Groups and plenary sessions.

F. Valdes participated in the development of the data model and metadata standards as they apply to spectroscopy. Valdes contributed two documents to the discussion in these areas: "Incorporating Spectra in the Next Phase of the Virtual Observatory" (http://iraf.noao.edu/projects/vo/dal/specsiap.html) and "A Virtual Observatory Data Model" (http://iraf.noao.edu/projects/vo/dal/datamodel.html).

D. De Young attended the NVO team meeting in April, and participated in weekly Executive Committee telecons, biweekly WBS Level 2 telecons, and in IVOA telecons. He worked with G. Helou on the science part of the project roadmap.

.

National Radio Astronomy Observatory
Most of the effort at NRAO this past quarter was associated with organization and initiation of the IVOA data access layer (DAL) working group, which D. Tody (NRAO) chairs. The IVOA DAL working group will generate international VO data access standards, which will also be the basis for the NVO data access portals and science demonstrations.

Specific activities included

- Calls for proposals for SIA enhancement.
- Preliminary whitepaper on a scalable data analysis framework.
- D. Tody attended the NVO team meeting in Pasadena (Apr 3-4).
- Presentation on DAL phase II at NVO team meeting.
- D. Tody chaired the DAL working group meeting in Cambridge (May 12).
- Questionnaire and inventory of spectral data archives.
- Discussions of UCDs and data models.
- Discussions of open data formats for radio astronomy.
- D. Tody visited CDS Strasbourg Jun 24-25 for SIA/UCD/DM discussions.
- D. Tody visited ESO/AVO Jun 27 - Jul 2 for discussions of simple spectral access and data analysis frameworks.

All VLA observations back to 1976 have now been loaded into the NRAO archive (2.5 TB total). Loading of the VLBA archive (10 TB in total, if we load it all) has commenced. New data from VLA, VLBA, and GBT is being loaded into the archive within several days of acquisition. Release of the NRAO archive to the public is planned for October 2003. Beginning in 2004 the proprietary period for new observations will be 12 months. Replication of the NRAO archive at NCSA is still planned: current activities include acquisition of a 1-2 TB disk box for sneaker-net replication of the archive.

Work has begun on a prototype pipeline for ALMA data. The intention is to use this technology later for EVLA and other NRAO telescopes as well. When completed, this will allow generation of calibrated visibility data as well as reference images for the NRAO and ALMA archives, with publication of all such data to the VO. The same technology could be adapted to provide on-the-fly imaging for VO data access to visibility data.

One of the priorities identified by the IVOA DAL working group was "first steps for event and visibility data." As part of this initiative a questionnaire and survey was generated by Peter Lamb (CSIRO) and Anita Richards (Jodrell Bank) and posted to the radiovo mail exploder. Several sites including NRAO have since responded. The ALMA project (F. Viallefond, Paris) has produced a draft specification for a formal data model for ALMA data based on the AIPS++ measurement set. This could serve as the basis for a future general data model for visibility data, which would consist of a general core data model plus telescope-specific attachments, possibly as a radiovo standard. External representations in XML (VOTable) and probably FITS could also be defined. A. Wicenec (ESO, ALMA) is preparing a proposal to modify VOTable to add support for binary attachments to support such applications.

Raytheon Technical Services Company
The Raytheon Technical Services Company (RTSC) provided support in the following activities:

Project-wide. RTSC staff participated in the NVO Project Team meeting at Caltech in April and at the IVOA registry workshops in Cambridge, UK. Staff gave talks on the VO

Query Language (VOQL) and on the NVO project CVS software repository that we implemented for the rest of the project team. Staff also participated in on-line discussion groups, including several IVOA-sponsored e-mail discussion lists. RTSC staff has taken a leadership role as co-chair of the VOQL working group, formed out of the IVOA registry meeting.

WBS 2: Data Models. RTSC staff have been working on the VOQL and on the data model for the IVO "quantity" object. The main activity has been the VOQL effort. RTSC staff have been developing this high-level query language to allow scientists to query the VO even when those scientists are not familiar with the various data centers and their particular organization of data. A schema for the high-level language is now well developed, but work is still ongoing in the area of transformation of the VOQL query into lower-layer data access within the VO distributed environment. RTSC staff have set up a collaboration with CADC staff, to address design, experimentation, and implementation of VOQL. In addition to these activities, RTSC staff created an extension of the VOTable schema that allows the data to be arranged by columns as lists of data rather than by rows of tagged data cells. This allows better XML typing and validation of the data, plus it results in a substantial reduction in the number of XML tags with concomitant reduction in data volume and transfer times. Staff further extended the schema to allow for arrays within a list of data.

WBS 3: Metadata Standards. RTSC staff continued to participate in and support the Metadata Working Group, including the weekly telecons, with particular emphasis on the NVO registries and the VOQL (VO Query Language). Staff participated in testing the Caltech version of the VO Registry by using actual data sets. Staff are participating on the NASA-wide XML working group, established by the NASA CIO, to develop a NASA XML implementation and work plan. As part of this latter effort, staff provided descriptions of the metadata activities of the NVO project and the corresponding XML applications.

WBS 10: Science Prototypes. RTSC staff are participating in the IVOA online discussions (RWP02) regarding the registry requirements that are needed to support science demos and science use cases. Under the auspices of other research funds, staff is continuing to investigate scientific data mining techniques for a specific astronomy research project, but with a long-term goal of applying these techniques within the NVO. Staff attended a Scientific Data Mining Workshop and the International SIAM data mining conference in May 2003, plus gave a talk at the SPIE data mining conference in April 2003. In each of these cases, staff presented the NVO project plans and applications of data mining technologies to the NVO. Staff also gave several talks on the topic of "Distributed Data Mining with the NVO": at the FDA Office of Drug Safety; at the University of Maryland-Eastern Shore department of mathematics and computer sciences; for the NASA Goddard Space Flight Center's ESDIS Project Office, Science Data Systems Branch, and Advanced Data Management Branch; and at the annual Science Data Centers Symposium [http://www.sci-datacenter.org/]. Staff also contributed to the NASA Minority Universities Research (MUCERPI) program by serving as a consultant on a minority university research proposal involving students in NVO-like data mining research exercises using variable star databases (e.g., MACHO). Staff also contributed NVO knowledge and insights to several other NASA VO-like projects, including: (a) the LWS (Living With a Star) and Magnetospheric VO projects; and (b) the NVO as an innovative concept for the Intelligent Archives of the Future project (funded by the NASA Intelligent Systems Program).

San Diego Supercomputer Center

SDSC continues support for the formation of an initial NVO testbed. The goal is to support large-scale analysis on replicas of collections that are located near the computational resources. The expectation is that consistency can be maintained across the replicated collections through use of the SRB data Grid technology. Tasks that have been completed in the last three months include:

- The registration of the USNO-B catalog into a SRB Data Grid is now underway. Disks are shipped to Flagstaff, where data is loaded, and then transferred to SDSC for installation in a Grid Brick.
- The SDSS DR1 data is being loaded into the SRB Data Grid, for high-speed access on the TeraGrid.
- A friendly user period on the TeraGrid is now in effect through the end of December. We plan to use Montage to re-project 2MASS images in a large-scale computation on the TeraGrid.
- Implementation of a test version of the SDSS catalog. V. Nandigam has created the schema in DB2 and worked through the porting issues for the tables, etc. He's been working on porting the stored procedures and triggers, but has had some difficulty. This was mostly due to calls to system tables for metadata about the tables that are not accessed in the same way in DB2. He has six of the stored procedures complete and is working on the remainder. The catalog has been implemented in DB2, on a 64-processor Sun server.
- Implementation of the USNO-B catalog. The USNO-B schema has been created and V. Nandigam has some test data loaded. The rest of the catalog will be loaded in July.
- A new release of the SRB has been created, version 2.1.1 and installed on the NVO testbed. This provides bug fixes, integration with GSI 2.2, new bulk load loading capabilities.
- We have registered a small number of MACHO images into a SRB Data Grid. This required the installation of a SRB server at ANU. Discussions with J. Smillie (ANU) are now underway for the selection of a collection for publication through the SRB. R. Hanisch examined the FITS header information that was provided with the MACHO images and had the following comments:
    - Provision of an SIA is needed to the collection. This requires world-coordinate information. There are the RA and DEC keywords (not FITS standard) but no information about pixel size (in degrees on the sky) or orientation. Perhaps this is the same for all MACHO images, and a proper WCS can be created.
    - The SRB could be used to update headers with this information, either on-the-fly in the SIA service or through back-end reprocessing.

- A minimalist WCS for a FITS image must include
  - CRPIX1 the reference pixel
  - CRPIX2
  - CRVAL1 the celestial coordinate value at the reference pixel (decimal degrees)
  - CRVAL2
  - CDELT1 the pixel size in decimal degrees, at the reference pixel
  - CDELT2
  - CTYPE1 the coordinate geometry and axis type, e.g., 'RA---TAN'
  - CTYPE2 'DEC--TAN' for a tangent plane projection with the point of tangency at CRPIX1, CRPIX2
  - CROTA2 the rotation angle of the coordinate system from north-up

Smithsonian Astrophysical Observatory
SAO continued to lead the Data Model design (WBS 2.1, 2.2) and the Metadata design (WBS 3.1) efforts.

Personnel attended or participated in the following meetings:
- J. McDowell and A. Rots attended IVOA conference in Cambridge, May 2003.
- J. McDowell led a DM Working Group meeting.
- G. Fabbiano was invited to present a talk on the Virtual Observatory at the Astronomy Roundtable of the SLA (Special Libraries Association) 2003 Annual Conference in New York City (June 9th 2003). The talk generated quite a bit of interest and it was urged that we keep the librarians in the loop.
- Regular team and metadata telecons.

Space Telescope Science Institute
STScI staff planned and participated in the project team meeting, April 3-4, at Caltech. STScI staff contributed to a paper for the Supercomputing 2003 conference entitled "Grid-based galaxy morphology analysis for the Virtual Observatory."

R. Hanisch attended and co-organized an International VO Alliance interoperability workshop in Cambridge, UK (May 12-16), which was attended by nearly 60 people representing the VO projects worldwide. Substantial progress was made in seven work areas: registries, data models, data access layer, VO query language, uniform content descriptors, VOTable, and web services.

An NVO prototype registry was jointly developed at STScI and JHU. This registry functionality was successfully demonstrated at the IVOA meeting in Cambridge. This registry was loaded with NVO cone search registry and SIAP services and utilizes SQL Server database. A harvester was written and integrated with this registry database to extract services from OAI publishing registries at Caltech and NCSA and import them into this full registry. It incorporates parsing of the emerging standard VOResource.xsd schema, a standard format for expressing metadata, which describes an astronomical resource or service. Several Web Services were written to load, query and update the registry entries.

The NVO registry prototype was then fully integrated into the Data Inventory Service. The registry will be demonstrated at the IAU in July. The registry now contains 95 unique astronomical resources described with standard VO Resource metadata retrievable in XML or XSL rendered formats. Coordination between JHU and STScI has been very successful in setting up a NVO mirror registry. While the primary SQL Server database is located at STScI, there is a backup and development site at JHU that is also fully operational. The STScI SIA services and HST cone services have been updated to meet requirements for the GSFC DIS service.

The Resource and Service Metadata definition document was updated (to Version 0.7) and distributed to the IVOA Registry Working Group. RSM V0.7 forms the basis for the registry services described above.

A draft documentation standards process was defined and distribution to the NVO and IVOA Executive Committees. The process is based on W3C standards, but adapted to VO project needs.

An agenda was set for the June 2 IVOA Executive telecon. In preparation, R. Hanisch met with the head of the European Grid for Solar Observations to discuss the nature of the working relationship between the astrophysics and solar VO initiatives.

F. Summers is assuming responsibility for the NVO EPO program. Discussions were initiated with collaborators at UC Berkeley in setting up an experimental EPO portal for NVO. EPO metadata was also incorporated into the Resource and Service Metadata definition document, following guidelines developed by M. Voit.

United States Naval Observatory
S. Levine attended the April team meeting. Preparations were made for shipping USNO-B catalogs to SDSC for installation in the Storage Resource Broker.

University of Illinois Urbana-Champaign/
National Center for Supercomputer Applications
R. Plante continues to chair the weekly telecons of the Metadata Working Group. The primary focus of the MWG agendas this quarter have been on

- Resource metadata and identifiers
- Resource registries
- Preparation for the IAU Assembly in July

R. Plante and R. Williamson continue to concentrate on research on metadata and resource registries. This has focused on three key fronts:

1. *Resource Metadata Definitions.* R. Plante continues to contribute to the evolution of the Resource and Service Metadata document (RSM). R. Plante and R. Williamson, likewise, continue to refine the VOResource XML Schema accordingly. R. Plante leads the Resource Metadata Work Package of the IVOA Registry Working Group

and is collaborating in the current review of the resource metadata model. R. Plante has begun work on a general style guide for metadata definitions in XML Schema.

2. *Resource Identifier Specification.* Through the IVOA Registry WG, R. Plante moderated the development of a specification for resource identifiers; he is currently editing a Working Draft for submission to the IVOA standards process.
3. *Registry Prototyping.* R. Plante and R. Williamson participated in the NVO Registry "Tiger Team," aimed at providing a registry prototype for the Data Inventory Service. R. Williamson developed a deployable, publishing registry package.

R. Plante continues to enhance the Galaxy Morphology Demo in collaboration with E. Deelman (ISI) and J. Annis (Fermilab). Plante demonstrated it during a keynote address at the NCSA Alliance All-Hands Meeting in May. The demo team wrote a technical paper describing the demo, which was submitted to and subsequently accepted by the Supercomputing 2003 conference. This paper will be presented at the meeting in October.

Finally, R. Plante contributes to the various IVOA working groups. In addition to the Registry Working Group, he is leading a data model development effort dedicated to scientific quantities.

University of Pennsylvania

P. Protopapas has implemented an SQL database to host all MACHO lightcurves using NVO standards; it has been populated with a small subset of the whole dataset. This database enables lightcurves to be accessed transparently, and for complex operations to be performed within the database. It supports SIAP, ConeSearch, and VOTables (for returned results). In addition, Protopapas is preparing a web service that supports the NVO framework, including a capability to provide the union of multiple VOTables.

Penn participated in the design discussions regarding metadata concepts. P. Protopapas participated in the discussion group for metadata standards and VOQuery language.

University of Southern California (ISI)

Tasks undertaken by USC/ISI during the April 2003-June 2003 quarter:

- Porting Montage v. 1.7 to the Pegasus framework. ISI is collaborating with IPAC to schedule the Montage computation onto the Grid. IPAC describes the workflow in abstract terms and passes it to Pegasus for mapping and execution. Currently the initial phases of the workflow can be efficiently scheduled.
- Adding new features to Pegasus to support large-scale applications, such as those targeted by NVO. ISI added a feature to Pegasus to interface to bulk operations of the Replica Location Service. This increases the performance of Pegasus by enabling queries for multiple replicas at the same time.
- Adding support in Pegasus to interface to the PBS scheduler. ISI is investigating the possibility of adding support to Pegasus, which will allow the workflows to be scheduled on resources controlled by PBS. Among such resources as the TeraGrid machines.

University of Wisconsin
No activities to report for this Quarter.

Publications and Presentations

Borne, K. D. 2003, "Distributed Data Mining in the National Virtual Observatory, SPIE Conference "Data Mining and Knowledge Discovery", Volume 5098, pp. 211-218.

Deelman, E., Plante, R., Kesselman, C., Singh, G., Su, M., Greene, G., Hanisch, R., Gaffney, N., Volpicelli, A., Budavari, T., Nieto-Santisteban, M., O'Mullane, W., Annis, J., Sekhri, V., Bohlender, D., McGlynn, T., Rots, A., & Pevunova, O. 2003, "Grid-Based Galaxy Morphology Analysis for the National Virtual Observatory," Supercomputing 2003, accepted.

Acronyms

AAS	American Astronomical Society
ADEC	Astrophysics Data Centers Executive Committee (NASA)
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
DAGMan	Directed Acyclic Graph Manager (Condor)
DAL	Data Access Layer
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DM	Data Model
DOE	Department of Energy
DPOSS	Digitized Palomar Observatory Sky Survey
DTD	Document Type Description
EDG	European Data Grid
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
GSC	Guide Star Catalog
HEASARC	High Energy Astrophysics Science Archive Center
HTM	Hierarchical Triangular Mesh
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
MACHO	MAssive Compact Halo Object
MAST	Multi-mission 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 Archive Initiative
OASIS	On-line Archive Science Information Services (IRSA)
OGSA	Open Grid Services Architecture
OIL	Ontology Inference Layer
PB	petabyte
PSL	Problem Statement Language
Q	quarter
QSO	Quasi-Stellar Object
RC	Replica Catalog
RDF	Resource Description Framework
RLS	Replica Location Service
ROME	Request Object Management Environment
RSM	Resource and Service Metadata
RTSC	Raytheon Technical Services Corporation
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
STScI	Space Telescope Science Institute
SWG	Science Working Group
TB	terabyte
UCD	Uniform 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 System 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