Annual Report
October 2001—September 2002

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement
AST0122449

NATIONAL VIRTUAL OBSERVATORY

Contents

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

Period covered by this report: 1 October 2001 - 30 September 2002
Submitted by: Dr. Robert Hanisch (STScI), Project Manager

Executive Summary

In the first year of this project, substantial progress has been made on all fronts: programmatic, technical, and scientific. As we begin the second year of the project, we are poised to make our first public science demonstrations, building upon substantial technical developments in metadata standards and data access protocols. We have been successful in engaging nearly all participating organizations in substantive work. The NVO project is co-leading international VO initiatives, including the formation of the International Virtual Observatory Alliance, for which the NVO Project Manager serves as chairperson.

NVO Science. At the spring project team meeting (Tucson, 16-17 April 2002) three scientific demonstration projects were selected from an extensive list of potential projects compiled by the Science Working Group. The demonstrations were chosen based on a number of criteria, including availability of necessary data, feasibility of completion by January 2003, and ability to show results in a matter of a few minutes (i.e., the time one can typically hold the attention of an astronomer passing by a display booth at an AAS meeting). The selected demonstrations are

- Brown dwarf candidate search.
- Gamma-ray burst follow-up service.
- Galaxy morphology measurement and analysis.

These are described in more detail in WBS 10.1 and 10.2 of this report.

Next year we will develop more complex science demonstrations, and these will incorporate data from our international partners. A major milestone is the August 2003 IAU General Assembly, where we will unveil a second round of demonstrations and participate in a Joint Discussion on virtual observatories and new large telescopes.

NVO Technology. In collaboration with the European virtual observatory development projects, AstroGrid and AVO, we released V1.0 of the VOTable XML formatting standard for astronomical tables. Using VOTable as a standard output product, some 50 "cone search" services were implemented by 7 different groups within the team. The cone search services respond to a request for information based on a right ascension, declination, and radius about that position. Four software libraries for parsing VOTable documents were written and made available via the team web site. Also, a JHU-based team developed a catalog cross-correlation service for SDSS, 2MASS, and FIRST using Microsoft's .NET facilities and won second place in a nationwide software development contest.

During the summer of 2002 we developed the specification for a Simple Image Access Protocol, and by the end of the first project year several implementations had been completed. By combining the cone search and SIA services we have the infrastructure necessary for implementing the science demonstration projects.

Substantial progress has been made on metadata standards, work that supports both the VOTable and SIA specifications. In addition, a standard for resource and service-level metadata has been developed based on the Dublin Core. This standard has been widely reviewed and discussed among the international VO projects.

Next year we will begin to explore methods for creating industry-standard Web Services, and for deploying our initial http and cgi-bin services through WSDL.

The NVO Project. Despite the somewhat lengthy negotiation process that was required to place all of the subawards under this project, participating organizations were generally able to start work within the first several months. It is a challenge to coordinate work and fully exchange information within a collaboration of this scale, but through a system of working groups, project status reviews, and regular team meetings we have established effective communication and cooperation. The project Executive Committee meets weekly by telecon to address issues as they arise.

The delays in getting all subawards issued led to cost underruns in Year 1. These will be rolled forward to Year 2, and further to Year 3, to help smooth out the strongly front-end loaded funding profile. Financially the project is in good shape.

Activities by WBS

1 Management

1.1 Science Oversight

The Executive Committee has taken a direct interest in the progress on the selected science demonstration projects:

- A brown dwarf candidate search
- A galaxy morphology analysis
- A gamma-ray burst follow-up service

These have been monitored closely, and when issues have arisen or progress been less than expected, the EC has intervened accordingly. It will be a challenge to complete all three demonstrations in time for the January AAS meeting, though we remain optimistic of attaining success.

Two members of the EC, R. Hanisch and D. De Young, team member G. Fabbiano, and EPO collaborator J. Mattei, are members of the Astrophysical Virtual Observatory Science Working Group. We are following the AVO science demonstration developments and will work with the AVO, AstroGrid, and other international VO projects to develop science demonstrations in the second year of the project that draw upon data resources and information services from all international partners.

1.2 Technical Oversight

The Executive Committee is also directly involved with technical development activities: metadata standards, interoperability protocols, and web services. We have been actively involved with the IVOA (International Virtual Observatory Alliance) to build a single internationally accepted Simple Image Access Protocol, a follow-on to prior success in establishing the VOTable standard.

We maintained a web site, http://us-vo.org, for the project. This includes a document management system that allows team members to publish documents directly, without going through a human web master. The system already has over 40 documents. The web page also contains archives of several active discussion groups that are associated with the NVO (http://archives.us-vo.org). These include the very active Metadata and VOTable discussion groups each with several hundred messages. A new discussion group, "semantics," has been set up to discuss application of Knowledge Engineering technologies such as DAML-OIL and Topic Maps to astronomy.

1.3 Project and Budget Oversight

Performance Against Schedule. We are on or ahead of schedule in most activities. The detailed project plan shows progress (estimated percent completion) to date. Some scheduled activities need to be modified to reflect changes in approach.

Performance Against Budget. We did not spend the full first-year funding for the project owing to complications in issuing subawards and the associated delays in hiring at many organizations. Many of our university-based team members operate on quarterly billing cycles, and have mechanisms for covering costs internally until invoices are issued and payments are received. It has been difficult, therefore, to have a very accurate picture of to-date spending. Based on invoices received and known commitments, we expect to carry forward approximately 40% of our first year budget. We have made some budget reallocations within the project, moving responsibilities and associated funding to organizations that have been the strongest contributors. One senior member of the team relocated from one participating organization to another, taking responsibilities and work areas with him; SOWs and budgets were adjusted accordingly.

2 Data Models

2.1 Data Models / Data Model Architecture

We established a mailing list for data model discussions (dm@us-vo.org) and began work on proposed nomenclature. J. McDowell visited Strasbourg for the interoperability workshop and held discussions with M. Louys and F. Genova to establish a collaboration with the AVO data model effort. A draft document on the data model architecture has been circulated among the team and to members of the international collaborations. Fruitful discussions at the April NVO team meeting in Tucson and at the VO conference in Garching have led to agreement on a basic approach, in which we will make small models of aspects of the data and agree on a mechanism for associating such models with datasets and representing them in formats such as VOTable. A document describing the modeling of spectral bandpasses was also written and circulated.

The SAO group has begun modeling existing datasets and elaborating the possible components of the data model. A detailed comparison of the CDS Aladdin image archive model and the CXC X-ray data model was carried out and distributed to the team to stimulate discussion.

2.2 Data Models / Data Types

We have established that both images and catalogs have many common attributes; the information content of the CDS catalog description file is closely matched by the information content required to describe image axes. Our investigations emphasize the need to support, at a fundamental level, mosaiced images such as those made by HST and modern ground based imagers.

We studied image data formats from the archives of participating organizations, and established the importance of unifying the different mosaic image formats (four main variants were identified). These issues and a proposed general approach were described in a talk at the Garching VO conference.

2.3 Data Models / Data Associations

During the Strasbourg discussions we addressed issues of data quality (WBS 2.3.4) as an important component of the VO that should eventually be supported at the level of datasets, calibration quantities, and individual data pixels.

Work in this WBS supports the Metadata Working Group in their definition of space-time metadata. From the data model point of view, it is important to ensure that the mechanisms used to associate the space-time metadata with a dataset are defined generically so that they can also be used with other kinds of metadata.

A collaboration between CACR, the Caltech Astronomy department, and the CDS Strasbourg has been using Topic Map technology to create tools that can federate metadata. We are leveraging the UCD (Uniform Content Descriptor) mechanism—which closely describes the semantic meaning of an astronomical datum—and the central role of UCD in the VOTable specification. Given that UCDs are already internationally accepted we can build these further semantic tools. Topic maps can be used to take a number of related astronomical tables and find the connections and commonalities between the attribute descriptors, so that effective federation and data mining can be assisted by machine.

3 Metadata Standards

3.1 Metadata Standards / Basic Profile Elements

The Space-Time metadata design (A. Rots) has progressed to a point where it is defined in terms of an XML DTD as well (and more usefully) as an XML Schema. Extensive discussions and experiments have led to various revisions. One final revision will be made before November 1, 2002. When that is done, we can concentrate on writing code to construct and interpret the Space-Time Coordinate objects, as well as to perform transformations. With the help of such tools the Space-Time Coordinate (STC) metadata can actually be used. The STC metadata project also has shown the path to a metadata generalization that will allow us to express other metadata following a similar design. As part of this work, we have contributed to the effort to define the proper place and use of Uniform Content Descriptors (UCDs).

There are a few issues left concerning the STC metadata. In particular, we will need to find a firm design for defining new coordinate frames such as, for instance, coordinate frames anchored to solar system objects. But these are not of immediate concern and we have ensured that the current design of the STC metadata allows such extensions. As a sub-issue in this area, we have provided a design for Spatial Region metadata. In the next year we will need to work on interfaces to this metadata design. Various experiments in the Metadata and Data Model groups have made us all realize the importance of such metadata.

R. Hanisch led the draft definition of Resource and Service Metadata (http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/ResourceServiceMetadataV5.pdf). An important result of this document has been its description of an architecture for understanding the role of resources and services in the VO. The architecture outlined by the Resource and Service Metadata document makes clear the need for an integrated approach to resource and service registration in which service descriptions "inherit" the metadata of resource that provides it. Such an approach will ultimately make registration easier for providers by minimizing the information they must provide as they register or extend more and more services.

3.2 Specific Profile Implementations

A white paper describing the relationship between existing metadata standards and the interactions between users and the VO was circulated. In the related NASA ITWG effort, preliminary WSDL profiles were written for services for several NASA archives and some simple Web services built on these profiles were prototyped.

While substantial work has been done in this area, the anticipated focus on specific metadata profiles in the early part of the VO development has been somewhat shifted to implementations of more generic metadata and transport protocols for the support of the VO demonstrations.

The relationship between this effort and the data models effort continues to be clarified. An image specification was nominally made in the data models area but was strongly influenced by the metadata discussion.

3.3 Metadata Representations and Encoding

The bulk of our work in this area has been oriented toward supporting the first year prototype demonstrations. The first major accomplishment in this area was the development of the VOTable XML definition, version 1.0, led by R. Williams (Caltech) and Francois Ochsenbein (CDS/AVO). Besides proving to be a critical component of the cone search and Simple Image Access interfaces, the VOTable demonstrated the process of developing standards through an open, international effort.

An important part of the Simple Image Access (SIA) specification is the handling of metadata used for locating and querying image servers. As part of the development of this specification, we identified the metadata required for the various forms of the service and matched them with existing CDS/ESO UCD tags. Where appropriate UCDs were not defined, we defined new UCDs within an experimental namespace (named VOX, for Virtual Observatory eXperimental). The specification enumerates which metadata are required and how they should be represented in the image query and the VOTable response. The specification also lists the metadata need for registering the service, which is a superset of the Resource and Service Metadata.

In addition to our short-term focus on the first year demos, we have put some effort into long-term metadata solutions. In particular, R. Plante has been assembling requirements and implementation ideas for a general metadata definition framework, resulting in a white paper (http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/fw-draft2.pdf). This framework will be further refined in collaboration with the Data Models Working Group.

Issues and Concerns: Since the release of version 1.0 of the VOTable specification, we have examined how the Space-Time Coordinate System metadata might be integrated into VOTable. We realized that this problem exemplified a more general need to associate detailed metadata information with one or more table columns. VOTable thus needs a hook for referencing arbitrary, external schemas so that new metadata can be easily inserted into the VOTable document.

The SIA specification is a prototype developed to support the first year demonstrations; thus, we expect to replace this specification. The approach used to develop the spec was to start by mirroring architecture of the prototype cone search specification, use existing VOTable capabilities and practices to express query result information, and use existing UCDs where ever possible. This uncovered various shortcomings of these technologies, and we departed from this approach accordingly.

With the coding version of the specification complete, we are now focusing on the caching of service metadata in registries. After the completion of the first year demos, efforts will shift to longer-term solutions; this includes:

- a comprehensive framework for registering data and services that minimizes redundant information and effort required of providers, and
- a general framework for defining metadata on which to base generic metadata software.

3.4 Profile Applications

In support of the first-year demonstrations, R. Williams, R. Hanisch, and A. Szalay developed a specification for a "Cone Search" interface for gathering information associated with circular regions on the sky from distributed catalogs. This specification has been implemented for over 50 data services to date (see http://skyserver.pha.jhu.edu/VoConeProfile/). Szalay has set up the registration service used to located compliant cone search services.

The Simple Image Access interface represents the image analog to the catalog cone search, however it harnesses a wider array of metadata. At the core is a rectangular region search. Since this interface can apply to cutout services as well as static image archives, additional data for describing and reporting precise spatial coverage is provided in both the query and response.

3.5 Metadata Standards / Relationships

No work scheduled during this period.

3.6 Metadata APIs

A number of libraries have been developed for getting information in and out of VOTables. Parsers are available in Perl (HEASARC), Java (CACR), and C++ (VO-India); a library for writing VOTables is available in Perl (NCSA). In addition, a VOTable-to-AIPS++-Table has also been developed.

The Simple Image Access (SIA) specification includes a mechanism, referred to as a metadata query, which allows implementing services to describe how they support image queries. In particular, they describe what input parameters they support and what columns they will return in the query result. While this functionality is accessible to end clients, it is primarily intended for use by the central registry service: when an implementing service registers itself, the central registry will send it a metadata query and cache the results. This will allow clients to use the registry to search for compliant services according the query-able parameters and the information they return.

Currently, the SIA metadata query mechanism does not return (in a standard, specified way) the resource and service metadata; that is, this information is only available via the registry. In the future, however, it is expected that querying the service directly should be the most authoritative way to get this information. Thus, a scheme must be worked out for dynamically gathering this information into registries for efficient access by clients.

It would be good to revise the cone search specification to adopt this metadata query framework. This would make it possible to better integrate cone search registry information with that of the SIA services. A registry for the SIA services is now being set up at JHU.

4 Systems Architecture

4.1 System Design

The system design for the NVO relies strongly upon the Grid technology that is being developed under the NSF NMI initiative and applied in the NSF Teragrid. The design has three main components: Web services support, data analysis support, and collection management support. The web services design has been primarily led by D. Tody and R. Williams. The data analysis support will be provided by the Globus toolkit. The collection management support is being provided by the Storage Resource Broker. Components of the NVO system include portals for accessing images, catalogs, and procedures; interactive web services; batch oriented survey processing pipelines; and grid services. While these components are oriented towards data and information management, a similar infrastructure is required for knowledge management that expresses the sets of operations that can be performed on a given data model, and defines the relationships between the UCDs that express exact semantics for physical quantities. The knowledge management tools are a current active area of discussion, with multiple options being considered.

The NVO system design document is primarily being driven by the technologies that are being used within the NSF Teragrid. The Teragrid will be the NVO testbed for both large-scale data manipulation and collection replication. Hence the NVO system design will closely follow that of the Teragrid. The data handling systems of the Teragrid are still being debated. Three environments are under consideration; persistent sky survey disk caches, high performance SAN-based data analysis disk caches, and deep archives. Versions of each of these environments either exist at SDSC, Caltech, or NCSA, or are being implemented. The specification of an architecture for the NVO will in part depend upon how the Teragrid decides to integrate these data management systems.

Similarly, the system design for the web services environment depends upon the competing standards from three communities: the Web Services Description Language environment being created by vendors, the Open Grid Services Architecture being developed by the Global Grid Forum, and the Semantic Web architecture being developed by the W3C community. There are efforts to merge the three environments. The challenges are the choices that will be made for authentication, for service discovery registries, and for service instantiation factories. The current Grid Forum Services architecture is not yet stable enough for production systems. We have the choice of going with WSDL based implementations, and then upgrading to the next generation technology, or waiting to see what the final architecture will look like. The services that are being implemented currently within the NVO are an important step, but they will require significant modification to interoperate with the Grid.

4.1.1 System-Level Requirements Definition. The system design for most of the NVO architecture is being driven by practical experience with test systems. Three categories of environments are in active test. They include services oriented towards processing a small amount of data (1000 records or 90 seconds access), data analysis pipelines that are scaled to process all the images acquired during one day of image collection, and large scale processing supported by the NVO testbed. The design of the testbed requires an engineering estimate of the computation capacity, I/O bandwidth, and caching capacity. To ascertain a reasonable scale of resources, we are continuing the implementation of a background analysis of the 2MASS collection, in collaboration with J. Good of IPAC. This will require a complete sweep through 10 TB of data, at an expected rate of 3 GB/sec. Good has created the initial pixel reprojection and background normalization routine, which is being applied at SDSC to the 2MASS data. The analysis is compute intensive, instead of data intensive. The complexity of the computation appears to be 9000 operations per pixel. The needed data access rates can be sustained from archives without use of high performance disk.

A second observation is related to the cut-out and mosaicing service that has been created by R. Williams for processing the DPOSS collection. Each DPOSS image is a gigabyte in size. The initial version of the service retrieved the entire image from the remote storage system, and applied the cut-out and mosaic generation locally. The time needed to generate the cut-out was dominated by the time needed to move a GB of data over the network. The analysis was implemented as a remote proxy in the Storage Resource

Broker by G. Kremenek. This eliminated the need to move the entire image. The cutout was generated directly on the remote storage system, and only the reduced image transferred to the user. The service then ran much faster.

A third observation is related to the replication of the DPOSS sky survey collection between Caltech and SDSC. The files were registered into a logical name space through execution of a script. The images were then replicated onto a second archive at a relatively slow rate limited by the network bandwidth. The management of data within the NVO testbed will need to rely heavily upon the use of logical name spaces rather than physical file names. The preservation of the NVO logical name space will be one of the major system design requirements. This will be a differentiating factor between the Grid and the NVO testbed. The Grid replica services are currently designed for short-term replication of data, rather than the long-term replication of entire collections.

These three experiments indicate the need to address latency management directly within the NVO system design. Fortunately, the mechanisms implemented in the Storage Resource Broker appear to be sufficient, namely data aggregation in containers for bulk data movement, remote proxies for I/O command aggregation, remote proxies for data sub-setting, replication for data caching, and bulk metadata manipulation. The latter is important for replicating collections onto compute resources.

Four TB of the 2MASS collection are replicated onto a disk cache at SDSC. We have completed the replication of the 2MASS collection onto the HPSS archive at Caltech. This is important to improve reliability by a factor of 10. When the HPSS archive at SDSC is off line, we are able to retrieve images from the Caltech copy. To support the automated replica fail over, we have installed version 1.1.8 of the SRB at Caltech.

We have done a test run of a re-analysis for the DPOSS collection, in collaboration with R. Williams of Caltech. This was done on a 64-processor Sun platform, accessing data from a disk cache. The computation was CPU-limited, taking 410 seconds to process a single 1-GB image on one processor. Using the entire platform, the re-analysis of the complete DPOSS collection could be done in 11 hours, at a sustained I/O rate of 135 MB/sec. This includes writing a new version of the entire collection back to disk, or moving 5.6 TB of data. The goal is to gain a factor of 10 in performance by moving to the Teraflops compute platform, and the large 30 TB disk cache.

We are also working on engineering estimates for the manipulation of large catalogs. J. Gray has shipped us a copy of the SDSS metadata (80 GB). We have dedicated disk space and compute resources to the analysis support requirements for this catalog.

4.1.2 Component Requirements, and 4.1.3 Interaction with Grid Components and Tools. E-mail exchanges were conducted on WSDL and OGSA interfaces to the grid environment, metadata management, data model specification, and knowledge management with E. Deelman, D. Tody, R. Williams, and R. Plante. SDSC is implementing a set of WSDL data management services for data discovery, data access, collection building, and data replication. The services are being integrated with Grid portal technology to support computations on shared data environments. We expect this approach to be a prototype for the NVO services that are integrated with Grid technology.

4.1.4 Logical Name Space. We upgraded the SRB server at Caltech to version 1.1.8, to support automatic fail over to an alternate replica. This will improve reliability of the system for image access by a factor of ten. This still requires testing the new version with the existing IPAC 2MASS portal.

4.2 Interface Definition

SOAP-based web services are becoming standard in the business community, and are expected to rapidly become the vehicle for sophisticated web applications like the Virtual Observatory. In Year 2 of the NVO project, we expect to start a transition to SOAP of many of the GET/POST based services that we have defined this year, such as the Cone Search and Simple Image Access Protocol. CACR has been creating simple SOAP Web Services from open-source Apache Tomcat and Axis software. This complements work at JHU, which is using the Microsoft framework for SOAP services. These alternate development paths are necessary to assure interoperability among various implementations. Also see WBS 3.4.

4.3 Network Requirements

Work not scheduled until Year 2.

4.4 Computational Requirements

Work not scheduled until Year 2.

4.5 Security Requirements

See WBS 6.2.

5 Data Access/Resource Layer

5.1 Resource and Information Discovery

Work has proceeded along several fronts in the area of resource and information discovery. Following the specification of the Cone Search service, we implemented a registration service that indexes these services. Fifty services are registered. This registry will be extended to include services supporting the Simple Image Access Protocol. Working in collaboration with CDS (Strasbourg), we assigned Uniform Content Descriptors (UCDs) to more than 1400 attributes in the Sloan Digital Sky Survey database. The SDSS database was amended to be compliant with the UCD physical units standards. In making the UCD associations, we noted several gaps in the UCD hierarchy that have since been added by the CDS. A template was developed that allows the mapping between UCDs and database relations to be incorporated directly into the archive, and thus to support queries based on UCDs. This, in turn, will allow the automatic creation of Topic Maps.

5.2 Data Access Mechanisms

5.2.1 Data Replication. In a wide area computing system, it may be desirable to create remote read-only copies (replicas) of data elements (files)—for example, to reduce access latency, increase robustness, or increase the probability that a file can be found associated with idle computing capacity. A system that includes such replicas requires a mechanism for locating them.

USC/ISI is developing a Replica Location Service, the next generation of the Globus Replica Catalog (RC). RC permitted a mapping from logical file names to the physical locations of the particular file. Although the functionality of RC in terms of the mapping was adequate the performance and the reliability of the system (a centralized server) was low. The new generation, the Replica Location Service, allows for the system to be distributed, and replicated. The RLS is extensible in that the users and applications can extend the information contained within it to other application specific attributes. The testing on the alpha prototype of the service is underway. As we progress in the development cycle, we will look forward to setting up a testing environment within the NVO framework. We are also in the process of integrating RLS into Chimera (see WBS 6.2).

The Replica Location Service is now in beta testing. During this period we are testing the functionality of the service as well as its performance. So far the results are encouraging in both areas however, further testing still need to be conducted.

5.2.2 Metadata Catalog Service. The Metadata Service provides a mechanism for storing and accessing metadata, which is information that describes data files or data items. The Metadata Service (MCS) allows users to query based on attributes of data rather than data names. In addition, the MCS provides management of logical collections of files and containers that consist of small files that are stored, moved and replicated together.
At this time, an initial design has been proposed and a Java API to access the catalog has been implemented.

Metadata Services require a high level of consistency. In the current design, we have implemented the service as a single centralized unit. Obviously this solution may not be scalable as the size of the metadata increases and the accesses to the catalog service increase. As a result, in the future, we may consider a more distributed architecture where we can have access to the information at various locations in the Grid, but still be able to rely on highly up-to-date information.

5.3 Data Access Protocols

Much of the work of the Metadata Working Group concentrated on a protocol by which image data could be published and retrieved—the so-called Simple Image Access Protocol. The word "image" in this context was restricted to sky-registered images—

images that have an actual or implied World Coordinate System (WCS) structure—and a single well-defined bandpass specification. However, the standard is capable of representing several publication paradigms:

- a collection of pointed observations,
- a collection of overlapping survey images covering a region,
- a uniform mosaic coverage of a region of the sky, and
- dynamically reprojected images with client-specified WCS parameters.

The Metadata Working Group has also discussed and defined XML data models for point, region, coordinate frame, bandpass, and other astronomical data objects.

5.4 Data Access Portals

The paper "Simple Image Retrieval: Interface Concepts and Issues" (July 2002) presented a conceptual design for implementing uniform image access via services supporting multiple access protocols. The document "Simple Image Access Prototype Specification" was released in late September following much discussion and several drafts.

Several implementations of simple image access were completed during interface development (by STScI, HEASARC, NOAO) and a number of others were in progress as of the end of the reporting period. A related image cutout service developed by Caltech and SDSC for DPOSS data uses scaleable grid services to provide access to massive all-sky survey data collections such as DPOSS.

The initial goal of simple image access was to support the NVO science demos while exploring the issues of providing uniform access to heterogeneous, distributed image data holdings. The simple image access service, along with the cone search service developed previously, provide early prototype data access portals. The simple image access interface has since drawn interest from our IVOA partners in Canada, Europe, and the UK, and future development will be a collaborative effort with these partners.

The next step will be to explore the use of web services for data access, and look into the issues of client access to such services. This will be done by demonstrating simple image services that simultaneously support both URL and WSDL/SOAP based access. A parallel effort is underway to develop a modular data model and metadata framework, which will be integrated with data access as it develops. The simple image access prototype already includes experimental data model components, e.g., for the image world coordinate system, and for characterizing the spectral bandpass of an image.

To keep simple image access "simple" and have it ready in time to support the science demos the SIA interface is based on simple URLs for requests, using FITS files to return science data. Future challenges will be to provide data access via a web services interface (WSDL/SOAP), and later, via grid-enabled interfaces such as OGSA or Condor-G. A concern is that the effort expended on implementing simple image services not be lost as we develop future, more sophisticated access protocols and services. A potential solution is to separate the access protocol from the service implementation.

This approach also has the advantage that a service can potentially support multiple simultaneous access protocols.

6 NVO Services

6.1 Computational Services

The work on computational services is proceeding on two broad fronts. The first is the development of compute and I/O-intensive services for deployment within the NVO architecture:

- Montage, an astronomical mosaic service funded by the Earth Sciences Technology Office Computing Technologies program. It will deliver science-grade mosaics where terrestrial background emission has been removed. Ultimately, Montage will run operationally on the Teragrid, and deliver on demand custom mosaics according to the user's specification of size, rotation, spatial sampling, coordinates and WCS projection.
- A general cross-matching engine funded by the National Partnership for Advanced Computing Infrastructure (NPACI) Digital Sky project; this service will have the flexibility to cross-match two tables in memory, or two database catalogs, and will have the option to return probabilistic measures of cross-identification of sources in the two tables.

A Software Engineering plan, Requirements Specification, Design Specification, and Test Plan have been completed for the Montage project. These documents are available on the project web site at http://montage.ipac.caltech.edu.

The design of Montage separates the functions of finding the images needed to generate a mosaic, reprojection, and transformation of images, background removal, and co-addition of images. Thus it is a toolkit whose functions can be controlled by executives or scripts to support many processing scenarios.

The heart of Montage is the reprojection algorithm. An input pixel will overlap several pixels in the output mosaic. We have developed a general algorithm that conserves energy and astrometric accuracy. It uses spherical trigonometry to determine the fractional overlap in the output mosaic pixels.

A fully functional prototype has been deployed for Solaris 2.8, Linux 6.x, and AIX ; it is available for download to parties willing to take part in validating the algorithms. The reprojection algorithm is slow—a single 2MASS image takes 4 minutes on a Sun Ultra 10 workstation. The algorithm can be easily parallelized, and we will use this approach to speed up the code. We have begun a collaboration with SDSC to parallelize Montage on the IBM Blue Horizon supercomputer. We have already run Montage on 64 nodes in parallel, where a 1 square degree area (55 2MASS images) can be run in under 3 minutes.

USC/ISI is also working closely with the IPAC team in the porting of Montage onto the Grid. ISI has also agreed to be an initial tester of the system. At present USC/ISI is learning about the structure of the Montage code with the hopes of using the Chimera system (WBS 6.2 to drive the execution of the Montage components. The main concern is the access to the data required for the Montage. Although we can use protocols such as GridFTP to access individual files, the data is currently stored in containers that can only be indexed by SRB. We are working on indexing the SRB containers so that the necessary data can be retrieved.

Montage will be used to deliver small image mosaics as part of the "Gamma Ray Transients" demonstration project.

The cross-match engine development has been largely geared towards the "brown dwarf demonstration project," which will cross-match the 2MASS and SDSS point-source catalogs. We have developed a design that is quite general and will support cross-matching between local files and database catalogs, and streaming from distributed catalogs. Our aim is, in fact, to stream the 2MASS and SDSS catalogs and cross-match them on the fly. Thus far, we have delivered code that will cross-match small tables that can be held in memory, and applies the probabilistic cross-match code used by the NASA Extragalactic Database to match sources. We are currently developing code that will handle database catalogs and streamed data.

6.2 Computational Resource Management

6.2.1 Computational Request and Planning. We are developing the Request Object Management Environment (ROME) to manage requests for compute- and time-intensive processing and data requests through existing portals; this middleware employs Enterprise technology already widely used in e-business. Most web services and portals employ Apache to manage requests. Apache is efficient and stable, but has no memory of requests submitted to it. Consequently, visitors to web services have no means of monitoring their jobs or resubmitting them, and the service itself has no means of load balancing requests. When large numbers of requests for time- and compute-intensive requests are submitted to NVO services, such functions are essential; without them, users will simply have to wait until job information returns. Users will not tolerate several days of waiting to learn that their job has failed.

ROME will rectify this state of affairs. It will deploy an Enterprise Java applications server, the commercial system BEA Web Logic, which accepts and persists time and compute intensive requests. It is based on e-business technology used, for example, by banks in managing transactions, but with one major change: ROME will have two components optimized for handling very time intensive requests. One, the Request Manager, will register requests in a database, and a second, the Process Manager, will perform load balancing by polling the database to find jobs that must be submitted and then send them for processing on a remote server.

We have delivered design and requirements documents, and have prototyped the following EJB components:

- UserRegistration—Creates user entry in the DBMS, user email address is used as user ID.
- UpdateUserInfo—A user contacts ROME to update the log-in information (e.g., machine name and port).
- RequestSubmission—Creates a request entry in the request DBMS table, and returns a request ID to user.
- UpdateRequest—Allows a user to send interrupt request to abort a job.
- GetStatus—A user fetches request status from DBMS.
- GetRequest—A processor thread asks ROME to search DBMS for the next request (of the specified application) in the queue.
- UpdateApplicationJobID—Once a processor thread started a job running successfully, it sends the job ID to ROME.
- SetMessage—A processor thread sends messages from the application to ROME.

Each of these components is a servlet/EJB pair. The servlet accepts an HTTP request from external entities (user/processor) and employs the corresponding EJB to write/retrieve the information to/from DBMS tables.

A Request Processor with multiple processing threads was built to process the requests. A simple dummy application program was used in the server to accept the request parameters and to send a sequence of "processing" messages to ROME (and on to the user).

This prototyping effort was aimed at understanding the challenges involved in using EJB technology under heavy load. We found that:

- An EJB container is very good at maintaining DBMS integrity. When two EJBs try to access a DBMS record simultaneously, the EJB container automatically deals with record locking and data rollback so that only one of the EJB instances will succeed in accessing the record, but it does not ensure that both updates are eventually processed successfully.

- When two processor threads contact ROME requesting the "next" job to process, ROME must ensure that the same request is not given to both of them.

We are also tracking technology that is being developed through the NSF GriPhyN project, the DOE Particle Physics Data Grid, the DOE SciDAC projects, and the NASA Information Power Grid, for the management of computational resources. The two central components are management of the computational resources, and management of the processes that are being run on the computational resources. The former is handled by the Globus toolkit, version 2. The latter is still a research activity. There are multiple versions of work flow management under development, including the Condor DAGman and associated data scheduling mechanisms, the survey pipeline processing systems used in astronomy, and an advanced knowledge-based processing system under development at SDSC for a DOE SciDAC project. We would expect to start with the current survey pipeline systems, switch processing to a grid managed environment under Condor when computer resources are exceeded, and then switch to the knowledge-based processing systems for complex queries. The advantage of the knowledge-based systems is their ability to dynamically adjust the workflow based upon results of complex queries to information collections. The conditional relationships between processing steps can be quite complex, as opposed to the simple semantic mapping of output files to input files for the DAGman system.

6.2.2 Authentication. USC/ISI has evaluated Spitfire, a database access service, which allows access to a variety of databases. Spitfire, developed as part of the European Data Grid project, consist of a server as well as client tools. The Spitfire server connects through JDBC to a database using predefined roles. The client can connect directly to the server through HTTP, and perform database operations. Even though Spitfire seems on the surface to be an interesting technology, it has many drawbacks in terms of security and support for transactions that span multiple database tables. For example, although Spitfire is based on the Globus Grid Security Infrastructure for Authentication, it exemplifies security problems in terms of authorization. In tests performed at USC/ISI, we are able to modify the database using a new version of the Spitfire server with an unauthorized client (a client from an earlier version of the code, which did not implement any security). Spitfire also does not currently support transactions that span multiple DB tables. The documentation was also inadequate, as it showed only examples of query operations and not example templates for create, update or delete operations. USC/ISI has communicated the authentication concerns to the Spitfire developers and is currently studying the possibility of adding transactional support to Spitfire. We are also following the developments within the UK e-Science program for any development in the area of grid-enabled interfaced to databases.

6.2.4 Virtual Data. USC/ISI is working with University of Chicago on a Virtual Data System, Chimera, which allows users to specify virtual data in terms of transformations and input data. The system is composed of a language and a database for storing the information needed to derive virtual data products. USC/ISI has focused on designing and implementing a planner which enables the translation between an abstract representation of the workflow necessary to produce the virtual data and the concrete steps needed to schedule the computation and data movement.

USC/ISI is currently working on the second version of the planner, which is part of the Chimera. The second version allows the planner to map the execution of the workflow onto a heterogeneous set of resources. Currently the planner is rudimentary and further research is needed to increase the level of sophistication of the planning algorithm as well as increase the level of the planner's fault tolerance. ISI is actively working with the AI planning community to increase the capabilities of the planner.

The Virtual Data System language (VDL), developed at University of Chicago is specified in both a textual and XML format. The textual version is intended for use in the manual creation of VDL definitions, for use in tutorial, discussion, and publication contexts. The XML version is intended for use in all machine-to-machine communication contexts, such as when VDL definitions will be automatically generated by application components for inclusion into a VDL definition database. The VDS-1 system, also known as Chimera, is implemented in Java, and currently uses a very simple XML text file format for the persistent storage of VDL definitions. Its virtual data language provides a simple and consistent mechanism for the specification of formal and actual parameters, and a convenient paradigm for the specification of input parameter files. VDS-1 has been released in the summer of 2002.

This planner takes an abstract Directed Acyclic Graph (DAG) specified by Chimera and builds a concrete DAG that can then be executed by Condor-G. In the abstract DAG neither the locations of where the computation is to take place nor the location of the data are specified. The planner consults the replica catalog to determine which data specified in the abstract DAG already exists and reduces the DAG to only the minimum number of required computations and data movements. Finally, the planner transforms the abstract DAG into a concrete DAG where the execution locations and the sources of the input data are specified. This DAG is then sent to Condor-G for execution.

7 Service/Data Provider Implementation and Integration

7.1 Service/Data Provider Implementation

Through the publication of the Cone Search and Simple Image Access Protocols, we have made it possible for service and data providers to begin to make information available through VO-compliant interfaces. Within the NVO project we brought some 50 Cone Search services on-line, and as the first project year came to a close several SIA services had already been implemented.

7.2 Service/Data Provider Integration (Hanisch/STScI)

Integration of Cone Search (VOTable) and Simple Image Access Protocol services is a challenge for the initial science demonstrations. As formal work in this area is not scheduled until 2003, the science demonstration teams are experimenting and developing prototypes that will, in time, migrate into additional integration tools and templates.

8 Portals and Workbenches

8.1 Data Location Services

Although formal activities in this area are not scheduled until later, the registration services for the cone search and simple image access protocols directly impinge on this area. Similar prototype efforts as part of the GRB demo project enable searching a hierarchy of surveys to find the “best” available survey image in a given wavelength regime.

8.2 Cross-Correlation Services

Work in this area is primarily funded by other resources. See WBS 6.1 for details.

8.3 Visualization Services

In anticipation of the need to be able to visualize correlations in complex data sets, such as joins between large catalogs, we have been evaluating several extant software packages that might serve as user front-ends. Foremost among these is *Partiview*, a package developed originally at NCSA and currently supported by the American Museum of National History/Hayden Planetarium. R. Hanisch and M. Voit met with program developers B. Abbott and C. Emmart to understand more about its capabilities. Partiview (particle viewer) was designed to render 3-D scenes for complex distributions of particles. It includes, for example, a full 3-D model of the Galaxy as a test data set. Our interest in Partiview is as a visualization tool for *n*-dimensional parameter sets, where one might plot an V magnitude on one axis, an x-ray magnitude on a second axis, and an infrared color index on a third axis. The ability to view such distributions from arbitrary angles, and to "fly" through and around the data, will be helpful in understanding correlations and in identifying unusual object classes. Partiview is freely available for Unix and Windows platforms.

We have also begun experimenting with *Mirage*, a 2-D plotting and data exploration tool developed by T. K. Ho (Bell Laboratories). Mirage provides a very flexible user interface and allows for rapid exploration of complex data. One can highlight objects in one 2-dimensional view, and instantly see the same objects in all other views. We expect to use Mirage as one of the visualization tools for the galaxy morphology science demonstration. Mirage is a Java application that installs easily on any Java-enabled platform.

Most recently, we have implemented some enhancements to the CDS Aladin visualization package, including the ability to overlay data directly from VOTables and to plot symbols in colors corresponding to an attribute in the VOTable. For example, objects in a catalog could be marked by position, and a third attribute such as spectral index or ellipticity could be encoded through the color of the plot symbol. Other encoding schemes (symbol size, vectors, etc.) will be explored in the future.

8.4 Theoretical Models

The inclusion of the US theoretical astrophysics community into the NVO framework continues to be a high priority item. In FY 2002 there were continued discussions among theorists interested in establishing a "Theory Virtual Observatory" (TVO) as a working prototype that could be incorporated into the US-NVO. These discussions focused primarily on the N-body codes being developed for simulation of the evolution of globular clusters, but discussions were also held with those groups working on N-body plus hydrodynamic codes, together with groups involved with MHD codes. The intent is to develop libraries of computationally derived datasets that can be directly compared with observations. In addition, there is interest in establishing sets of commonly shared subroutines and software tools for post-processing analysis. Throughout the fiscal year the "TVO Website," located at http://bima.astro.umd.edu/nemo/tvo has been maintained and updated.

Work that will lead to incorporation of theoretical astrophysics into the general NVO structure has been initiated in collaboration with J. McDowell (SAO). This effort will begin the definition of the metadata for simulation archives and will design the path needed to implement the publication and archiving of both theory datasets and theory software.

9 Test-Bed

9.1 Grid Infrastructure

We are engaged in initial experiments based on Grid services at USC/ISI, UCSD, SDSC, and NCSA, building upon the TeraGrid collaboration's infrastructure. See WBS 6.1 for details.

9.2 User Support

Work not scheduled in this area until 2003.

9.3 Software Profile

Work not scheduled in this area until 2003.

9.4 Data Archiving and Caching

Work not scheduled in this area until 2003.

9.5 Testbed Operations

Formal activity in this area not scheduled until 2003, though some use of the Grid testbed is planned for the early science demonstrations.

9.6 Resource Allocation

See WBS 6.2.1.

9.7 Authentication and Security

See WBS 6.2.2.

10 Science Prototypes

10.1 Definition of Essential Astronomical Services

We have defined a set of Core Services for astronomical web services. These include metadata services, basic catalog query functions, basic image access functions, survey footprint functions and functions for cross-identification. URL-based definitions of these functions have been developed in the Metadata Working Group. Based upon the above tentative specifications, JHU team members have built a prototype multi-layer Web Services application, called SkyQuery, which uses archive-level core services to perform basic functions, using the SOAP protocol. Proper WSDL descriptions have been written for these services, and the services have been successfully built for the SDSS, FIRST, and 2MASS. Templates for these Web Services have been used successfully by other groups (STScI, AstroGrid Edinburgh , Institute of Astronomy Cambridge).

JHU and STScI are in the process of creating a prototype footprint service that can be used to automatically determine overlap areas between several surveys. JHU and STScI have successfully built a simple SOAP-based web service interoperating between the .NET and Java platforms. JHU staff have built a web-services template to turn legacy C applications into Web Services. In collaboration with A. Moore (CMU), we have built several data-mining web services. We are currently building a C# class around the CFITSIO package, which will enable an easier handling of legacy FITS files within web services.

10.2 Definition of Representative Query Cases

In order to facilitate the functionality of the NVO and to test software developments, a clear need exists for implementation of representative query cases. In addition, the early demonstration of this capability to the US astronomical community will inform astronomers in general about the NVO and its ability to enhance scientific inquiry. Thus in FY 2002 the NVO Science Working Group (SWG) was given the task of developing an appropriate suite of scientific queries that would serve both to test the NVO structure and to demonstrate its capability. The membership of this Working Group is C. Alcock, A. Connolly, K. Cook, R. Dave, D. De Young (Chair), S. Djorgovski, G. Evrard, G. Fabbiano, J. Gray, R. Hanisch, L. Hernquist, P. Hut, B. Jannuzi, S. Kent, B. Madore, R. Nichol, M. Postman, D. Schade, M. Shara, A. Szalay, P. Teuben, and D. Weinberg. Through the process of many e-mail exchanges and telecons the SWG finally converged on a set of 13 well-defined scientific inquiries that would be appropriate to the NVO and would yield interesting and timely scientific results.

These 13 queries were than presented to an NVO Team meeting held in Tucson 16-17 April 2002. One of the major objectives of this Team Meeting was to converge on a set of three or four Science Demonstration Projects that could be developed in time for presentation at the AAS meeting in January 2003. Discussions at the Team Meeting thus focused not only on the scientific merits of the 13 inquiries but also their technical feasibility and their appropriateness to the NVO concept and architecture. At the end of the Tucson meeting, three of the 13 science queries had been chosen, and their technical requirements had been largely defined. These three science demonstrations are: 1) a brown dwarf candidate search project; 2) a gamma ray burst follow-up service; and 3) a galaxy cluster morphology and evolution survey project. The NVO Executive Committee held a number of meetings with the team members identified to lead each of these demonstrations, and progress in their development has been closely monitored.

10.3 Design, Definition, and Demonstration of Science Capabilities

Gamma-Ray Burst Follow-up Service Demonstration: The GRB demo comprises several distinct elements:

- Automated response to the discovery of a GRB: A request to include this demo in the Gamma-Ray Burst Coordinate (GCN) network has been submitted. This will inform the service of bursts within a few (typically < 2) seconds of initial GRB trigger in satellite flags. Occasional triggers are being received today, but these will become common with the launch of Swift. The GCN provides software to receive these reports and this software has been modified to initiate the retrieval of data.

- Querying and caching of results. Preliminary scripts for the querying and caching of results have been developed and tested. These are currently being changed to use the SIA and Cone search protocols for resources that support them. A more formalized caching mechanism needs to be developed.

- Initial notification page. A design for the initial notification of a burst was circulated and comments received. The actual implementation of this page is underway.

- Initiation of user interfaces. Neither of the user interfaces to be used in the demo, Aladin or OASIS, directly supports the VOTable format. Scripts for chopping data into appropriate pieces to start these programs have been developed but will need further work.

Galaxy Morphology Demonstration: This demonstration, which examines the relationship between galaxy morphology and cluster evolution, has been designed to illustrate some of the key functionality of VO infrastructure, including access to data through standard interfaces and grid-based analysis (http://bill.cacr.caltech.edu /cfdocs/usvo-pubs/files/morphdemo-plan2.txt). Development of this demo has progressed along several fronts:

1. Science goal development: With advisement from the Science Working Group, R. Plante, J. Annis, and D. De Young developed the overall plan for the demo.
2. Development of the Simple Image Access interface: With its development led by D. Tody and contributions from the Metadata Working Group, this interface will be used to access image data used by the demo.
3. Identification support of input data sets: E. Shaya and B. Thomas (NASA ADC/Raytheon) have implemented access to ADC catalog data via VOTable cone search interface. This service will provide various data about the target clusters.

The exact image data that will be used will depend on which of candidate datasets can be made available via the SIA interface. Candidates include the DSS survey, 2MASS, and the HST WFPIC2 data from the Canadian Data Center. X-ray data will come from the Chandra data archive through a specialized service that can return calculated fluxes; A. Rots and J. McDowell have implemented the basic service. Galaxy catalog data will come from either CNOC1 catalog from the CDC or the DSS catalog at NCSA.

4. Assembling the Grid-based data management and computing infrastructure: A special working group made up of J. Annis, E. Deelman, and R. Plante was formed to work on this. R. Plante has been testing the use of the mySRB tool for managing the data workspace where data for the demo can be collected. J. Annis and E. Deelman have been defining the technology components required to launch the grid-based analysis of the galaxy images.

11 Outreach and Education

11.1 Strategic Partnerships

NVO Outreach Workshop. The NVO project held an outreach workshop in Baltimore on July 11-12, 2002 that brought together a diverse group of education and outreach experts to identify critical features of NVO that would enable effective outreach. Twenty-six people attended, representing the NASA outreach community, the NSF ground-based astronomy community, museum professionals, amateur astronomers, planetarium builders, and developers of desktop planetarium software. The recommendations emerging from this meeting are setting the agenda for the development of the NVO outreach infrastructure.

Education and Outreach Requirements Document. The document *Enabling Outreach with NVO* collects and prioritizes the recommendations of the outreach community that were identified at the July workshop. The most critical need is for infrastructure development that will 1) lead non-astronomers who visit NVO to services and information that are most likely to be of interest to them, and 2) simplify the development of education and outreach resources by our partners. We will develop a metadata vocabulary for identifying and categorizing EPO services; work in this area has already been applied to the *Resource and Service Metadata* document.

Amateur Astronomy Image Archive. We have been working on a feasibility study of an Amateur Astronomers Deep Space Image Archive that would encourage amateurs to publish and request images using NVO protocols. This pilot is in collaboration with Sky and Telescope magazine.

11.2 Education Initiatives

No education initiatives were planned for this year.

11.3 Outreach and Press Activities

No outreach and press activities were planned for this year.

Activities by Organization

California Institute of Technology/Astronomy Department
S.G. Djorgovski, R. Brunner, and A. Mahabal participated in the discussions on the development of science demonstration cases. Work was also done in the following areas:

1. Preparation of the DPOSS data (one of the selected data sets) for the various VO uses and demonstration experiments. Image data reside at both CACR and SDSC. Catalog data are served via:

http://dposs.caltech.edu:8080/query.jsp

VOTable format is supported. A cone search service is under development. Most of this work was done by R. Brunner, with contributions from A. Mahabal.

2. Most of the effort supported by this grant was focused on the exploration of the Topic Maps technology for a VO. Most of the work was done by A. Mahabal. In collaboration with CACR and CDS Strasbourg, we have been using Topic Map technology to create tools that can federate metadata. We are using UCDs (Uniform Content Descriptors) in astronomical catalogs as PSIs (Published Subject Indexes) to relate columns from different tables to each other. The tool that we have built allows a user to choose a set of existing UCD-enabled catalogs and build a Topic Map out of the metadata of those tables. That Topic Map is then available for the community and can be used as a data discovery tool. Users can explore combinations of different catalogs to look for compatibility, overlap, cross-matches, and other scientifically enabling activities. As needed, for instance, users can generate and query different Topic Maps for X-ray, IR, and optical regions by combining metadata for those catalogs.

We have also started adding meaningful external links as part of the basic Topic Map. These include access to the catalogs, doing statistics on individual columns, plotting histograms etc. While these tools are neither part of Topic Maps nor necessarily developed by us, providing them in this fashion is a right step in semantically connecting the VO tools.

The main Topic Map page can be seen at:

http://www.astro.caltech.edu/~aam/science/topicmaps/ucd.html

The Topic Map generator is accessed by going to:

http://avyakta.caltech.edu:8080/topicmap/

California Institute of Technology/Center for Advanced Computational Research
CACR implemented the Simple Image Access Protocol for the 3 TB DPOSS sky survey, with cutouts generated dynamically from gigabyte-sized plate images. We have also begun the implementation for the Virtual Sky collection of co-registered multi-

wavelength surveys. We have been working with the NASA Extragalactic Database (NED) to create SIA services for their large and diverse image holdings.

Caltech is one of the four core sites of the NSF-funded Teragrid project (http://www.teragrid.org), which is designed to bring the scientific community towards the new paradigm of Grid computing. Much of the funding of this project goes to high-performance clusters of 64-bit Itanium processors, as well as large “datawulf” style disk storage systems. One of these systems will become part of the NVO testbed, with 15 terabytes allocated for storing large astronomical datasets such as DPOSS, 2MASS, and SDSS. These will be available online—with no delay as tapes are loaded—and under NVO access protocols. In this way, we hope to increase acceptance in the astronomical community of these protocols.

CACR is a collaborator in the NASA-funded Montage project for creating scientifically credible image mosaics from sky surveys such as 2MASS. Montage allows accurate image reprojection, thus creating federated multi-wavelength images. CACR, with SDSC and USC/ISI, is working on efficient parallel and grid implementations of Montage. (http://montage.ipac.caltech.edu)

In collaboration with SDSC, we have implemented a cutout service for the DPOSS archive that takes the data from the nearest of any number of replications of the archive. The SRB (Storage Resource Broker) that underlies the service provides this location transparency—DPOSS is currently at both SDSC and CACR. The SRB also provides protocol transparency, so that the archive can be stored with different mass-storage software (HPSS, Unitree, Sun-QFS, Posix, etc.). The image service responds to requests based on sky position, then finds and opens the relevant image file and extracts the desired pixels. Further processing of the cutout creates a valid FITS-World Coordinate System header (for sky registration) from the polynomial Digitized Sky Survey plate solution.

California Institute of Technology/Infrared Processing and Analysis Center
During the past year, IPAC personnel:

- Delivered 12 cone search services to the prototype NVO services registry.
- Began work on making IPAC services VOTable compliant.
- Assumed technical leadership for the “brown dwarf” demonstration project; delivered project work and technical description.
- Made substantial progress in deploying a general cross-match engine, to be used in the “brown dwarfs” project.
- Developed mature prototype of the Montage image mosaic service, and successfully ran it on 64 nodes of the IBM Blue Horizon supercomputer.
- Developed design and requirements for ROME; began prototyping efforts.

Canadian Astronomy Data Centre/Canadian Virtual Observatory Project
The Canadian Virtual Observatory (CVO) Prototype system has been developed and tested with WFPC2 catalogue content. Deployment of the CVO prototype has been delayed because of unacceptable database performance. Funding has been secured for a major upgrade in database hardware and software. Several months have been invested in identifying the most effective purchasing strategy. A new database system will be in place by March 31, 2003.

The Canadian Astronomy Data Centre (CADC) has committed to participation in the NVO demo project on Galaxy Morphology and will supply catalogues, cone search, WFPC2 image cutout service, and WPC2 image retrieval service for that demo.

The Canada-France-Hawaii Telescope Legacy Survey represents valuable content for the Virtual Observatory. CADC is designing and implementing the data processing and distribution system in collaboration with CFHT and TERAPIX. The initial goal is to effectively deliver archive services for this data and full integration into the Canadian Virtual Observatory (CVO) prototype system will be started.

Storage capacity at CADC will reach 40 Terabytes and a 40-node processing array will be deployed in early 2003. Hiring of 1.5 FTE of new staffing for CVO has been initiated.

Carnegie-Mellon University/University of Pittsburgh
The NVO NSF funding we received was used to support P. Husing, a programmer working with the Autonlab group at Carnegie Mellon. Husing was tasked with three NVO related problems that he succeeded in during this year. First, he has created simple and complete web documentation of our fast and efficient data-mining applications (see http://www.autonlab.org/astro/). These pages provide the code and examples of how to use the code and the various inputs and outputs. Second, he succeeded in making the EM Mixture Model code (see Connolly et al. 2000) command-line based as well as making the code much more modular in nature. This is vital for creating a web service out of this algorithm as well as providing users with the underlying kd-tree technology. Thirdly, he was successful in working with the JHU SDSS database group in interfacing our EM Mixture Model code with the SDSS SQL database. This was done in the Microsoft .NET architecture where an http request was sent to the server to a) extract some SDSS data from the DB, and then b) run our EM algorithm on that data and return as a web-page the parameters of the best fitted Gaussians. This implementation was very rudimentary and is not a true web-service as the data and algorithm reside on the same machine.

Husing has unfortunately left this project, but remains at CMU (as a CS PhD student) and therefore, we have not lost his experience and knowledge. In the coming year we plan to expand on this work in the following ways. First, we will make our other algorithms command-line stand-alone programs (so they can become web-services). Second, we will expand on the foundation made by Husing in adding more functionality to the EM Mixture Model within the .NET environment (at present, the code takes on one parameter input, so it is severely limited). Finally, we need to make our algorithms real web-services, i.e., they reside at CMU as a service that anyone can access through well-defined protocols. This will require us to learn about SOAP, JavaScript and the Web Service Description Language. Hopefully, we can find significant experience within the NVO to help us with these tasks and share resources and experience.

We continue to develop and refine our algorithms, but this effort is funded separately from the NVO efforts discussed above.

Fermi National Accelerator Laboratory
Fermilab has contributed in the areas of metadata and resource discovery (WBS 3 and 5), testbed (WBS 8) and science demos (WBS 10). It has hired one person, V. Sekhri, to work part-time on NVO and part-time on iVDGL (International Virtual Data Grid Laboratory) duties.

S. Kent is a member of the Metadata Working Group and has regularly participated in the weekly working group phone conferences.

J. Annis has worked on development of the Galaxy Morphology science demo (WBS 10.3.1) and on its implementation. The demo involves combining multiple datasets on galaxy clusters from the optical and X-ray bands, computing galaxy morphology parameters on the fly, and comparing them with optical and X-ray maps of the clusters. A set of five datasets with the necessary catalog and atlas image cutouts were identified for use in the demo. FNAL's unique contribution to this demo is the use of grid computing toolkits developed for the Griphyn/iVDGL projects to generate "virtual data" from existing raw data sources. J. Annis has developed the algorithm for computing the morphology of a galaxy and interfacing the code to Chimera, a facility in the grid computing environment that is used to track virtual data catalogs and launch applications. V. Sekhri has worked on the grid infrastructure, installing the Globus/Condor/Chimera toolkits on the Fermilab computing cluster and integrating a C++ VOTable parser into the Fermilab SDSS software framework. As part of the iVDGL project, Fermilab is operating a five-node analysis cluster.

Fermilab has had representatives at all three NVO collaboration meetings.

Johns Hopkins University
JHU staff participated primarily in activities related to WBS 5.1 and 10.1. Additional activities included:

- Implementing a Condor testbed to experiment with linking databases to the Computational Grid
- Experimenting with horizontal and vertical partitioning of databases, measuring the performance implications
- Experimenting with various fast cross-identification algorithms
- Collaborating with IPAC on the brown dwarf science demo and with STScI and FNAL on the galaxy morphology demo
- Overhauling and extending our HTM library to prepare for the footprint services and larger scale cross-identification in the NVO framework

- Working out an efficient way to perform spherical polygon intersections in SQL in collaboration with J. Gray
- Participating in LSST working groups to understand the requirements for NVO from the LSST perspective
- Working on various science projects that integrate archives (with millions of objects) and the data analysis software. These will help us understand the real science requirements arising from future applications.

Presentations by A. Szalay:

- Invited talk, Workshop on Computational Grids, LISHEP2002, Brazil, Feb 2002
- Astronomy Colloquium, Princeton, Feb 2002
- Invited talk, AMPATH conference, Chile Apr 2002
- Keynote Address, SIAM Workshop on Scientific Data Mining, Arlington, VA, Apr 2002
- Colloquium, NRAO, Soccoro, May 2002
- Invited talk, Hungarian Academy of Sciences, May 2002
- Presentation, Sigmod Conference, Madison,WI, June 2002
- Presentation about the SDSS, AAS Meeting, Albuquerque, June 2002
- Presentation about the NVO, AAS Meeting, Albuquerque, June 2002
- Invited talk, NVO conference, Garching, June 2002
- Invited talk, Sky Surveys workshop, Leiden, June 2002
- Plenary talk, Conference on Hyperbolic Geometry, Budapest, July 2002
- Presentation, Microsoft, Redmond, July 2002
- Seminar, Institute for Astronomy, Hawaii, Aug 2002
- Invited talk on Data analysis, SPIE conference, Hawaii, Aug 2002
- Invited talk on Data mining, SPIE conference, Hawaii, Aug 2002
- Invited talk on Web Services, SPIE conference, Hawaii, Aug 2002
- 2 lectures at the D. Chalonge School of Astroparticle Physics, Palermo, Sep 2002
- Seminar, Observatoire Strasbourg, Sep 2002
- Talk, NASA AISRP conference, Mountain View CA, Oct 2002

Microsoft Research

J. Gray worked closely with the Sloan Digital Sky Survey on building an online version of the First Data Release (DR1, SkyServer.sdss.org). In addition he worked with R. Mann and others at the Royal Observatory-Edinburgh to get the Synoptic Sky Survey online. These are all preparatory to having online archives that can contribute to the virtual observatory. Gray also worked with T. Malik and others at JHU on SkyQuery (SkyQuery.org), which is a step beyond a cone search: it allows general queries across an SDSS+First+2MASS federation of web services. In particular, Gray wrote an image renderer that rectifies and decorates SDSS images attribute data (it is part of SkyQuery.org).

National Optical Astronomy Observatories

NOAO was involved in many aspects of the NVO project this year through the efforts of D. Tody (who relocated to NRAO at the end of the year) and D. De Young. D. Tody was active in the Metadata and Systems Working Groups, and co-led the development of the

Simple Image Access Protocol (WBS 5.4). D. De Young led the Science Working Group and carries out the responsibilities of Project Scientist. In addition, De Young has been coordinating the development of requirements for the NVO for theoretical astrophysics.

In addition, M. Fitzpatrick began work on a prototype VOTable browser in support of the first year science demos as well as legacy software (IRAF) integration with NVO. Thus far this work has focused on design and initial development of the browser, plus related extensions to the IRAF environment including extensions to the GUI toolkit (new widgets) and development of a preliminary XML parser for use with IRAF. An early implementation was demonstrated at the 2002 ADASS conference, with delivery of a baseline version expected early in Q4 for possible use in the NVO science demos. R. Plante and T. McGlynn expressed interest in helping test the VOTable browser for possible use in the galaxy morphology and GRB demos.

P. Warner and M. Fitzpatrick implemented a Simple Image Access image cutout service for the NOAO Science Archive. New image data will be made available via this service as it is ingested into the archive. R. Allsman made improvements to the USNO cone search service previously released. G. Chisholm attended the NVO Team meeting in Urbana, July 29-30.

National Radio Astronomy Observatory
NRAO participated in the Metadata Working Group.

Work at NRAO in 2002 was limited by the lack of a person dedicated to our NVO funded work. Fortunately, Doug Tody will be joining NRAO in November 2002 to fill this position and to lead NRAO's NVO work.

Even without a dedicated NVO person, we have been able to make progress on making connections between AIPS++ and web resources (including services).

- We now have a connection between Glish (the scripting language used in AIPS++) and any web page. This is being deployed in our prototype archive server.
- We have a working SOAP server that connects to AIPS++-based data sources. This is being packaged for wider use within the NVO.
- We are also working to enable connections from Glish to and from Java, with the goal of eventually providing a simple way of accessing and providing web services.

This work is heavily leveraged by the use of AIPS++ in the radio astronomy community. Any such facility added to AIPS++ automatically becomes available for a wide range of different data sources.

Complementing this work, under a different NSF grant, we have developed and deployed a prototype radio astronomy archive server using the above Glish-Web page connection.

Doing this required development of a first pass at meta-data for radio astronomy data sets: specifically meta-data for any data set stored in an AIPS++ MeasurementSet.

Raytheon/NASA Astronomical Data Center
Raytheon/ADC staff contributed to the follow work areas:

WBS 2 Data Models. Raytheon/ADC staff wrote a separate white paper on a Data Model for space systems data and telemetry (tXDF). Staff also provided analysis and commentary on various NVO data model documents. Staff provided expertise to the project team in various XML developments, including the eXtensible Data Format (XDF) and the scientific Article XML Markup Language (AXML).

WBS 3 Metadata Standards. Raytheon/ADC staff participated most significantly in the Metadata Standards (WBS 3) activities, which included: (a) participation in weekly telecons of the NVO Metadata Working Group; (b) development of science use cases and the science demos; (c) analysis of various NVO metadata framework documents and resource description documents; (d) research into Resource Description standards (DAML+OIL) to see what applications are useful for NVO; (e) worked on improvements of VOTable and its API; (f) made improvements to the proposed NVO Space-Time schema and validation of its DTD; (g) reviewed and provided comments on the simple image access prototype; (h) investigated applications of Web Services (UDDI, WSDL, SOAP) to NVO prototypes; and (i) participation in the VOSemantics discussion group. Raytheon/ADC staff member K. Borne attended the first team meeting (Caltech, December 2001) where many of these metadata working group (and data model) goals were first discussed and established.

WBS 7 Service/Data Provider Implementation and Integration. Raytheon/ADC staff worked to define the ADC metadata and to set up a prototype WSDL service demo. By the end of the reporting period, staff had developed a WSDL/SOAP service demo for astronomical data catalogs and worked on a cone search service for such catalogs. Staff supported the "cluster-galaxy morphology" demo by designing a port to NVO for astronomical catalogs that does coordinate search and provides data.

WBS 10 Science Prototypes. Raytheon/ADC staff members K. Borne and E. Shaya attended the Project Team meetings in Tucson (April 2002) and at NCSA (July 2002), with travel expenses covered by this grant. At those meetings and subsequently, Raytheon/ADC staff contributed to the development of two of the science demos: the Gamma-Ray Burst Follow-up and the Galaxy Morphology Demo.

WBS 11 Outreach and Education. Raytheon/ADC staff participated in the E/PO workshop at STScI in July 2002, and gave a presentation on "Data Mining with the NVO" as a data-intensive NVO exploration tool for students and amateur astronomers. Staff also developed a public web site for "NVO Information Resources" at http://nvo.gsfc.nasa.gov/nvo-index.html.

NASA decided to terminate the operations of the Astronomical Data Center at the end of FY 2002. Raytheon technical staff will continue to support the NVO project in their areas of expertise, however.

San Diego Supercomputer Center

SDSC is supporting formation of an initial NVO testbed that includes replication of collections onto environments where large-scale re-analysis operations can be performed. This includes working with collections to demonstrate replication, collaborating on re-analysis tasks to understand computation requirements, and collaborating on the NVO system architecture design. SDSC has implemented the following NVO resources:

- 2MASS sky survey archive in HPSS at SDSC
- 2MASS sky survey replica in the HPSS archive at Caltech
- Disk cache of 4 TB of the 2MASS collection to support the SC 2002 demos
- Data grid linking SDSC and Caltech based on the Storage Resource Broker. The data grid has been upgraded to version 1.1.8. This version supports automatic fail-over to alternate data sources when a repository is unavailable. The result is a service that appears substantially more reliable, and that provides interactive response when the data is on a disk cache. This style of data grid is an important system design requirement since users expect interactive response through the web.
- DPOSS sky survey replication into HPSS at SDSC
- DPOSS sky survey replication onto a high performance SAN at SDSC to support the SC 2002 demos
- Cutout service for DPOSS based on the SRB remote proxy capability to provide interactive response
- Analysis testbed for re-projection and background normalization of the 2MASS sky survey. The services will be demonstrated at SC 2002.
- SDSS catalog replication onto DB2 database technology. This effort is continuing to work on building correct representations of the twenty-one "canonical" queries that were originally implemented.
- Initiated replication of the USNO-B proper motion catalog onto resources at SDSC. A 120-GB disk has been shipped for retrieval of a copy of the catalog. We will replicate a copy of the USNO-B images onto HPSS by shipping data to SDSC on disks.
- Initiated a persistent disk collection cache for storing the USNO-B images. The cache is being assembled out of grid bricks, at a cost of $3,500 per TB. Ten grid bricks will be purchased for a total capacity of 10 TB. The grid bricks will be managed with the Storage Resource Broker to provide a uniform name space, manage user access, manage the file system, and manage user authentication. A grid brick has been ordered from Datel, but has not been received because of the scarcity of 160-GB drives. As soon as the system is received and tested, the remaining grid bricks will be bought.

  Analyses of the performance of grid bricks have been done both locally and through expert consultants on the design of disk caches. The choice of IDE drives to minimize cost is subject to some risk. IDE drives do not provide the same degree of head movement control as done by more expensive SCSI disk drives. IDE drives have only a single head synchronization track on each side of a recording track, instead of the two head synchronization tracks provided on SCSI disk. This means IDE drives are more susceptible to vibration. Fortunately, for read-only environments, such as data caches for sky surveys, the vibration risk is much lower, and the low-cost drives may turn out to be an acceptable solution for permanent caching of sky surveys. Performance testing at SDSC indicates data retrieval rates in excess of 100 MB/sec per grid brick.

- The Storage Resource Broker has been provided to NCSA to support creation of NVO data sharing environments, and will be used for demonstrations at SC 2002.
- The NVO testbed will use the Teragrid resources at SDSC. A 500-TB disk cache is being installed over the next 6 months to support large-scale data analyses. A high performance tape system has already been acquired. Data transfer rates of over 800 MB/sec have been demonstrated to the tape drives. We will experiment with data transfers from the archive to the SAN disk cache for the ability of the system to support data intensive analyses.

Smithsonian Astrophysical Observatory
The primary activities at SAO included:

- Leading the Data Model design (WBS 2.1, 2.2) and the Metadata design (WBS 3.1) efforts.
- NVO concepts were discussed with colleagues at CFA and several archives and sample data sets were analyzed (Optical, Radio, X-ray) to identify their unique and common characteristics.
- The SAO team monitored various design and implementation proposals related to DM, Metadata, and Grid. Sent comments on VOTable version 0.99; reviewed Web Services, SOAP, and UDDI.

Meetings:
- G. Fabbiano, A. Rots, and I. Evans attended the Pasadena kick-off meeting and contributed to, respectively, the science panel, the metadata/data model panel, and the testbed panel.
- J. McDowell and A. Rots attended the Opticon Archive Interoperability meeting in Strasbourg, in January, and had additional discussion with the staff at CDS.
- G. Fabbiano, J. McDowell, and A. Rots attended the Tucson team meeting in April.
- G. Fabbiano participated in the AVO meeting at ESO in June and the AVO Science Working Group meeting.
- A. Rots and J. McDowell attended team meeting in Champaign, IL, on July 30, 31.
- A. Rots and G. Fabbiano attended SPIE meeting in Waikoloa, HI, August 25.
- J. McDowell hosted a meeting at CfA with M. Louys from CDS/Strasbourg to work on issues of common interest: bandpasses, data quality, and image models.
- J. McDowell and A. Rots participated in the weekly Metadata Working Group telecons and the biweekly project telecons. Both are heavily involved in almost all

discussions that concern Metadata and Data Models, as well as in various aspects of coordinating and managing the project as a whole.
- J. McDowell and A. Rots took an active part in the discussions that led to the design of two prototype interfaces (the Cone Search and the Simple Image Access Protocol) and the design of the VOTable XML standard.

Talks/Workshops/Demos:

- G. Fabbiano, J. McDowell, and A. Rots presented talks at the June VO meeting in Garching; I. Evans et al. presented a poster paper.
- A. Rots and J. McDowell worked on the Chandra interface for the Galaxy Morphology demo.
- J. McDowell planned a VO Data Model Technical Meeting to be held at SAO, October 11/12.

Space Telescope Science Institute
R. Hanisch, the Project Manager, has organized team meetings, Executive Committee telecons, and has been the primary interface to the International Virtual Observatory Alliance (IVOA). Project management work has included development of the project management plan, the general project schedule, the project budget, and quarterly and annual reports.

Technical Activities:

Cone search and image access web services for GSC-I and II, DSS-I and II were developed, tested and entered into the NVO public registry. With these catalog services, profiles were published although they will be updated to meet with standard metadata resource and service specifications.

As part of the Metadata Working Group and the ongoing development of NVO services at STScI, staff actively participated in the development of VOTable XML standardized document exchange, regularly attended the telecoms, and also attended the team meeting at NCSA in Champaign-Urbana, Illinois.

Coordination between STScI technical and scientific staff and the JHU SDSS Science Archive team was initiated for the purpose of technical exchange on NVO technologies currently in development. Several areas have been identified where key software and architecture components will be beneficial to both sites and assist in NVO science demonstrations.

STScI has expanded the computer infrastructure to support .NET technology. This has provided a development environment where high performance features from integrated database and web service technology are used for prototype NVO discovery services. With this capability, the HST pointing catalog and associated exposure fields for WFPC2, FOC, and STIS have been deployed in a SQL Server database with NVO cone search and VOTable XML formatted data retrieval. This service and database is currently a testbed for cross-correlation between the SDSS science archive and sparse HST pointings using the Hierarchical Triangulated Mesh (HTM) spatial partitioning scheme. This is helping to identify fundamental requirements for NVO interoperability between archives by providing rapid sky coverage information and metadata discovery.

STScI staff are actively developing a C# VOTable parser for public distribution to the NVO resource/service participants in coordination with the JHU SDSS Science Archive group.

University of Illinois-Urbana/Champaign/National Center for Supercomputer Applications

R. Plante has been involved in the following NVO activities:

- Moderates the weekly Metadata Working Group telecon meetings. Early in the project, this group carried out much of the science scenario development by assembling "questions to the VO." Since then, this forum has been used to discuss and refine the various specifications developed to support the first year prototypes, including the VOTable XML format, the Cone Search interface, the Simple Image Access (SIA) interface, and metadata for resource and service descriptions.
- Collaborated with D. Tody on the development of the SIA specification, focusing primarily on the metadata specification and the metadata query mechanism. Plante also worked on an early test implementation.
- Developed a prototype Perl library for writing VOTables. Developed a VOTable to AIPS++ Table converter.
- Leading the Galaxy Morphology Demo (see Sect. 10.3): with J. Annis and D. De Young, developed demonstration plan, prototyped use of SRB in demo, coordinated the availability of target data through standard interfaces.
- Researching a long-term metadata definition framework; proposed approach in white paper (http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/fw-draft2.pdf).

University of Pennsylvania

The Penn group has concentrated on issues related to time-domain astronomy: temporal variability and proper motion. Discovery and examination in this domain will often require the comparison of images drawn from different surveys, with the attendant opportunities for discovery and risk for false detection. One obvious example of this type of activity is the combined use of the MACHO and OGLE III image data to search for moving objects: the baseline is now over a decade, but inter-comparison is presently not possible.

The Penn group has developed a preliminary "straw-man" design of a new standard for the incorporation of time-series data into a federated database system. It appears that an extension of FITS (with NVO compliant metadata) will be the most readily accepted proposal. (Alternative, non-FITS models may also be successful, but have been explored less extensively.) We have started work on a proposal to the FITS community.

We have concentrated on the issues of data provenance, especially in federated time series databases which may be assembled in semi-automated data pipelines. This is a particular concern with moving objects, which may be detectable only with analyses of multiple image datasets. Automated federation of heterogeneous image databases for the discovery of time-variable or moving objects has great potential for the discovery of new phenomena, but ensuring uniform, or at least well-understood selection criteria is not straightforward. These issues will have consequences for metadata standards for use in the time domain that are not yet understood.

The Penn group has designed a modular analysis pipeline for use with time series of images.

University of Southern California/Information Sciences Institute
USC/ISI has been active in several areas of technology evaluation:

- Virtual Data Systems (Chimera)
- Request Planning and Execution
- Data replica location services
- Authenticated remote database access
- Metadata Catalog Service

Also, in collaboration with the iVDGL project, we are evaluating tools for testbed set up and management.

Technical Report:
Ann Chervenak, Ewa Deelman, Carl Kesselman, Laura Pearlman, Gurmeet Singh, "A Metadata Catalog Service for Data Intensive Applications", GriPhyN 2002-11, http://www.griphyn.org/documents/document_server/uploaded_documents/doc--275--mcs.pdf

United States Naval Observatory
Work has progressed in making the USNO-FS (Flagstaff Station) image and catalog server compatible with the NVO interchange protocols. Specifically, the next release version of the server will have the capability to return data in the VOTable XML format, and should be able to handle automated queries in a standardized format. The USNO-B catalog (work done outside of the NVO aegis) was finished, and as a part of the NVO cooperative, arrangements have been made to supply a copy to SDSC as well as to several of the other major data centers (CDS, CADC) for general use. One of the first uses will be as part of the brown dwarf search science demonstration being done with 2MASS and SDSS data. Arrangements are in the works to have the USNO-FS image archive replicated at SDSC as well.

USNO team members participated in the April and July NVO team meetings.

University of Wisconsin
We assisted JHU in setting up a Condor pool. Since then, our primary interactions with the NVO community have mainly concerned bulk data transfers. We are continuing to

enhance our software to address what we believe are the future needs of NVO pipelines. We have not yet deployed real clients using this technology.

University Space Research Association/NASA High Energy Astrophysics Science Archive Research Center

Activities focused on three areas, metadata and data models, access services and the VO demos.

In the metadata area, a catalog metadata profile was released and extensive comments were made on a variety of proposals including the draft of the VOTable document, the image data model, the space-time coordinates document and discussion of data registries.

Both of the simple data access services: the Cone Search service and the Image Access Service were implemented. Considerable effort was made in reviewing these protocols, especially the later.

The HEASARC assumed leadership of the gamma-ray burst demonstration. Several iterations of documentation describing the demo were presented to the collaboration. A crude demonstration of what the demonstration would do was presented at a team meeting and this is the current primary focus of activities for the HEASARC.

USRA/HEASARC has implemented the cone search protocol on top of its standard Browse catalog interface. While all catalogs are currently queryable using the cone search interface only a few of these catalogs have been registered. It is unclear whether it is appropriate to register each of the hundreds of available catalogs.

USRA/HEASARC has implemented the Simple Image Access protocol on top of the SkyView Virtual Telescope. This provides access to approximately 20 large-scale survey datasets.

The USRA PI, T. McGlynn, has moved to a government position. S. Drake has taken over as USRA PI, but Dr. McGlynn is still intimately involved with this effort.

Related Tasks. The HEASARC is involved in two major VO related tasks. The ClassX collaboration is building an automated classifier for X-ray objects. This is intended as a VO prototype to get practical experience in some of the issues that the VO is wrestling with. As part of this task a complete Perl VOTable library was written and has been used, e.g., in the HEASARC's simple image access service. The HEASARC is also active in a collaboration of NASA archives working to improve interoperability of the NASA archive system. This group has explored WSDL and other promising technologies for describing their on-line astronomical services.

Publications

Berriman, G., Kong, M, & Good, J. C. "OASIS: A Data Fusion System Optimized for Access to Distributed Archives," AAS Meeting 200, 60.01, 2002

Borne, K. D., & Cheung, C. Y., "Science Data Mining Resources for the National Virtual Observatory (NVO)," AAS Meeting 199, 10.05 (2002)
http://nvo.gsfc.nasa.gov/nvo_datamining.html

Borne, K. D., "A National Virtual Observatory (NVO) Science Case: Properties of Very Luminous IR Galaxies (VLIRGs)," AAS Meeting 200, 97.10 (2002).

Brunner, R. J., Djorgovski, S. G., Prince, T. A., & Szalay, A. S., "The National Virtual Observatory," ASP Conf. Proc. 254, 383, 2002

Brunner, R.J., Djorgovski, S. G., Prince, T. A., & Szalay, A. S., "Massive Data Sets in Astronomy," Handbook of Massive Data Sets, J. Abello, P. Pardalos, & M Resende, eds., Kluwer Academic Publishers, 931, 2002

Davis, L. E., "Tools for Automated Astrometric and Photometric Calibration of Survey Data," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

De Young, D., "Theoretical Astrophysics and the US-NVO Initiative," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Djorgovski, G., "Data Mining Challenges in a Virtual Observatory," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Djorgovski, S. G., "The National Virtual Observatory Science Definition Team: Report and Status," AAS Meeting 200, 87.01, 2002

Evans, I., DePonte Evans, J., Cresitello-Dittmar, M., Germain, G., Glotfelty, K., Plummer, D., & Zografou, P., "Chandra Pipelines—Meeting the Science and Engineering Challenges," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Fabbiano, G., "CfA VO Activities and Data Model Testbed," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Genova, F., Benvenuti, P., De Young, D. S., Hanisch, R. J., Lawrence, A., Linde, T., Quinn, P. J., Szalay, A. S., Walton, N. A., & Williams, R. D., "International Collaboration for the Virtual Observatory," AAS Meeting 200, 87.03, 2002

Gray, J., “Computer Science Aspects of the Virtual Observatory—How Can We Federate All These Archives?” Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Gray, J., Chong, W., Barclay, T., Szalay, A.S., & Vandenberg, J., “TeraScale SneakerNet: Using Inexpensive Disks for Backup, Archiving, and Data Exchange,” Microsoft Technical Report, MS-TR-2002-54, 2002

Gray, J., & Szalay, A., “The World Wide Telescope: An Archetype for Online Science,” MSR TR 2002-75, pp. 4, June 2002

Gray, J., Szalay, A.S., Thakar, A.R. Stoughton, C., & Vandenberg, J., “Scientific Data Publication Pragmatics,” Proc. SPIE Conference on Advanced Telescope Technologies, in press, 2002

Gray, J., Slutz, D., Szalay, A., Thakar, A., Kuntz, P., & Stoughton, C., “Data Mining the SDSS SkyServer Database,” MSR TR 2002-1, pp.1-40, 2002

Hanisch, R. J., “Building the Infrastructure for the National Virtual Observatory: An Information Technology Research Initiative of the National Science Foundation,” Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Hanisch, R. J., “Building the Framework of the National Virtual Observatory: Status Report,” Proc. SPIE, 4846, in press, 2002

Hawkins, I., & Mattei, J., “Education and Public Outreach Summary from the NVO Science Definition Team Report,” AAS Meeting 200, 60.08, 2002

Mahabal, A. A., McDermott, J. P., Babu, G. J., Feigelson, E. D., Djorgovski, S. G., Nichol, R., & Wasserman, L., “Implementing Astrostatistics in the Virtual Observatory,” AAS Meeting 200, 60.05, 2002

Mahabal, A. A., Djorgovski, S. G., Brunner, R. J., & Williams, R. E., “Topic Maps as a Virtual Observatory Tool,” Proc. SPIE, 4477, 161, 2001 (astro-ph/0110184)

Mahabal, A. A., Djorgovski, S. G., Brunner, R. J., & Williams, R. E., “Topic Maps for Custom Viewing of Data,” Proc. SPIE, 4486, in press, 2002 (astro-ph/02110404)

McDowell, J., “Towards an Image Data Model for the VO,” Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

McGlynn, T. A., Angellini, L., Corcoran, M., Drake, S., Pence, W., Winter, E., Hanisch, R., Suchkov, A., White, R., Postman, M., Donahue, M., Genova, F., Ochsenbein, F., Fernique, P., & Wenger, M., “Classifying the High Energy Universe with ClassX,” Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

McGlynn, T. A., Suchkov, A., Angellini, L., Corcoran, M., Drake, S., Pence, W., Winter, E., Hanisch, R., White, R., Postman, M., Donahue, M., Genova, F., Ochsenbein, F., Fernique, P., & Wenger, M., "Classifying X-Ray Sources Using Multi-Wavelength Data," AAS Meeting 200, 60.03, 2002

McGlynn, T., Borne, K., Thomas, B., Eichhorn, G., Accomazzi, A., Rots, A., Berriman, B., Good, J., Kimball, T., & Mazzarella, J., "Linking NASA's Astronomy Data Archives," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Moore, R. W., "Data Grids for Collection Federation," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Moore, R., & Baru, C., "Virtualization Services for Data Grids," for publication as part of a collection on "The State of the Grid 2002"

Ochsenbein, F., Williams, R., Davenhall, C., Durand, D., Fernique, P., Giaretta, D., Hanisch, R., McGlynn, T., Szalay, A., & Wicenec, A., "VOTable—A Proposed XML Format for Astronomical Tables," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Plante, R., "Scalable Metadata Definition Frameworks," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Rots, A., "Space-Time Metadata for the VO," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Schade, D., "The Canadian Virtual Observatory Online: A VO Prototype System at the Canadian Astronomy Data Centre," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Shaya, E., Borne, K., Thomas, B., & Cheung, C. Y., "Publishing Scientific Articles in XML," AAS Meeting 199, 10.09 (2002) http://rings.gsfc.nasa.gov/~borne/axml-poster-mar2002.html

Szalay, A., "Analyzing Large Data Sets in Astrophysics (Distributed Computing and the Virtual Observatory)", Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Szalay, A., & Williams, R., "The NST ITR Project: Framework for the National Virtual Observatory," AAS Meeting 200, 87.02, 2002

Szalay, A.S., Budavari, T., Connolly, A.J., Gray, J., Matsubara, T., Pope, A. & Szapudi, I, "Spatial Clustering of Galaxies in Large Datasets," Proc. SPIE Conference on Advanced Telescope Technologies, in press, 2002

Szalay, A.S., Budavari, T., Malik, T. Gray, J., & Thakar, A., "Web Services for the Virtual Observatory," Proc. SPIE Conference on Advanced Telescope Technologies, in press, 2002

Szalay, A., Gray, J., Thakar, A., Kuntz, P., Malik, T., Raddick, J., Stoughton, C., Vandenberg, J., "The SDSS SkyServer—Public Access to the Sloan Digital Sky Server Data," ACM SIGMOD 2002, MSR TR 2001 104

Szalay, A.S., Gray, J., Vandenberg, J., "Petabyte Scale Data Mining: Dream or Reality?" Proc. SPIE Conference on Advanced Telescope Technologies, in press, 2002

Szalay, A.S., & Matsubara, T., "Analyzing Large Datasets in Astronomy," Statistical Challenges in Modern Astronomy III, E. D. Feigelson & G. J. Babu, eds., New York:Springer-Verlag, in press, 2002

Tody, D., Davis, L., Fitzpatrick, M., & Valdes, F., "A Prototype VO Data Access Portal," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Voit, G. M., "Enabling Outreach with Virtual Observatories," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Williams, R., "The Uphill Battle of Semantic Interoperability," Towards an International Virtual Observatory, Garching, Germany, June 10-14, 2002

Williams, R. D., "Grids and the Virtual Observatory", for publication as part of a collection on "The State of the Grid 2002"

NVO Publications Repository
http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/list.cfm

Date	Title	Uploaded by (email)	Category	Author(s)	**Abstract/Description** (may be truncated but is recorded in full)
2002/10/11	Resource and Service Metadata for the Virtual Observatory	hanisch@stsci.edu	Technical Reports	Hanisch, Robert, et al.	Version 5.
2002/10/07	Simple Image Access Prototype Specification	dtody@nrao.edu	Technical Reports	Doug Tody, Ray Plante, et. al.	This document specifies an interface for simple image access via a simple URL-based interface. The SIA prototype provides a uniform interface for access to a wide range of image services, including
2002/10/07	Resource and Service Metadata for the Virtual Observatory	hanisch@stsci.edu	Technical Reports	Hanisch, Robert, et al.	This is an update to V3.0 of this document.
2002/10/07	Simple Image Retrieval: Interface Concepts and Issues	dtody@nrao.edu	Technical Reports	Doug Tody	Conceptual design for simple image access.
2002/09/30	Quarterly Report, Apr-Jun 2002	hanisch@stsci.edu	Periodic Reports	Hanisch, Robert	Quarterly Report for Third Quarter, FY 2002.
2002/09/24	NVO Movie (condensed)	hanisch@stsci.edu	Presentations	STScI Audiovisual Lab	This is a condensed, compressed version the NVO "movie" that was produced by the STScI Audiovisual Lab as an aide to explaining the NVO concept. It has no sound, and was intended to allow a speaker t
2002/09/13	NVO Poster SC2002	hanisch@stsci.edu	Presentations	Hanisch, Robert and Szalay, Alex	This is a poster presentation intended for display at SC2002

					and the January AAS meeting.
2002/08/23	Astronomy Data Online	voit@stsci.edu	Education & Outreach Reports	Raddick, J.	This is a PPT presentation given at the NVO EPO Workshop, July 11-12, 2002, on how SDSS is enabling the public to access and understand their survey data.
2002/08/23	SegNVO Project	voit@stsci.edu	Education & Outreach Reports	Hawkins, I., Spitz, R., Craig, N.	This PPT presentation on assessing the NVO-related needs of the arts community and science museums community was given at the July 11-12, 2002 NVO Outreach Workshop in Baltimore.
2002/08/23	Astronomy Data Online	voit@stsci.edu	Education & Outreach Reports	Marschall, L.	This is a PPT presentation given at the NVO EPO Workshop, July 11-12, 2002, on how project CLEA's approach to developing a virtual observatory for undergraduate education.
2002/08/22	Metadata Working Group Minutes 2002-08-15	rplante@ncsa.uiuc.edu	Meeting & Telecon Minutes	Steve Kent	Agenda: 1. Roll call & News 2. Simple Image Access Document 3. Review of Action Items
2002/08/22	Metadata Working Group Minutes 2002-08-08	rplante@ncsa.uiuc.edu	Meeting & Telecon Minutes	Good, John and Plante, Ray	Agenda: 1. Roll call & News 2. Simple Image Access Document 3. Review of Action Items
2002/08/22	Metadata Working Group Minutes 2002-08-22	rplante@ncsa.uiuc.edu	Meeting & Telecon Minutes	McDowell, Jonathan and Plante, Ray	Agenda 1. Roll call & News 2. Simple Image Access Document * update on spec * implementation experiences 3. Cone Search Service Extension Proposal: VOTable inputs

					(Tom) 4. Review of A
2002/08/13	Producer-Archive Interface Methodology Abstract Standard	tam@lheapop.gsfc.nasa.gov	Standards	CCSDS	The object of this document is to regulate the relationships and interactions between an information Producer and an Archive. It defines the methodology to allow all the actions to be structured the
2002/08/05	Demonstration Project for the NVO: A Search For Brown Dwarfs	gbb@ipac.caltech.edu	Technical Reports	Bruce Berriman, John Good, Alex Szalay, and Davy Kirkpatrick	Scientific goals, technical description of project, and deliverables
2002/08/02	Minutes, 30-31 July 2002 Team Meeting	hanisch@stsci.edu	Meeting & Telecon Minutes	Hanisch, Robert	Minutes of the 30-31 July 2002 Team Meeting, held in Urbana, Illinois. The minutes include copies of presentations given at the meeting as well as notes on the discussions.
2002/07/22	Space-Time Coordinate Specification for VO Metadata	hanisch@stsci.edu	Technical Reports	Rots, Arnold	The following table attempts to explain the various elements in the Space-Time DTD and their use in each of the five areas of application: 1. Resource description (especially the coverage portion)
2002/07/19	Lessons Learned from Data-Rich Science Education Projects	voit@stsci.edu	Education & Outreach Reports	Pompea, Stephen & Isbell, Doug	This PPT presentation on developing data-rich astronomy education projects was given at the July 11-12, 2002 NVO Outreach Workshop in Baltimore.
2002/07/19	Enabling Outreach with Virtual Observatories	voit@stsci.edu	Education & Outreach Reports	Voit, Mark	This PPT presentation was given at the June 2002 International VO meeting in Garching and the July 11-12, 2002 NVO

					Outreach Workshop in Baltimore
2002/07/19	Lessons Learned from Integrating Experimental Technologies into Education	voit@stsci.edu	Education & Outreach Reports	Thakkar, Umesh	This PPT presentation on integrating experimental technologies into education was given at the July 11-12, 2002 NVO Outreach Workshop in Baltimore.
2002/07/19	The International Virtual Observatory Alliance: A Mission and Roadmap Statement 2002-2005	hanisch@stsci.edu	International VO Alliance	Quinn, Peter, and Hanisch, Robert	The International Virtual Observatory Alliance to facilitate the international coordination and collaboration necessary for the development and deployment of all tools, systems and organizational stru
2002/07/19	The National Virtual Observatory and Hands-On Universe	voit@stsci.edu	Education & Outreach Reports	Pennypacker, Carl	This PPT presentation on integrating NVO content into the Hands-On Universe education program was given at the July 11-12, 2002 NVO Outreach Workshop in Baltimore.
2002/07/18	Relationship of Galaxy Morphology to the Intra-cluster Medium: an NVO Demonstration	rplante@ncsa.uiuc.edu	Science Reports	Plante, Ray and Annis, Jim	The NVO is developing three demonstrations of VO principles as part of its first year efforts. This document describes the "Galaxy Morphology" demo, summarizing the science and technical goals.
2002/07/18	Resource and Service Metadata for the Virtual Observatory	hanisch@stsci.edu	Technical Reports	Hanisch, Robert and NVO Metadata Working Group	An essential capability of the Virtual Observatory is a means for describing what data and computational facilities are available where, and once identified, how to use them. The data

					themselves have
2002/07/18	NVO and the Digital Planetarium	hanisch@stsci.edu	Education & Outreach Reports	Lantz, Ed	This is a PPT presentation given at the NVO EPO Workshop, July 11-12, 2002, on the topic of how NVO might support digital planetariums.
2002/07/18	Metadata Working Group Minutes 2002-07-18	hanisch@stsci.edu	Meeting & Telecon Minutes	Hanisch, Robert	Agenda: 1. Roll call & News * DPOSS cutout service * Image service specification 2. Generalized Queries We may have finished this last week, but we'll see if there are any
2002/07/18	Topic Maps as a Virtual Observatory Tool	aam@astro.caltech.edu	Science Reports	Mahabal, A., Djorgovski, S.G., Brunner, R., Williams, R.D.	We describe Topic Maps and a specific example, and how they can be implemented within the VO Framework.
2002/07/18	Data Models for the VO, Part II: Metadata Objects for the VO	jcm@cfa.harvard.edu	Technical Reports	Jonathan McDowell et al	In this document, we identify some specific metadata quantities that may need to be modeled as objects.
2002/07/18	Data Models for the VO	jcm@cfa.harvard.edu	Technical Reports	Jonathan McDowell et al	This `white paper' describes the proposed VO data model. First, we describe the general architecture for data models. Then, we give examples of image and table data models. The data model will prov
2002/07/18	GRB Follow-up Datasets and Sources	tam@lheapop.gsfc.nasa.gov	Science Reports	Tom McGlynn	Describes the interfaces and datasets to be presented to the user for the GRB follow-up demo including the sources for these.
2002/07/17	EPO Workshop Presentation.	hanisch@stsci.edu	Presentations	Hanisch, Robert	This presentation gives an overview of

	11 July 2002				the NVO initiative and the NSF-funded NVO infrastructure project. This was presented at the 11-12 July 2002 NVO EPO workshop held at JHU/STScI.
2002/07/17	A Scalable Metadata Framework for the Virtual Observatory	rplante@ncsa.uiuc.edu	Technical Reports	Plante, Raymond L.	This white paper presents a framework for defining and evolving metadata for use across the different contexts of the VO in a community-based way. This paper was presented at the 2002 Garching confere
2002/07/16	Annual Report 2002	hanisch@stsci.edu	Periodic Reports	Hanisch, Robert	This is the summary document for our first Annual Report, submitted to NSF on 16 July 2002. The full report also includes copies of the first three Quarterly Reports, a spending and budget analysis,
2002/07/15	Initial Science Demonstration Schedules	hanisch@stsci.edu	Management Documents & Schedules	Hanisch, Robert; McGlynn, Thomas; Plante, Ray; & Berriman, Bruce	Schedules for initial science demonstrations (GRB follow-up, brown dwarf candidate search, galaxy morphology). (PDF format)
2002/07/15	Master Schedule	hanisch@stsci.edu	Management Documents & Schedules	Hanisch, Robert	Version 2 of the project master schedule. (File is in MSProject format.)
2002/07/15	Quarterly Report, Oct-Dec 2001	hanisch@stsci.edu	Periodic Reports	Hanisch, Robert	Quarterly Report, October-December 2001
2002/07/15	Quarterly Report, Jan-Mar 2002	hanisch@stsci.edu	Periodic Reports	Hanisch, Robert	Quarterly Report for January-March 2002.
2002/07/12	Virtual Observatories and the Grid	roy@cacr.caltech.edu	Science Reports	Williams, Roy	We consider several projects from astronomy that benefit from the Grid

					paradigm and associated technology, many of which involve either massive datasets or the federation of multiple datasets. We cove
2002/07/12	Data Grids for Collection Federation	moore@sdsc.edu	Presentations	Reagan W. Moore	Presentation given at the AVO meeting in Garching, Germany in June, 2002, on the use of data grid technology to federate multiple data collections.
2002/07/12	Toward a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan	roy@cacr.caltech.edu	Science Reports		White paper outlining the vision of the National Virtual Observatory
2002/07/12	VOTable: A Proposed XML Format for Astronomical Tables	roy@cacr.caltech.edu	Standards	Williams, R., Ochsenbein, F., Davenhall, C., Durand, D., Fernique, P., Giaretta, D., Hanisch, R., McGlynn, T., Szalay, A., Wicenec, A.	The VOTable format is a proposed XML standard for representing a table. In this context, a table is an unordered set of rows, each of a uniform format, as specified in the table metadata. Each row is
2002/07/12	Building the Framework for the National Virtual Observatory	roy@cacr.caltech.edu	Management Documents & Schedules	Szalay, Williams, Hanisch et. al.	Project overview for the NSF-funded collaboration. a

Acronyms

AAS	American Astronomical Society
ADC	Astronomical Data Center
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)
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
EPO	Education and Public Outreach
ESTO	Earth Science Technology Office (NASA)
ESTO-CT	ESTO Computational Technologies (NASA)
FIRST	Faint Images of the Radio Sky at Twenty Centimeters
FITS	Flexible Image Transport System
FNAL	Fermi National Accelerator Laboratory
FTP	File Transport Protocol
FY	fiscal year
GB	gigabyte
GLU	Générateur de Liens Uniformes (uniform link generator)
GRB	Gamma Ray Burst
GriPhyN	Grid Physics Network
HEASARC	High Energy Astrophysics Science Archive Center
HTTP	HyperText Transport Protocol
IPAC	Infrared Processing and Analysis Center (Caltech)
IRAF	Image Reduction and Analysis Facility (NOAO)
IRSA	Infrared Science Archive (IPAC)
ISI	Information Sciences Institute (USC)
ITWG	Information Technology Working Group (NASA data centers)
iVDGL	International Virtual Data Grid Laboratory
IVOA	International Virtual Observatory Alliance
JDBC	Java Data Base Connectivity (Sun, Inc., trademark)
JHU	The Johns Hopkins University
MAST	Multimission Archive at Space Telescope (STScI)
MB	megabyte

MOU	Memorandum of Understanding
MWG	Metadata Working Group
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputer Applications
NOAO	National Optical Astronomy Observatories
NPACI	National Partnership for Advanced Computational Infrastructure
NRAO	National Radio Astronomy Observatory
NSF	National Science Foundation
NVO	National Virtual Observatory
OASIS	On-line Archive Science Information Services (IRSA)
OGSA	Open Grid Services Architecture
OIL	Ontology Inference Layer
PB	petabyte
Q	quarter
QSO	Quasi-Stellar Object
RC	Replica Catalog
RDF	Resource Description Framework
RLS	Replica Location Service
ROME	Request Object Management Environment
SAO	Smithsonian Astrophysical Observatory
SAWG	Science Archives Working Group (NASA)
SAWG	System Architecture Working Group (this project)
SciDAC	Scientific Discovery through Advanced Computing (DOE)
SDSC	San Diego Supercomputer Center
SDSS	Sloan Digital Sky Survey
SDT	Science Definition Team
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
WBS	Work Breakdown Structure
WSDL	Web Services Description Language
XML	Extensible Mark-up Language
2MASS	Two-Micron All Sky Survey