

**NVO Team Meeting
30-31 July 2002
UIUC/NCSA
Champaign-Urbana, IL**

Minutes

News and Logistics

Attending: Ray Plante, Markus Dolensky, Tom McGlynn, Greg Chisholm, Jonathan McDowell, Kirk Borne, Stephen Levine, David Schade, Ed Shaya, John Good, Doug Tody, Ani Thakar, Vijay Sekhri, Steve Kent, Robert Brunner, Roy Gal, Reagan Moore, Gretchen Greene, Niall Gaffney, Arnold Rots, Bob Hanisch, Jim Annis, Dave De Young. By telecon: Bruce Berriman, Barry Madore, Ewa Deelman, Mark Voit. Guests: Dick Crutcher, Stuart Levy, Lindsey Davis.

Bob reported on status of second year renewal with NSF. Thanks to all for getting inputs in to Quarterly Report, which formed the basis of the renewal package. Indications from NSF are positive.

International collaboration is solidifying, with creation of International Virtual Observatory Alliance at the Garching VO Conference in June. Bob is chairing the IVOA initially, with Peter Quinn as co-chair.

The Garching conference was successful, with much more substantive talks. Critics were present but not all that critical. Peebles and Ostriker gave supportive presentations.

Forthcoming meetings include:

- SPIE
- ADASS XII
- Supercomputing 2002
- IAU Joint Discussion

Goals for this Meeting

- Review progress on science demonstrations, resolve open issues, consider contingency plans
- Review technical progress
 - Data models
 - Metadata standards
 - System architecture
- Recap EPO activities and first EPO workshop
- Review team working relationships—what should we change/improve based on first 8 months of experience?

Technical Status Reports

Metadata Working Group Activities (Ray Plante)

- Weekly telecons continue
- Priority given to topics related to first year demos

Space-Time Coordinate Metadata

- Led by Arnold Rots
- DTDs for Coordinates and Regions
- MWG discussing how to integrate into metadata into various metadata contexts:

- Registration, searching, **VOTable response**
- Goal: proposal for change to VOTable (as necessary)
- Speaks to more general VOTable question:
 - how can we add extra metadata descriptions of groups of columns?
 - Generalized "Frames"; e.g. Filter descriptions
 - Column errors
 - Expected approach:
 - Supplementary DTDs that specialize in describing specialized objects
 - Minimize changes to VOTable; just need a "hook" for inserting supplementary tags

Resource and Service Metadata

- Strawman proposal for general description of resources & services
 - Led by Hanisch
 - Includes working definition of what a resource and service is in the architecture of the NVO
 - Presented at Garching Interoperability WG meeting
- Use of a registry in first year demos
 - Important part of NVO: discovering what's there
 - Plante has worked out opportunities for discovery in Galaxy Morphology demo.
 - Total amount of information desired might be extensive;
 - how do we make it easy to register?

Relationship to query standard: what can I query on?

- Will present something do-able this month.

Image Services

- Needed by first year prototypes (particularly Galaxy Morphology demo)
- Architecture of image service standards
 - Led by Doug Tody; white paper available
 - Components:
 - Registration
 - Image Query
 - Simple Image Retrieval (synchronous, single-image)
 - Advanced Image Retrieval (asynchronous, multiple images)
- First-year version of Simple Image Access
 - Analogous to Cone Search Spec, but based on general architecture
 - Draft by Plante and Tody in progress.
 - Need implementations by Sept. 1.
- Cut-out services
 - New one by Roy Williams: DPOSS, uses SRB at SDSC
 - Several existing services

Data Models

- Two white papers by McDowell et al.
- Two data model meetings planned: August and October
- Global discussion

Miscellaneous

- Web Services
 - Several experiments underway within project
- OGSA -- Open Grid Services Architecture
 - A Globus initiative to define and implement Grid services in terms of Web Services

Discussion:

- Terminology: “data model”, “metadata”, “access methods”, “applications” – what do we mean specifically by these terms? We need to agree on definitions and relationships, need to document these definitions and relationships, and provides examples. **Action to MWG to provide such documentation.** [Following the meeting, Ed Shaya provided the following:

“...here are the definitions according to ISO/IEC SC32

<http://metadata-stds.org/Document-library/Draft-standards/11179-Part1-Framework/11179-Part-1.doc>

3.23 data model: A description of the organization of data in a manner that reflects an information structure.

3.41 metadata: Data that defines and describes other data.

That is, our work on metadata is description of data that is used to describe metadata. This description, a metametadata if you will, need not include any organizational info (order, hierarchy, containment, etc) because the data model is where that arises. One may choose to do both at the same time, and it often makes sense to, but the data model should also contain generic descriptions of how metadata is held in general.”

- Resource and service metadata. Bob wants to get back to this.
- Registry to support initial demos:
 - Ray will present something do-able this month (early August)
 - Reagan has software to do this
 - Discussed role of the service registry
 - For discovery purposes only?
 - For complete specification of how to use the service?
- Image services
 - Doug has specified the general architecture
 - Doug and Ray working on specification for URL-based interface
- Data models
 - Mirelle Louys (CDS) visiting CfA in August

System Architecture Working Group (Reagan Moore)

NVO Requirements

- Capability to off-load large computational jobs onto NVO testbed
 - Request management to differentiate between interactive jobs and batch jobs
- Ability to evolve infrastructure to encompass large-scale computations
- Mechanism to publish and share results
- Knowledge management mechanisms
 - Physical quantity equivalent forms
 - VOTable structural relationships
 - UCD semantic relationships

Brown Dwarf Search Science Demo

- Standard query language for interacting with entire catalogs.
- Information repository abstraction for characterizing interactions with catalogs in databases

- Standard information interchange format based on VOTable.
- Web based portal to control input parameter selection.
- Standard access mechanism based upon SOAP protocol and WSDL services.
- Service registry for publishing service descriptions, and supporting a service directory
- Method registry for methods that can be applied against data models.
- Standard display mechanism for results.
- Procedural support for composing results from multiple catalogs.
- Interactive response time for browsing tables and columns in databases.
- Modular design to enable rapid prototyping.
- Integrate existing systems and tools
- Persistent interactive access to image archives

Galaxy Morphology Science Demo

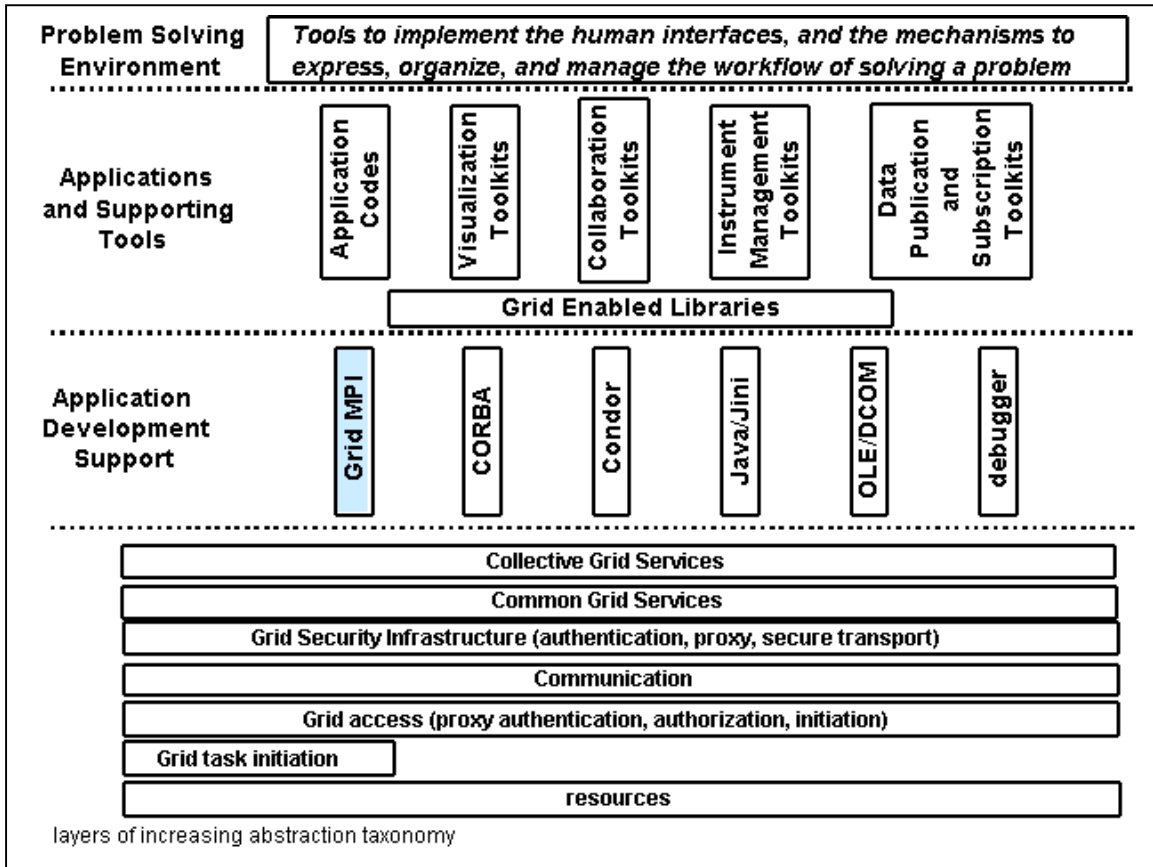
- Distributed joins across multiple catalogs
- Catalog Cone Search and positional cross-matching of extended sources
- Image retrieval from image archives, including partial images or cutouts
- Support for analyses of images, including galaxy morphology analysis
- Support for visualization services including contouring.
- Grid computing interface to support high-throughput analyses
- Mechanism for formatting data into a standard catalog format – VOTable
- Support for publication of new tables
- Service registry for catalogs
- Portal for defining input parameters, specifying processing chain, interacting with catalogs and image archives, interacting with grid computing, and formatting results.

Gamma Ray Burst Science Demo

- Cone search across multiple catalogs
- Standard information exchange format based on VOTable
- Support for mapping from current format to standard format
- Support for exception handling.
- Support for registration of interest in desired events.
- Support for tracking astronomical events.
- Support for distribution of notifications to registered users.
- Support for publishing results
- Support for interactive display of results, including parsing of VOTable format.
- Mechanism to specify relationships within a complex VOTable as a concept space

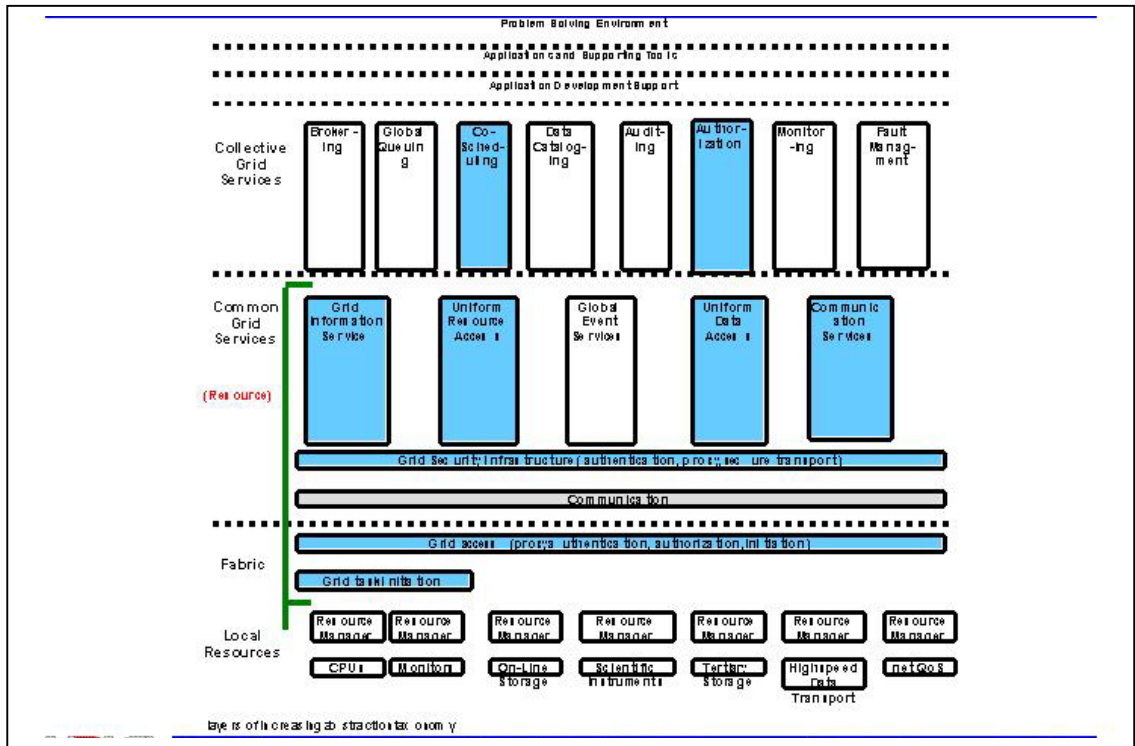
System Components

- Portals – the user interfaces that manage interactions with the multiple NVO services. Portals encapsulate domain knowledge by associating explicit mediators or transformation programs with access to the chosen catalogs and archives. Portals also provide interfaces to control image processing pipelines.
- Analysis and support tools – the NVO supplied applications for processing image and catalog records. These tools provide capabilities such as conversion to a standard data format, visualization mechanisms for 2D and 3D displays, metadata extraction mechanisms for populating attributes in catalogs, and algorithms for processing data.
- Application support platforms – these typically provide an environment for interactive execution support. Examples include Java execution, Corba object management, DCOM integration, WSDL/SOAP.
- Grid support environment – the set of collective operations used to build a common data management and job execution environment across administration domains. Grid services manage distributed state information.



Selected System Components

- Portal
 - SkyServer
- Analysis and support tools
 - Mosaic
- Application support platforms
 - Data access layer
- Grid Environments
 - Globus
 - SRB



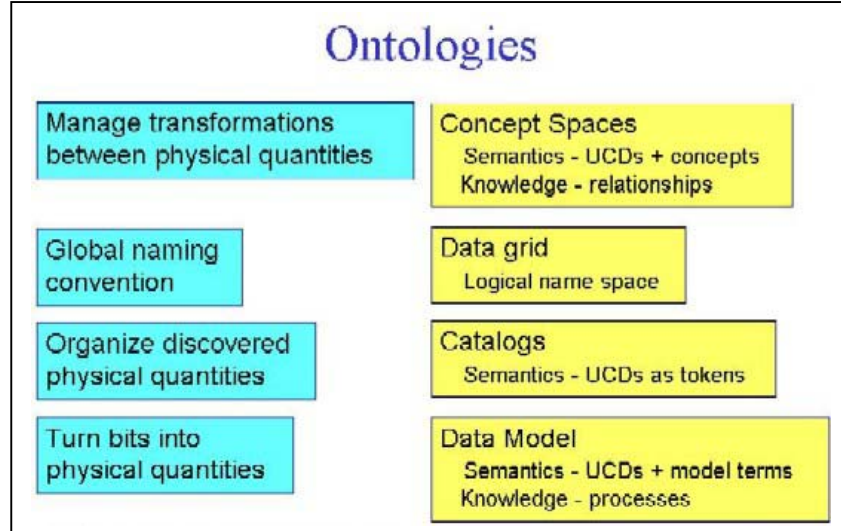
NVO Grid Requirements

- Support for interactive processes
 - Cutouts on large scale collections
- Transition from interactive to batch processing
 - Job management – Rome
- Knowledge management within the process workflow
 - Data model methods to transform physical quantities
- Multiple authentication levels
 - Anonymous / Shibboleth - group / GSI - individual

NVO Test Bed

- Survey portals
- Process management systems
- Data access layer for data model methods
- Data grid for logical name space
- Interactive cache for image archives
- Computational grid

Ontologies



Discussion:

- Reagan did a demonstration of the Storage Resource Broker (SRB). It worked! See <http://srb.npaci.edu>. MYSRB could be a vehicle for ingest of legacy data sets. Can easily browse and select attributes in a data collection.
- Web-based services are okay for managing ~100 datasets/calls. We'll want to push large comp. jobs onto comp. testbed. Looking for feedback on assessment of demo requirements.
- Testbed -- need a way to replicate data to where more resources can be harnessed.

Complementary projects

- ROME (John Good)
 - Focus is on managing requests and providing a view of request and processing status to the user
- Montage image mosaic service (John Good)
 - All components now prototyped successfully
- NASA ADEC interoperability efforts (Tom McGlynn)
 - Interoperability Technical Working Group working toward making all NASA archives interoperable
 - NASA data centers already have many interlinked services, but all are custom implementations
 - By object position
 - By observation ID
 - By bibcode
 - Discussing building service registries and gaining experience with WSDL; see <http://heasarc.gsfc.nasa.gov/itwg/sampleForm.html>
 - Working with journals to implement uniform data set identifiers in the literature, e.g., through new tags in AASTeX
- ClassX project (Tom McGlynn)
 - Project is now about half way done. Idea is to prototype VO functions in the context of a science application (identification of x-ray sources). See <http://heasarc.gsfc.nasa.gov/classx/>.

International Projects

AVO and AstroGrid (Markus Dolensky)

Work Area 0 – Programme Management (P. Quinn, ESO)

- Enabling a Virtual Observatory for Europe EVOE
- Submitted expression of interest for the 6th framework programme for integrated projects.
- Purpose: R&D alliance between all major astronomical data centres in Europe.
- EIROforum Grid Group EGG (chair P. Quinn)
- 1st Meeting of working group on March 22, 2002
- Members:
CERN, EBI/EMBL, EFDA/JET, ESA, ESO, ESRF, ILL
- Gridstart: Meeting this week (G. Rixon)

International Virtual Observatory Alliance IVOA

- Was formed at meeting on June 13, 2002
- 2 preparatory telecons on March 6 + May 2 about mission statement and road map
- Members:
 - AstroGrid–Australian Virtual Observatory eAA
 - Astrophysical Virtual Observatory AVO
 - Canadian Virtual Observatory CVO
 - German Astrophysical Virtual Observatory GAVO
 - Virtual Observatory India VOI
 - National Virtual Observatory NVO
 - Russian Virtual Observatory RVO

Recruitment Status

	<i>Job Title</i>	<i>Name</i>	<i>Institution</i>
AVO-1	Senior System Engineer	Markus Dolensky	ESO
AVO-2	System Engineer	Marco Leoni	ESO
AVO-3	AVO Scientist	TBD	ST-ECF
AVO-4	DB Tech. Softw. Devel.	TBD	Univ. Edin.
AVO-5	GRID Tech. Softw. Devel.	TBD	Univ. Edin.
AVO-6	Software Engineer	Marc Allen	CDS
AVO-7	Software Engineer	Thomas Boch	CDS
AVO-8	Software Engineer	TBD	Terapix
AVO-9	Software Engineer	Anita Richards	JBO

Work Area 1 – Science (P. Benvenuti, ST-ECF)

- Science Working Group Meeting on June 12, 2002 in Garching
- Formed subgroup to define use cases for demonstration in Jan. 2003:
N. Walton (chair), R. de Grijs, D. Egret, M. Kontizas, P. Ortiz, T. Prusti
- Two use cases are currently discussed:
 1. Cosmology Case: GOODS
 2. Nearby ‘Galactic’ Case: The Magellanic Clouds
- Tools envisaged to support both cases:
 - Spectral Energy Distribution fitting tool (SED)
 - Colour selection tool

Science Cases for 01/2003

1. GOODS (Great Observatories Origins Deep Survey)
 - Tools for SED extraction and analysis
 - Select objects for follow-up spectroscopic studies
 - Bands: X-ray, UV, optical, near-IR, mid-IR
2. Magellanic Clouds

- Analysing time variation of stellar color-color diagrams
- Bands: optical, IR, radio

Suggested Tool Capabilities

- Visualize data-cube
- Cut-out service for field selection
- Retrieve selected data
- Re-extract SED (e.g. sextractor)
- Create source catalog
- Generate colour-colour plots

Work Area 2 – Interoperability

OPTICON working group released VOTable V1.0,
A Proposed XML Format for Astronomical Tables,
on 15 Apr 2002 and right on schedule

Roy Williams , California Institute of Technology, USA	NVO
François Ochsenbein , Observatoire Astronomique de Strasbourg, Fr	AVO
Clive Davenhall , University of Edinburgh, UK	AVO
Daniel Durand , Canadian Astronomy Data Centre, Canada	CADC
Pierre Fernique , Observatoire Astronomique de Strasbourg, France	AVO
David Giarretta , Rutherford Appleton Laboratory, UK	STARLINK
Robert Hanisch , Space Telescope Science Institute, USA	NVO
Tom McGlynn , NASA Goddard Space Flight Center, USA	NVO
Alex Szalay , Johns Hopkins University, USA	NVO
Andreas Wicenec , European Southern Observatory, Germany	AVO

Document repository: <http://cdsweb.u-strasbg.fr/doc/VOTable/>
Comments: VOTable@us-vo.org

VOTable

- Early releases of VOTable APIs for Perl and Java are available for download:
 - JAVOT: Java parser for VOTable (R. Williams)
<http://www.us-vo.org/VOTable/JAVOT/>
 - Perl package VOTABLE (T. McGlynn, E. Winter)
<http://heasarc.gsfc.nasa.gov/classx/pub/votable/>
 - C++ package VotableParser: (S. Kale, VOI) <http://vo.iucaa.ernet.in/~voi/html/iucaawriteup.htm>–Perl module (R. Plante)
- Conducted experiment: Used perl package to transfer observation logs of ESO and HST instruments from ESO/ST-ECF to CDS in VOTable format (1 Mio. rows, ~500 MB of XML data). Included UCD tagging of columns.

Work Area 3 – Infrastructure (A. Lawrence, AstroGrid)

- Coordination Meeting in Leicester attended by AstroGrid representatives and WA3 WP manager A. Wicenec.
- AVO expressed interest in participating in AstroGrid network tests
- Next Generation Archive System NGAS:
 - Disk configuration tests
 - Implementing management software

Next Steps

- Finalize recruitment process
- Finalize definition of use cases for Jan. 2003 demo
- Define and implement specific WPs to meet use case requirements
- Decide on venue and format of Jan. (?) 2003 demo
- Set up collaborative tools (e.g. Wiki)

- Improve Web presence
- Prepare for discussion of IVOA standards and interoperability tools

Discussion:

- Conversion of HST and ESO observation logs to VOTable: about 1 million rows, or ~100 MB. Worked quite well. Some problems encountered with special characters. Large data volume shows some inefficiencies.
- Science demonstrations, final details TBD. Will prioritize components.
- AstroGrid meeting of workgroups held on June 26-27.
 - Taxonomy (Tony Linde)
 - Data warehouse (Clive Page)
 - OGSA (Guy Rixon)

CVO (David Schade)

- CADC has now obtained ~\$C4.5M for CVO activities
- Results-oriented program
- VO prototypes, building from the pixels up
- Local services only so far
 - Hide internal complexity of archive from the user
 - Characterize all data by energy (bandpass), observation time, and position. Some information about “filling factors” also provided (spatial, spectral, temporal coverage and resolution)
- WFPC2 associations project
 - Covers ~25 sq deg
 - ~25M sources extracted and cataloged
- CFHT legacy survey
 - 500 nights over 5 years
 - 20-30 TB

Science Demonstration Status Reports

Gamma ray burst follow-up service (Tom McGlynn)

•Purpose

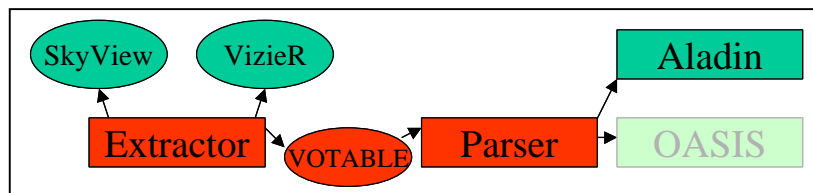
- Illustrate unification of data from disparate sources
- Use of multiple user interfaces to same underlying data
- Utility of common data standards

•Revised description provided 7/17

- No comments received...
- Generate multiple overlays of the same region giving multi-wavelength images and catalog overlays.

Current Status

- Simple testbed using SkyView images and Vizier catalogs.
- Optical, IR, Radio and X-ray images
- Several catalog overlays.
- Generate composite VOTable table from these components
- Parse VOTable and load information into Aladin



Demo Issues

- Standardized access to resources
- Aladin limits
 - Labels
 - Mixing scripts and interactive access
 - Customization of catalog overlays
- No standardization of metadata for complex, multi-resource VOTables.
 - How is relationship of one resource to another described?
- Caching versus direct access to remote resources
- OASIS access/limits

Development Issues

- No formal configuration management, testing, coding standards
- Where is code to be delivered?
- How can effort be distributed?

Schedule

- 8-1 to 9-15
 - Use Image cutout and cone search protocols
 - New services: 2MASS Mosaic, DSS at ST?, FIRST at LLNL, VizieR catalogs, IMPRESS
 - Refine Aladin Interface (possible small modifications to Aladin itself)
 - Explore OASIS interface
 - Integrate ‘pipeline’ into single Web interface
- 8-1 to 9-15
 - Use Image cutout and cone search protocols
 - New services: 2MASS Mosaic, DSS at ST?, FIRST at LLNL, VizieR catalogs, IMPRESS
 - Refine Aladin Interface (possible small modifications to Aladin itself)
 - Explore OASIS interface
 - Integrate ‘pipeline’ into single Web interface
- 11/1-12/15
 - Testing and verification
 - Development of demo scripts
 - HTML interface?
 - Logging, feedback capabilities
 - User customization?
 - User publication?

Discussion:

- Re/ request for code repository... we had agreed to hold off on this until after the first demos were done, but now people seem more eager to have some place to deliver their software. General agreement about using CVS. Several groups are potentially willing to host a CVS repository.
- David Schade suggested that the CFHT legacy survey might be a possible data source, though the quality of the astrometry is not very good.

Galaxy morphology/x-ray flux (Ray Plante)

Demo Description

- Version 2.0 on-line
- SWG feedback on Science Summary
 - 2.0 focused on relationship b/w galaxy morphology and cluster medium, as probed by x-ray
 - Request to shift more emphasis to cluster evolution questions.
 - Version 2.1 will integrate feedback.
- Technical Goals

1. Demonstrate how **diverse data resources** can be incorporated into **grid-based computational analysis**
 2. Demonstrate the **discovery of catalogs** relevant to a scientific question from **diverse locations**.
 3. Demonstrate the generic acquisition of **image cutouts** from various sky surveys.
 4. Demonstrate how a table interchange standard can **allow users to integrate their own data** into NVO-based analysis and publishing.
- Care w.r.t. what is do-able.
 - Target datasets: requires support for standards

Required development

- Richer registry for discovery of data & services
flexible in how far we take this
- Cone search support for target galaxy catalogs (SDSS, 2MASS, others?)
- Client for VOTable driven cone searches
- X-ray flux service at CXC
 - Interface worked out
 - 2-stage development plan
- Cross-match service (can use Roy's general service)
- **Condor-based computing scheme**
 - Mechanism for submitting job (on choice of machines)
 - How to handle data transfer (pre-stage?)
 - Deploying software
- Client-side plotting
 - Will use a package familiar to developer (WIP)
 - Table to image gridder out there?

Discussion:

- Gretchen Greene: Can we bring in smaller catalogs and data sets? Not just GSC, 2MASS, etc. David Schade suggested CNOC-1 (16 rich clusters of galaxies), and agreed to try to get this data set available quickly. See <http://www.astro.utoronto.ca/~cnoc/cnoc1.html>.
- Reagan Moore: What will be done with results from demo? Will view interactively and write into VOTable. Could do more, i.e., save image cutouts as a data collection in SRB that can be shared easily with collaborators.
- Resource/service metadata needs to characterize the type of service
 - High-level integration (e.g., NED)
 - Original source (e.g., DSS, GSC)
 - Processed or reprocessed (e.g., SkyView version of DSS)
- Schedule requires having the following implemented within the next month:
 - Image access service
 - X-ray flux service
 - Condor computational environment

Brown dwarf search (John Good, Bruce Berriman)

This demonstration is basically an extension of existing services.

Discussion:

- Could demo use SDSS first large release data prior to full release? EC will look into this with SDSS management.
- Resolved SDSS z-band-only detection questions.

Science Demonstration Recap

- Need stronger science description in the brown dwarf demo write-up. Bruce will update within a week.

- No one losing sleep over concerns.
- Ray will make a list of specific data availability expectations and circulate.
- Bruce would like to get more new science out of BD demo, through access to first SDSS data release.

EPO workshop and update

Goals of the Workshop

- Identify key capabilities that outreach developers want incorporated into NVO infrastructure
- Inform potential outreach partners about NVO status and future plans
- Begin to develop an NVO outreach community

Workshop Participants

- STScI
 - Voit, Hanisch, Stanley, Kakadelis, Stoke, Eisenhamer
- Space Science Institute, UC Berkeley
 - Hawkins, Spitz, Craig, Schultz, Mendez
- NOAO
 - Isbell, Pompea
- Chandra Science Center
 - Lestition
- Sloan Digital Sky Survey
 - Raddick
- NASA ADC / Raytheon
 - Borne
- Spitz Planetariums
 - Lantz
- Starry Night / Space.com
 - Whipps
- CLEA / Gettysburg
 - Marschall, Snyder
- NSCA
 - Thakkar
- AAVSO / Hands-On Astrophysics
 - Mattei
- Hands-On Universe
 - Pennypacker
- Loch Ness Productions / Sky & Tel.
 - Peterson
- AMNH / Hayden Planetarium
 - Liu

NVO E/PO Environment

- Follow the model of overall NVO project
- Many entry points tailored to needs of diverse communities
- A constellation of web services that enable effective outreach

Recommendations: Software Tools

- Protocol for outreach queries
- Popularity/suitability index for famous objects
- Metadata for outreach purposes
- WCS for press-release images
- Amateur photography archive
- Directory of educational resources
- Custom mosaic service that builds public-quality images
- Current-event service

- “Internet Sky Database”
- Mechanism(s) through which students can access “fresh” data
- Amateur photography archive
- NVO Object Log
- Capability of searching on object type
- Tool to convert VOTable to ASCII table
- Robust FITS to TIFF (etc.) converter
- Quality-control metadata

Recommendations: Community Development

- E/PO mailing list & discussion board
- Workshops on developing NVO-based outreach products
- Recruit amateur astronomers to be outreach partners
- Resolve image-credit issues

Recommendations: Funding Vehicles

- Tailor NVO outreach programs to existing NSF/NASA grant opportunities
- Plan for outreach component to be part of NASA-funded NVO support
- Seek funding for development workshops

Schedule/scope review, team working relationships, processes

Walked through entire WBS and identified schedule updates. In many areas we are ahead, having started work not originally planned until next year. Bob will update master schedule and distribute.

WBS 2 (Data Models) and WBS 3 (Metadata) still have a lot of overlap. Jonathan, Ray, and Doug will discuss and suggest how to further integrate these WBS elements.

Desire was expressed for greater visibility into the international projects. Encouraged all team members to regularly visit the project web sites and bring items of interest to our various working groups via e-mail.

Next meeting

Current preference: Nov 6-7, followed by external advisory committee on Nov 8. Baltimore.

Alternates: week of Nov 18-22, three consecutive days.

Start time will be 1pm to allow morning travel on the first day, and keep meeting sessions from getting too long.

Action Items

1. Metadata Working Group: Provide definitions of “data model”, “metadata”, etc., and document these and show examples. 3 Sept 2002.
2. Bob: Resume work on resource and service metadata document. 15 Sept 2002.
3. Ray: Develop service registry sufficient to support initial demos. 9 Aug 2002.
4. Bob: Identify site for project software library and designate code librarian. 15 Sept 2002.
5. David: Make CNOC-1 and WFPC2 associations available for science demos. 1 Oct 2002.
6. Bruce: Update science rationale for brown dwarf demo. 9 Aug 2002.
7. Ray: Make a list of specific data availability expectations for galaxy morphology demo and circulate. 17 Aug 2002
8. Jonathan, Ray, Doug: Make suggestions for better integration of WBS 2 and 3. 17 Aug 2002.
9. Team: Regularly visit international VO websites and send around notices of interesting developments. Ongoing.
10. Executive Committee: Investigate possible access to SDSS data prior to full public release.
11. Executive Committee: Finalize membership selection for external advisory committee.
12. Bob: Update master schedule, following completion of AI 8. 30 Sept 2002.