NVO Project Plan Update
February 2005

As the NVO development project continues in Year 4, and will be formally concluding about 20 months from now, it is important to keep our primary scientific and technical goals in mind. We want to complete a core NVO infrastructure before the conclusion of the project, leaving a follow-on NVO Facility with a solid base. The NVO infrastructure is developed in concert with the IVOA, meaning that NVO development schedules must recognize the necessity of coordination with the IVOA working groups.

For 2005 the NVO's first priority high-level science capabilities are:

1) Large-scale cross-correlations
2) Large-scale image mosaicing and stacking
3) Footprint/sky coverage-intersection services
4) Variability/change detection
5) Putting VO tools and software development environment into the hands of the community, and ensuring scalability
6) Incorporating VO capabilities into legacy software systems

These are to be done as we

7) Transition from open-loop development to facility operations

We will ask the Science Steering Committee to review and confirm these priorities, and to help define specific applications and use-cases.

These science capabilities motivate an emphasis on the following technical initiatives.

- We need to manage large-scale computations and large-scale data storage. This calls for the development of **VOSpace** (a user environment for controlling applications and interacting with distributed computation and storage facilities) and **VOStore** (the underlying distributed data storage management system. VOStore might be implemented with SRB, CASJobs, a local file system, etc. AstroGrid MySpace merges data management and workflow into one environment/system.

  VOSpace/VOStore is the next major piece of the VO infrastructure. This is required for large-scale jobs, sequencing, and utilizing the Lego pieces that are software components and web services. VOSpace encompasses issues such as registration, authentication, and authorization. This activity cuts across at least two NVO WBSs and will require a new WG and work plan.

- Scalability. The NVO will need to address how services scale. This means that users see a smooth transition from synchronous to asynchronous services, and we need to define how the latter is implemented. It means a security model, so that access to private data and large-scale resources can be controlled, logged, and traced back to a

human client. Systems like Astrogrid-CEA, Kepler, Pegasus, Condor need to be examined as ways to construct complex applications workflows and implement scalable services.

- Footprint/region services. The tools are "ready to hatch" based on the current maturity of STC and the nearly-complete footprint/intersection service developed by Alex Szalay and Jim Gray. Will need a way to convert a FITS WCS to a region specification. Similar services exist already, e.g., in AVO/Aladin. Arnold Rots will coordinate efforts in this area.

- OpenSkyQuery/OpenSkyNode. The Java full SkyNode implementation is almost done. The SkyNode specification is incomplete and difficult to follow for people outside the original SkyNode development group. OSQ/OSN need to support larger queries, which links back to the need for VOSpace/VOStore.

- Registry. **Curation** is the main topic for this year. We need to implement automated verification and validation tools, plus incorporate human quality checks. We must capture curation metadata and identify methods for authorizing changes. The issue of registry granularity will be debated further (in the context of completeness/ correctness). We also need to register software and identify necessary metadata extensions (WSDL, parameter mechanism). Other activities include finishing various odds and ends: the registry interface standard, integrating STC into the VOResource schemas, documentation of VOResource, and a general tutorial.

- Data Access Layer. First priority tasks are completion of the SSAP and Spectral DM. These are coming along; a 3rd draft should be ready in mid-February for limited circulation. SIAP/DM/data characterization and SIAP metadata extensions are also in progress (with F. Bonnarel). ADQL integration and reference implementations will follow.

  A second thrust in the DAL area is the integration of VO data access services with data analysis systems. Work is required on the architecture, the component-container model and execution framework, the VO client interface, and on wrapping data analysis code to produce VO web services (e.g., M. Fitzpatrick's work on IRAF).

- Variability and events. As significant new facilities such as PanSTARRS, Palomar QUEST, and LSST are designed and constructed, and as extant datasets such as the MACHO collection are published to the NVO, we need to assure that the NVO infrastructure (e.g., metadata, data access protocols, and event notification services) meets these projects' needs.

- Logging. A logging specification has been drafted and is ready for prototyping. We should pick ~5 services and test harvesting and collation. This activity is the responsibility of the Web and Grid Services WG.

In addition, we must reach closure on a code management system and pay attention to user support issues (monitor feedback@us-vo.org and consider a more sophisticated call-tracking system).