The Report of the NVO Advisory Committee

November 20-21, 2002

Executive Summary

The NVO Advisory Committee met for the first time on November 20-21, 2002, at Johns Hopkins University in Baltimore. Committee members in attendance included Gerry Gilmore, Tony Hey, John Huchra, Carl Lagoze, Eve Ostriker, and Sidney Wolff. Presentations were made by key members of the NVO team.

The Advisory Committee was very positively impressed by the progress made in the short time since funding was received. In particular, we highlight the following achievements:

> The very challenging task of setting up a highly distributed project with partners in a range of institutions has been met with a very high level of proficiency.
>
> The team appears to be working together very smoothly with an admirable division and coordination of distributed tasks.
>
> The work to date shows a strong commitment to leveraging numerous efforts in the web and general information community, including metadata work, web services, and the semantic web.
>
> The NVO has made very considerable progress in developing specific tools and capabilities. One good example is VOTable, but there are many others.
>
> Real implementations to scientific problems are already in place and will be demonstrated to the community at the AAS meeting in Seattle in January.
>
> The NVO staff are participating actively in, and providing significant leadership to, international VO efforts, including the development of internationally accepted standards for astronomical metadata. The NVO Project Manager (Hanisch) is serving as the Chair of the International VO Alliance (IVOA).
>
> The presentations demonstrated considerable progress based on a well-grounded understanding of the technical problems that the project faces.

We make several recommendations for the future development of the program:

> The project should provide a 5-year roadmap for the project, including education and outreach. This roadmap would define the goals of the program and would specify yearly milestones that can be used to measure progress toward those goals.
>
> The NVO is very likely to become an essential and powerful component of the infrastructure of research in astrophysics, and it will be necessary to plan for its the long term sustainability. The NVO staff should develop recommendations as to possible long-term NVO implementation and operation strategies, in order to

allow preliminary planning for the longer term. This document is likely to evolve as more experience arises both in NVO and internationally.

In parallel with their development of prototype NVO services, the project should become pro-active in nurturing “early adopters” to conduct new types of astronomical research through the use of NVO technology.

The project should conduct an analysis that enumerates the external developments upon which the results of the project are dependent, determines the potential risk areas for each, and provides reasonable fall-back courses of action for those that present unusually high risk.

Finally, we note that there is no strategy in place for archiving ground-based optical/IR data at either the national or the large private observatories. We believe that a review committee should be tasked with deciding what if any of the ground-based data should be archived, what appropriate rules of use and ownership are, and what levels of support for archiving will be required. One strategy for encouraging broad community support of the recommendations of this review would be for the NSF to sponsor it, and the review might be conducted by the Committee on Astronomy and Astrophysics of the National Research Council.

1. Structure of the NVO

The NVO is in the early stages of development, an ideas-led phase of the project and one that necessarily requires substantial creativity. The current project structure is appropriate to this phase. The work is very highly distributed, with partners in a range of institutions bringing complementary expertise. The direct links with the Grid and supercomputing communities are an exceptional strength of NVO, and one example of ways in which the NVO project is significantly in advance of other international VO efforts. Another strength of the distributed approach is that it ensures that a broad community will be aware of the NVO and its potential applications.

While the strengths are real, this project structure does inevitably impose considerable demands on project management. Since the many partners act as willing contributors to a joint project, rather than formal (sub-)contractors, there is a risk that potentially important contributions will not be made—or will not be made in a timely fashion. Some means to identify ‘back-up’ contingency effort may be desirable, based on the actual rate at which deliverables are received from the various partners. For example, it is apparent that rather little effort has actually been delivered to date on time-series data and their handling.

Fortunately for NVO, the key participants are highly motivated, clearly fully committed, and exceptionally knowledgeable and talented. Both the volume and the quality of the advances made so far are exceptional.

2. Project Management

Project management at the present stage of NVO is relatively informal, and this is appropriate at this stage of the project. The current approach has allowed creative scientists to explore options in a fast-changing environment and to remain fully involved with those developing what will become the standards, software tools, and infrastructure environment that will be relevant as the project evolves and develops into a valuable entity of long-term benefit to a wide community. Such informality is well-matched to the academic research phase of the NVO and is endorsed by the Advisory Committee. Because the NVO is being developed in a rapidly changing environment, a strong element of flexibility and adaptability will be required throughout the project.

As the project develops, however, the evolving balance between creativity, new concepts, and specific deliverables will require greater use of the formal elements of software engineering: documentation, supportable code, and standards. Inevitably, an evolution of the project structure will be required. The project team should develop a management plan to handle this evolution.

The very considerable experience available in the various data centers and within the NVO itself should provide the basis to ensure that a suitable balance between creative informality and adequate systems engineering is met.

Our assessment of project management performance to date is clear: The NVO project management is to be congratulated for achieving a very promising start, remaining focussed, avoiding drift into an unfocused ‘everything for everyone’ mode, and already demonstrating beta-operational science test applications of considerable technical challenge and huge scientific research potential. The recommendations here address possible enhancements and evolutionary developments--they are decidedly not a criticism or suggestion of inadequacies.

3. NVO Roadmap

The NVO staff presented very clearly the achievements of the first year of operation and the plans for the second year. It would be easier to assess these achievements if they had been placed in the context of an overall outline of the goals for the entire five years of the program. Specifically, the Advisory Committee recommends that the NVO project leaders provide a high-level “Roadmap” for its efforts over the next four years of the project. In the Annual Report documentation provided to the Committee, a detailed schedule of tasks/accomplishments was presented. Working from this detailed tabular material, together with the original scope of work defined in the proposal for the project, NVO management should develop an NVO Roadmap (not exceeding a few pages) that presents in an easily-readable form for community distribution:

> Introduction/mission statement for the NVO project, e.g. who is directly involved in NVO; who are its research partners and sponsors; what the “end product”' of this project will be; how the U.S. astronomy community may contribute to the VO effort; what the overall development strategy is, including how core standards, services, and sample science studies will be interleaved.
>
> A summary of both yearly milestones and the final goals of the Data Standards development effort.

A summary of both the yearly milestones and final goals of the Core Services development effort.

A description of yearly milestones and final goals in other areas, including: 1) the plans to provide web-accessible documentation/tutorials on VO science-user and data-center prototype tools; 2) running research community "missionary'" events/activities (e.g. ADASS special sessions) for information/training in data conversion to VO standards and use/incorporation of VO tools in software development for science studies; 3) establishing partnerships with public-outreach efforts sponsored by other organizations.

4. International VO

The NVO Advisory Committee was particularly impressed with the active role that NVO staff have taken within the larger VO community. In Year 1 of this project, the International VO Alliance (IVOA) was established. This body has representatives from all the VO initiatives worldwide, with the mission of facilitating the international coordination/cooperation needed for achieving interoperable VO data standards and services. In his role as Chair of the IVOA, NVO Project Manager Hanisch has been communicating with other IVOA members on a bi-monthly basis. NVO has actively collaborated with the European VO consortium (AVO) and the UK VO program (AstroGrid) on the prototype for the core data format, VOTable, as well as other aspects of establishing standards/classification schemes (e.g. the UCD) for data models and metadata. NVO has also worked with European/UK colleagues on the prototype for the Simple Image Access archive retrieval protocol. The NVO team has made an effort to leverage/incorporate successful services already developed by international organizations, e.g. the Aladdin data viewer engine operated by CDS (Strasbourg).

5. Standards

Standards are critical. Access to distributed data would be quite impossible without an agreed set of standards defining the very large number of pieces of information required to find, extract, and use data. NVO illustrates well the commitment to standards on which the future is being built. A specific success is VOTable, which has been adopted internationally as a standard for astronomical data representation. Definition and adoption of VOTable as a standard was built on a huge amount of work over many countries. It involves a large number of people, considerable knowledge, and a great deal of thought. Its rapid adoption underpins many applications, and allows much progress. It has been designed to be compatible with the future. The NVO team, especially Roy Williams, made a very significant contribution to the VOTable standard.

Additionally, NVO realised immediately that a standard without an application is less valuable. Thus, real implementations are already in place, with others under development. This progress--definition of a global standard, its implementation, and real use in a research astronomical application--is a remarkable achievement so early in the NVO project. Arguably even more impressive is the explicit intention to monitor , modify, and extend VOTable on the basis of experience. This outward and open attitude, always looking for improvement, remodelling when new information becomes available, and working with others to improve common tools is illustrative of the very considerable strength of the NVO project to date.

6. Metadata Strategy

The committee was impressed with the metadata strategy presented by the project team. It demonstrated a good understanding of the types of structured data needed for the intended services. We do have two comments for consideration:

- The group presented a metadata format that was based on the Dublin Core element set. The format presented had a number of qualifications of the DC elements that make it incompatible with other expressions of Dublin Core. This observation is based on the definition of qualification at http://www.dublincore.org/documents/2000/07/11/dcmes-qualifiers/, which (paraphrased) states that qualifiers should narrow the semantics of the respective element. In particular, the team presented some qualifications of the "coverage" element, which clearly extend rather than narrow semantics. As stated in the meeting, this is a matter of concern if the group expects to export these admittedly localized Dublin Core records external to the project. In that case, external handlers of the metadata will effectively misinterpret the portions of the records that stray from Dublin Core guidelines. These extensions are not a problem if metadata records are exchanged only internally to the project. In the long run, the group may find it advantageous to move away from Dublin Core elements in the case of these semantic violations, and create elements in a separate name space.
- Metadata quality and cost is a thorny issue that was not adequately addressed in the presentation. The metadata presented was clearly analyzed from the perspective of data needs of services intended by the project. However, it was not entirely clear where responsibility for the metadata rests. If production of the metadata is internal to the project, it is important that the group has some recognition of the per record costs of metadata production. These can be non-trivial, especially when a large number of metadata records are expected. If production of the metadata is external--e.g., by distributed partners--the group needs to recognize the inherent quality problems of distributed metadata production. The best metadata designs are meaningless in the face of sloppy individual record production.

7. Data Provenance

The success of the project is strongly dependent on the quality of data from external sources. There are a number of potential risks with such dependence that should be addressed by the risk analysis recommended below. One means of mitigating such risk is to pay some attention to issues of data provenance--who created the data sets, when and by whom they were modified, etc. The group should look at some of the current work on data provenance with an eye to understanding its applicability in the project. Some of the best work in this area is by Peter Buneman, formally at U. Penn, with a web page at http://db.cis.upenn.edu/Research/provenance.html.

8. NVO Sustainability

Perhaps one of the most significant project management challenges still to be addressed fully is the transition from the present creative development phase of NVO, which is funded for 5 years, into a long-term, sustainable, and supportable component of the essential infrastructure that supports astronomical research. There are many possible models for the long-term future of NVO, including continued existence as a group defining standards in collaboration with the rest of the International VO, through to a more material existence as an organisation supporting standards, software, and community data-center developments. Some possible models imply support scientists, career structures, and consideration of the relative merits of independent existence or association with some (group of) other organisations. Other possible models exist. The NVO project team has very considerable wisdom, technical skills, and experience directly relevant to outlining the future of the NVO program, both in its implementation within the NASA-NSF context and in its interrelationships with the continuing IVOA alliance. It is desirable for the NVO project to make the "lessons learned" during its evolution available, as a report, with recommendations as to possible long-term NVO implementation and operation strategies in order to allow preliminary planning for the longer term. This might be an evolving document, as more experience arises both in NVO and internationally.

The NSF also needs to give thought to how the NVO will be supported in the long term when the major activities will not be R&D but software maintenance and service support.

9. Creation of NVO User Community

Although the NVO collaboration is only funded to create the framework for the NVO and not the NVO itself, there is a clear need for the project to begin the development of a user community. In parallel with their development of prototype NVO services, the project needs to be pro-active in nurturing "early adopters" of NVO technology to do new astronomical research. These adopters might be individual users or the data centers that are providing data to the NVO. Although these early adopters will have to put up with unreliable software and services in the early years of the NVO project, unless there is an embryonic community of enthusiastic users who have successfully used the NVO to do new science there is a real risk that the NVO will not be able to make the transition to a persistent NVO service when the project ends. This process could perhaps be assisted by NSF small grants as an incentive for NVO early adopters. The importance of developing "NVO science champions" for the use of NVO technology to do new astronomy cannot be over-estimated: "technology champions" without tangible evidence that the technology enables new science will never be as convincing.

The Advisory Committee was much impressed by the progress made toward developing a set of NVO standard services. The development of a set of interesting services that are easy to use and deliver science results that could not otherwise have been obtained without significant effort is a key strategy for the project. In this context the science demonstrations will play a very important role for the project and should be used to attract some early adopter users.

10. Risk Analysis

A project such as this, which involves dependencies on both distributed project participants and external technologies, faces numerous risks. The convergence of the Grid Services and Web Services in terms of WSDL specifications is a helpful evolution for the project. It should make interoperability easier and reduce the risks of technology obsolescence. However, since this is a fast moving area, the project will need to continue to track developments closely and be prepared to undertake several re-implementations as the OGSA standards evolve.

Many of the standards that the group has decided to base their work on--for example, the semantic web--are fragile or in considerable flux. There is nothing the group can do to ensure that standards remain stable. However, it would behoove the group to undertake some standard risk analysis of their work. Such analysis should enumerate the externalities upon which the results of the project are dependent, determine the potential risk areas for each, and provide reasonable alternative courses of action for the areas where the risk is high. The goal here is not to guarantee success, which is clearly impossible in an innovative project such as this, but to account for potential problems and demonstrate readiness to adjust plans in the face of external failures.

Another potential challenge lies in the time required for cross-matches involving distributed data archives. The ability to deliver specified sustained bandwidth and latencies will be an important issue for NVO services and will certainly be dependent on the user location. More help should be sought from network experts, and the project could perhaps consider the deployment of monitoring tools to look into this network "Quality of Service" issue.

11. Education and Outreach

The NVO will be an important resource for formal and informal education and public outreach. In order to assess this potential, the NVO sponsored a workshop to identify potential users and to suggest products and services that will serve this very diverse community. The product of this workshop was a prioritised list of software tools and protocols.

In this as in other areas of potential investment, it is important to remember that the task of the NVO project is to create a framework that supports a variety of applications and not to carry out directly all of the functions enabled by the NVO. As noted in the report from the workshop, most of the education and outreach activities will be carried out by experts from the communities that are being served.

The list of tasks devised by the attendees at the workshop is extensive, and it is unclear how much of this can or should be accomplished by the NVO itself and how much should be the responsibility of partners. We recommend that the outreach specialists leading the NVO effort develop a roadmap for their five-year effort, just as we have recommended for the other portions of the NVO program. This roadmap should identify a time sequence of tasks matched to resources allocated for outreach. Clearly one high priority task is to establish metadata standards that will serve the education and outreach communities. Tools to construct queries of the kind likely to be submitted by non-professional astronomers will be needed. Workshops to inform educators, outreach developers, art and science museum professionals, etc. about the potential of the NVO and how to access it will provide leverage. Simple software tools to manipulate images will be needed. However, beyond these basics, the program could go in many directions. A roadmap will be helpful in making the necessary choices.

12. Archiving Ground-Based Data

An important issue for the NVO is the availability of data archives for ground-based astronomy. This is also a critical issue for the NSF, which provides much of the funding for research and instrumentation at ground-based telescopes, including at the large private observatories. While general data archives are in place for many of the major foreign observatories (e.g. ESO) and for some of the larger and more well planned surveys done in the US (2MASS, SDSS, NOAO), the vast majority of more specialized data taken with ground-based optical/IR telescopes is not preserved in any form accessible to the NVO. Despite considerable discussion of this topic for well over a decade, there are currently no real plans to do so.

The NVO Advisory Committee believes that this issue should be addressed soon, as the design of the NVO depends on the quantity and quality of the data and of the number of nodes expected to provide data and other information. As this issue transcends the NVO, it would be appropriate for the NSF itself to address it, perhaps through an independent yet broad organization such as the NRC's Committee on Astronomy and Astrophysics.

Such a review should be tasked with deciding what if any of the ground-based data should be archived, what appropriate rules of use and ownership are, and, if necessary, what reasonable levels of support for archiving at both the public and private observatories should be. We note that NASA has successfully implemented a general one-year maximum proprietary period for space data and also maintains through its data centers the archives necessary to make that data publicly available after the proprietary period is over. However the level of support necessary for this effort is not small and may never be realizable for the archiving of the general data flow at ground-based observatories. If that is indeed the conclusion, it would be important to reach it soon and move on.

We also note that a different solution to the general data problem might simply be for both the funding agencies and the journals to require that authors deposit usable forms of their data with existing data and knowledge bases such as the NSSDC, the CDS, MAST, or the journals themselves. That will also require the identification of resources, but places most of the burden of data quality control and reduction on the scientific community.

While addressing this issue may technically lie outside the charter of the NVO Advisory Committee, we do request that this issue be highlighted when our report is transmitted to the NSF.