

Princeton University

Department of Astrophysical Sciences
Peyton Hall
Princeton, New Jersey 08544-1001

Telephone: (609) 258-3808

Email: strauss@astro.princeton.edu

January 18, 2007

Dr. George Helou, IPAC Executive Director
Dr. Joseph Mazzearella, NED Project Scientist
and Task Lead
Dr. Barry Madore, NED Senior Scientist
Infrared Processing and Analysis Center
Caltech/JPL, MS 100-22
770 South Wilson Avenue
Pasadena, California 91125

Dear George, Joe, and Barry,

As chair of the NASA Extragalactic Database (NED) User's Committee, I am writing to summarize the recommendations that came from our very productive meeting with the NED team in Pasadena on 2006 October 26 and 27. We heard detailed presentations about the current status of NED and the challenges facing it, and had in-depth discussions about possible future directions for the project. This letter summarizes our findings; a second letter gives more detail on some of the specific NED features you had asked us to comment on.

Our first, and most important conclusion is one that you've heard consistently from the User's Committee in the past, namely that over the last decade or more, NED has become an absolutely essential part of the toolbox that all extragalactic astronomers use. The almost 5000 citations to NED in the literature to date are an indication that NED is being used extensively for a broad range of science investigations. The extragalactic literature is vast, and NED is the tool that we all rely on to use it to its full potential. NED, with its object-centered orientation, is much more than an archive of high-level data products (although that is a very important aspect of its operations; see below); the human expert vetting of the data that are taken from the literature often make NED a more useful repository of data than the original papers themselves. Our first conclusion therefore is that **NED must be given the resources it needs to carry out its core mission: to ingest the extragalactic data of the refereed journals and the high level products of the major surveys of the sky.**

Now that the NED enterprise has matured, there is a variety of ways it is looking to expand, and much of our discussion was focused on priorities and where resources should be placed. We felt strongly that NED should continue keeping its focus on its core mission, namely the incorporation of the astronomical extragalactic literature and major survey catalogs in a *complete* and scientifically coherent way. This completeness is paramount: a researcher needs confidence that when NED finds no near-infrared photometry for a given galaxy, for example, they can conclude that these data do not exist in the refereed literature. NED is not in itself a data analysis tool, and therefore additional capabilities currently available in NED, such as spectral plotting tools, are of lower priority.

The incorporation in an intelligent way of data from the literature and from major surveys requires the work of PhD-level astronomers. Because of the wide range of types of data to be incorporated, the different conventions, calibrations and reference frames used, and the various errors astronomers make in compiling tables in the literature, this process cannot be completely automated, but often requires astronomically expert human invention. However, much of the routine aspects of searching through and ingesting the data from the literature has been automated to an impressive extent, and it is hard to see how the efficiency of the system can be further improved. While the NED team are encouraged to continue improving the automation of this process, there is no more fat to be trimmed. **The astronomical expertise the team brings to the process of incorporating literature and survey data into the NED database cannot be cut back further, and probably needs to expand by another FTE.**

The work NED does incorporating data from the astronomical literature makes clear the inadequacies in vetting data in the refereeing and publishing process. While the scientific content of a paper is scrutinized carefully when it is refereed, there tends to be less scrutiny of the data that are published as part of that paper. We learned that many published data tables do not follow accepted rules for naming objects or fully identifying the quantities listed. Indeed, the journals themselves do not give full guidance on what is appropriate here, and the referees are given minimal guidelines as to what is needed. The NED team understands what standards a well-constituted data file should adhere to. We therefore recommend that **NED work closely with the major US journals on developing criteria and guidelines for authors, editors, and referees for what constitutes a good data table.** One could imagine an automated web-based service for uploading data tables to test that they are NED-compatible.

As we have already indicated, the fact that NED is playing a long-term stewardship role for astronomical data found in papers is absolutely crucial. In addition to data tables, the data for a paper can include extensive images and spectra, which are often not published, given the limitations of page charges. NED could serve as a long-term repository of such data. We therefore recommend that **NED explore with the journals the possibility of being a data archive for images and spectra that accompany papers published in the journals. This would be a substantial broadening of NED's responsibility, and NED would need additional resources to take this on.**

We understand that the support of the NASA data centers are set by the Senior Review, and that this review has historically considered both data centers and active space missions together. This gives an unbalanced view, in part because space missions have finite lifetimes, unlike the long-term archives of data centers. We therefore applaud the suggestion that the next Senior Review consider data centers and space missions separately.

We had extensive discussions of NED and its relation to NVO. NED should be commended for developing much of the functionality that NVO has promised, and having made these tools and the data behind them available and scientifically useful to the community. Thus NED is in effect a data portal to the astronomical literature for the extragalactic community, and has thus been addressing many of the NVO goals for some time. Moreover, the NVO already uses many of the NED tools (such as the name resolver), and one can imagine a future in which NED and NVO become yet more closely linked, including the ability to carry out NED queries from the NVO front end. We strongly encourage this sort of cooperation, and are heartened to see that several key NED personnel are taking part in NVO meetings and management decisions, and that NED data archives are packaged in such a way as to be NVO-compliant.

The planned NVO will be very powerful, and will include many capabilities beyond NED's scope. However, NED does several crucial things that NVO needs, but does not do, nor does it intend to do, itself:

- Archiving of data. NVO is a gateway to pre-existing data sets available on the web. However, it does no archiving of its own, and takes no stewardship responsibilities for the relevant datasets. Moreover, the NVO does not currently provide any sort of direct access to the astronomical literature (as opposed to large stand-alone datasets). Given the importance of archival data in astronomy, the long-time stewardship of high-level data products from the literature is an absolutely crucial task that NED carries out. Similarly, the long-term future of data archives from surveys such as SDSS is not completely clear, and NED has the potential to play a leadership role in long-term archiving of major survey databases.
- Vetting of data. The data given in the literature are often poorly documented, with missing details on bandpasses for photometry, epoch for astrometry, reference frame for radial velocities, and so on. Inconsistencies in the naming and referencing of objects abound. The careful work that the NED team does on fully documenting all aspects of the data ingested, and sorting out the discrepancies, make the data much more valuable to the community than it otherwise would be.
- Source matching. The careful matching of source lists and the astronomical source name resolver that NED provides will be a continuing need for data in the astronomical literature as well as large sky survey catalogs. NED's highly automated procedures for establishing high fidelity cross-matches (and statistical associations where appropriate), improving upon them as new information emerges, and serving them as a basis for cross-matches to new observations, will be an important component of the NVO facility.

There will soon be an Announcement of Opportunity (AO) for the development of a formal NVO facility. NED, together with IPAC as a whole, are very well-placed to play a major role in such a facility.

This leads to our next recommendation: **NED should continue to work closely with the NVO community. Given the aspects of NVO that NED is already successfully serving to the community in a scientifically useful way, and given those aspects of data vetting and long-term stewardship that NVO needs but does not see as part of its core mission, the NED team can and should take a leadership role in the NVO effort in general, and in response to the NVO facilities AO in particular.**

NED is of course focused on extragalactic science, while most of these recommendations are relevant for stellar astronomy as well. NED does not currently have the resources to allow it to expand into stellar astronomy, but the NVO will have all of astronomy in its purview, and there will be a call from the astronomical community to have the capabilities of NED for stellar sources available from the NVO. The NED team are best suited to include stellar sources in their database, but *only with a very substantial increase in their resources*. Thus we recommend that **if there is a call from the astronomical community to have the capabilities of NED for sources within the Milky Way, that the NED team be given the opportunity to lead this effort, and be given a substantial increase in resources to do so.** In the meantime, we recommend that as large datasets such as SDSS are incorporated into NED, that the division between stars and galaxies need not be made strictly, given that star-galaxy separation is often ambiguous for quasars and for objects close to the magnitude limit. Therefore we recommend that **NED incorporate all high-level data products from large surveys, stars and galaxies combined, but endeavor for completeness in the literature for extragalactic sources only.**

In summary, NED continues to be a crucial and absolutely necessary resource for the entire extragalactic astronomy community. It is important that it keep its focus on its core mission of archiving the literature and major surveys, with a strong emphasis on quality, completeness, and data integrity. NED should work with the refereed astronomical journals to improve the refereeing of data, and to become a long-term archive for images and spectra that accompany a scientific paper. Finally, NED has already become a critical part of the NVO, and is strongly encouraged to be a major player in a proposal to respond

to the impending AO for the NVO facility.

Yours Sincerely,

Michael Strauss

NED User's Committee Membership:

Dr. Daniela Calzetti, Space Telescope Science Institute and University of Massachusetts

Dr. James Condon, National Radio Astronomy Observatory

Dr. Sarah Gallagher, UCLA

Dr. David Hogg, New York University

Dr. John Mulchaey, Carnegie Institution of Washington

Dr. Andrea Prestwich, Center for Astrophysics, Harvard University

Dr. David Sanders, University of Hawaii

Dr. Michael Strauss, Princeton University