

December 23, 2016

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

Dear Joe, George and members of the NED team,

This is a report of the face-to-face meeting of the NASA/IPAC Extragalactic Database (NED) User's Committee (NUC) held in Pasadena on December 6, 2016.

The most important message from our meeting is continued support for the basic NED mission. As you described in detail in your presentations, NED is a resource used throughout the extragalactic community, and is key to the success of on-going and planned NASA missions, including the Hubble Space Telescope, the James Webb Space Telescope (JWST), the Wide-Field Infrared Space Telescope (WFIRST), and many others. The NED user and citation statistics from NED speak for themselves, with well over 700 refereed papers per year citing NED, and hundreds of thousands of web requests per day. NED has become an irreplaceable and essential tool for essentially the entire world-wide community of extragalactic observational astronomers. We especially like the quote you showed us by George Paturel (director of the Lyon Extragalactic Database, LEDA): "In the field of extragalactic databases, the most important event has been the creation of the NASA Extragalactic Database (NED) created by G. Helou and B. Madore... We [LEDA] are happy to have slightly contributed to this remarkable database."

In what follows, we describe our recommendations on a number of issues that came up in our meeting. We put specific emphasis on two topics: the role of the NASA Astronomical Virtual Observatories (NAVO) in NED, and the status of the NED user interface. More detailed issues are covered in the Appendix to this letter. A common theme throughout is the importance of the database refactoring that the NED team is now undertaking. It is crucial for NED's ability to ingest the next generation of surveys. The refactoring is a huge and important job which impacts all aspects of NED's future, including the interoperability of the NASA data centers, as well as the work on NED's user interface.

NED and NAVO: We were asked to comment on the importance of interoperability of the NASA data centers: NED, HEASARC, ADS, IRSA, and MAST. This is certainly desirable, and will become increasingly important given the relationships between their holdings. We are, however,

concerned whether the use of the IVOA protocols as a framework for interoperability will allow the NASA Astronomical Virtual Observatories (NAVO) to build and maintain the services that scientists need in a reasonable timeframe with the resources available to the NED team. This concern stems from the complexity of the protocols themselves, and the fact that the previous incarnations of the virtual observatory (VOA and NVO), using earlier versions of the same standards, never resulted in services widely used by the astronomical community. With this in mind, we encourage NED, and all the NASA archive centers, to develop a clear plan for how the common implementation of the IVOA machine interface protocols will actually be used, both by the archives themselves, and by the astronomical community.

NED is in the process of a several-year effort to bring its database to a level of sophistication and flexibility to allow it to ingest the next generation of large surveys. This database refactoring is perhaps the single most important goal of NED over the next few years, and it is important that the NAVO effort be done in a way compatible with that refactoring and not to take resources away from it. The NAVO effort also depends on the completion of the refactoring, for an efficient and maintainable implementation within NED. The NUC members do not have sufficient database expertise to comment on the technical aspects of this question, but we feel that the push to implement the IVOA standards within NAVO only makes sense within NED to the extent that it is compatible and synergistic with the database refactoring and related effort.

At the moment, making NED work within the NAVO framework is requiring the extensive NED databases and query interfaces to be reconfigured to support data access through IVOA protocols. In the opinion of the NUC, NAVO can only succeed if it generates concrete well-justified products that are of real need to working researchers on finite timescales.

We considered a hypothetical service that would allow NED users to identify all archived observations obtained with NASA astrophysics missions for a given object. This capability currently does not exist, and would be quite valuable. For images, this task would mainly exercise the simple image access (SIA) protocols of NAVO. In 2016, NED delivered an initial version of its SIA service, mostly incorporating FITS images contributed by authors of journal articles. By leveraging the SIA services at the mission archives, those images could be presented along with the results of the mission archive SIA queries and integrated together in an easy-to-use service for users.

In this context, the implementation of the Table Access Protocol (TAP) that NED has proposed is a good plan, and a step down the NAVO road, potentially addressing much of the “power user” needs from the point of view of supporting SQL-type database queries as a web service. Another way of addressing those power user needs would be to deliver client side tools themselves (possibly adding modules to astroquery in python for example). But further development of power user tools are likely to require additional resources. Perhaps IPAC will eventually want to deploy something akin to the

JHU SciServer Compute program, which currently uses jupyter notebooks.

We were asked to address a specific set of questions related to the NAVO effort. We list and respond to them here.

- To what extent are you aware of the Virtual Observatory (VO) interfaces to the services of IRSA, NED, MAST, HEASARC, and ADS?

Response: We were largely unaware of them before the NUC meeting. One specific example we found that makes significant use of the NED holdings is the VO-based Iris SED fitting tool. We are also aware that the VO Data Discovery Portal has been folded into the MAST Portal with VO search capability, including cone searches and SIA-compliant image access to NED.

- Do VO activities further the science goals of NED?

Response: The primary goal of NED is to make highly vetted and cross-matched extragalactic data available to the community. As described above, the interoperability of all NASA data centers is desirable for a variety of reasons, and if NAVO is the best way to enable this, then yes, it is important for the NED science goals. However, as we've emphasized, the highest priority for NED development is the database refactoring. The NAVO effort should be done in a way that is compatible with that effort, and does not take resources away from it.

- Are you aware of using these interfaces directly or indirectly?

Response: As described above, we were not aware of these interfaces before the NUC meeting.

- Are there additional priorities in providing machine based APIs to our archive services that we should be looking at?

Response: Given the past experiences with less than fully successful VO efforts, the NUC does not recommend or encourage expansion to new VO APIs, but would encourage the continued focus on the current set of IVOA protocols until all NASA archives are in full compliance with the current set.

In conclusion, we strongly urge the NED team to take a project approach to implementing IVOA protocols, in which there is a fixed realistic development timeline, goals, milestones, and deliverables that directly advance and demonstrate progress towards the motivating science research problem to be solved.

NED's User Interface: The report from the previous face-to-face meeting of the NUC (dated January 15, 2014) put a great deal of emphasis on the design of the NED User Interface. We were glad to see that many of our detailed recommendations are being incorporated in the "simple search"

webpage. However, this work has been in progress for several years, and we were concerned that there was not a clear plan, with milestones, for finishing it expeditiously. Because this new interface is still under development, the user is presented with a rather inconsistent look and feel to the website; different items on its pull-down menu take you to very different-looking pages, confusing the reader. Our immediate recommendations are to:

- Develop a clear plan, with milestones, for finishing the website design expeditiously.
- Not wait until everything is perfect before incorporating new features into the website.
- Include prominent labels on the website, to indicate that it is work in progress (like the “Under Construction” signs that are often seen on websites), and include a description for the user of what can and cannot be done on the website, and the plan for finishing the work. This description should include a description of the refactoring process, on which much of the redesign is resting.

We understand that this work cannot be finished until the refactoring is completed; with this in mind, it is important that the schedule for finishing the interface be compatible with the refactoring effort, as well as the ongoing daily responsibility of ingesting data from the literature.

In our previous report we gave a large number of recommendations for improvements on the “classic” NED page, many of which could be incorporated right away without needing to wait for the refactoring. An immediate example is the existence of two links from that first page that are almost redundant: “Extinction calculator” and “Extinction-Law Calculators” (following one of the links from the latter, <https://www.astro.uni-bonn.de/hisurvey/profile/> takes you to a webpage that is labeled “obsolete”). Another example is the link to thesis abstracts; the most recent thesis there is from 2004, and given that modern thesis abstracts are largely available through ADS, they should not be included in NED.

The growth of NED: As we discussed in detail in our meeting, NED is facing real challenges as the data it needs to ingest grow. NED incorporates information from two main sources: the refereed literature and major astronomical surveys. The steady growth of the relevant literature is no surprise; NED is now processing 4000 journal articles per year (over 10 per day). We are excited to hear about your innovations on using machine-learning techniques to automate the ingestion of the relevant data.

The real growth is in the size of the current and next generation of surveys. There are now multiple galaxy surveys with photometric information for hundreds of millions of objects, and this will be dwarfed by the surveys of the 2020’s, including Euclid, WFIRST, and LSST, whose object catalogs

will contain tens of billions of objects. The current database refactoring that the NED team is carrying out is designed to make the system more flexible as its holdings continue to grow. This refactoring is key to NED's future success, and we are excited to see it come to fruition. We recommend that NED prepare an explicit plan or demonstration that the combination of database technology and scheduled hardware upgrades will be adequate to allow NED to ingest LSST-sized datasets a decade from now.

We were very pleased with the recent development of automatic cross-matching algorithms for NED, including careful statistics on false positives/negatives. Such efforts are clearly key for ingesting the upcoming very large datasets. The main recommendation we have on this point is the need for NED to communicate more clearly its cross-matching procedures with the wider astronomical community (see Appendix for details).

NED and the journals: A major recommendation of our previous reports has been the development of the "Best Practices" document, which explains to authors and editors of journal articles how best to present and describe their data tables to be clear and easily ingestible by NED. The document that NED has prepared is very clear and well-written, but although it is available as a link from the NED and MNRAS websites, it has not had a noticeable effect on the quality of the way data are presented in published papers. The journals are quite effective in enforcing quality standards in bibliographic references, but there seems to be little interest in doing the same for astronomical objects and their attributes in tables. Here the onus is not on NED, but on ourselves, the NUC. We will reach out to the head of the American Astronomical Society publication board to express our concerns (which go well beyond NED-specific recommendations). We hope eventually to bring NED back into the conversation: if it is possible to make public some of the software tools that NED uses to validate tables published in papers, then authors and journals could confirm the integrity of those data before they are published.

NED's outlook for the future: NED's plans for the future are well thought-out, making a clear distinction between reasonable goals over the next three years, and longer-term goals into the beginning of the next decade. We're excited that the next generation of major catalogs of galaxy photometry, including AllWISE, the 2MASS Point Source Catalog, and source catalogs from HST, XMM, and Chandra, will soon be incorporated into NED, increasing the number of objects in the catalog by an order of magnitude in just a few years. A real challenge in this work is deriving consistent photometry, given the different apertures used to measure galaxy brightnesses in different surveys. With this in mind, we recommend putting at first priority indexing NED's holdings using colors derived within a given (multi-band) survey, avoiding the problems of defining consistent colors between different surveys.

In closing, NED continues to be a crucial and key resource for the worldwide community of extragalactic astronomers. It is handling the challenges of an exponentially growing set of data, and increasingly

Princeton University **Department of Astrophysical Sciences**
Peyton Hall
Princeton, NJ 08544-1001

sophisticated demands on it by its users, with aplomb, and has a well-justified set of goals over the next years to continue its key role in supporting the research of the astronomical community. We, the NED User's Committee, are delighted to offer our support and advice to further NED's efforts.

Yours sincerely,

The NED User's Committee:

Dr. Michael Blanton, New York University
Dr. James Condon, National Radio Astronomy Observatory
Dr. Tod Lauer, National Optical Astronomy Observatories
Dr. Andrea Prestwich, Center for Astrophysics, Harvard University
Dr. Anna Sajina, Tufts University
Dr. Mark Seibert, Carnegie Observatories
Dr. Ohad Shemmer, University of North Texas
Dr. Michael Strauss, Princeton University (Chair)

Appendix: More detailed recommendations

Here we list additional recommendations and suggestions on specific issues that came up during the review.

- A particularly valuable dataset not yet incorporated into NED is the sample of redshifts from the SDSS spectroscopic survey since DR6 (the last SDSS dataset that NED has ingested). There are almost 2 million redshifts that could be included in NED quite straightforwardly, and doing so would be valuable for the astronomy community.
- We encourage NED to raise funds to support a visiting scientist program, whereby a scientist using NED for his or her research is resident with the NED team at IPAC, allowing him or her to give useful feedback on all aspects of the program.
- The User Committee does not have the technical expertise to comment on the details of the database refactoring that is underway. It is very important to get the details of this right, as it will be key for the growth of NED in the future. In this context, we recommend that NED have a formal technical review of the database design, bringing in database experts both from within the astronomical community and beyond.
- We were very impressed by the care that has been taken to rigorously match NED's holdings with new large datasets as they are incorporated. Detailed statistical analyses of the false positive and false negative matching rates have been done for the recently ingested large surveys, and are key for interpreting NED's holdings. We recommend that this information be made available to the NED user. In addition, information on associations should also be made clear, e.g., with a clarifying sentence in the documentation something like the following: "The cross-matching does not include potential associations that did not meet NED's rigorous cross match criteria. If a user is interested in obtaining all potential associations, we recommend running a cone search around the source position."
- We are enthusiastic about NED's plans to incorporate footprint information for the various large-scale surveys in its holdings. This will allow users to determine whether the absence of a measurement in NED from, say, the Sloan Digital Sky Survey means that the object is fainter than the magnitude limit of that survey, or if it simply lies outside the area of sky covered by the survey. Each survey defines its footprint using a different formalism, and it will be a challenge for the NED team to put them into a common framework.

- We remain enthusiastic about incorporating time-variability information into the NED photometric and spectroscopic holdings. Where available, the NED data should include the date on which the photometry was taken. Surveys like LSST will include hundreds of epochs of photometry for all its roughly 20 billion galaxies. Including all epochs in NED would require reproducing a substantial part of the LSST database, which we do not recommend, but including flags or statistical measurements of variability would be very appropriate.
- We encourage NED to create a program of “NED Ambassadors”, active NED users who are familiar with the publishing best practices and who know how to use NED to its full potential. If every major astronomy program around the world had such named individuals, they would be able to educate their colleagues and increase the usefulness of NED to the broader community.
- To clarify a recommendation from our previous report, we would like to be able to feed a list of references that NED produces directly to ADS, in order to use the powerful ADS filtering tools to explore that reference list further. For example, NED lists 2255 papers citing the galaxy M81; if that list could be examined within ADS, the user could find those papers by a given author, or published in a given time range, or with the highest citations, or any number of other tasks.
- NED might consider making a set of static web pages for a local volume set of nearby and/or well-known galaxies, such as all those in the Messier or NGC catalogs. That way, when search engines look for objects like M83, NED will eventually become featured prominently.