

Princeton University

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

Telephone: (609) 258-3808

Email: strauss@astro.princeton.edu

June 6, 2012

Dr. Joseph Mazzearella, NED Project Scientist
Infrared Processing and Analysis Center
Caltech/JPL, MS 100-22
770 South Wilson Avenue
Pasadena, California 91125

Dear Joe and members of the NED team,

This is an informal report of our phone meeting that we held on May 24, 2012. We wanted to give you feedback on the new initiatives that NED is taking, in response both to our recommendations from our last face-to-face meeting (January 2011) and the recommendations from the Senior Review. The approach of making the presentation material available well before the meeting, allowing us to have a real discussion rather than just being in “receive” mode, was a good one: it allowed us to get through a great deal of material in only three hours, and allowed us to give you substantial feedback. Having said this, we did rush somewhat, and we should perhaps slow down a bit next time to make sure that everyone has an opportunity to comment.

Let us start by saying that we were pleased indeed that you heard us loud and clear on two of the most important recommendations we made, namely a restructuring of the NED website, and the development of a “best practices” document to guide authors of journal articles to make their data clear, unambiguous, and straightforward to ingest into NED. We spent most of the meeting discussing these topics. In what follows, we organize our comments following the major topics we covered in the meeting, and finish by responding to each of the discussion questions you had posed for us.

The Prototype New Web Interface

We were happy with the general approach of the web interface that you are developing; this is just the sort of thing we had in mind in our recommendations from our January 2011 meeting. We spent quite a bit of time talking about the balance between the desire to make this streamlined, and the desire to make this accessible and not confusing to those familiar with “classic NED”. As a general remark, we encourage the NED team to avoid the temptation to add features to high-level interface that may have minimal or obscure use. Simplicity is a strong goal. Complexity should be added only when there is a strong benefit for doing so.

Much of our discussion was centered on the separate “Search” and “Data” tabs on the web interface. The distinction between the two is unclear; the user is conducting searches to get data under both tabs. As we understand it, the only difference is that the “Data” searches can offer a modest shortcut to some data types, but in fact most of us use the “Search” tools anyway to, e.g., find the spectrum of NGC 1068. Having two almost identical paths to the same material leads to confusion and clutter. We thus recommend eliminating the “Data” tab, leaving a clear and unambiguous path to the search functions. Of course, the NED team should confirm that every one of the current “Data” functions are indeed doable (perhaps with a few more clicks) via a search function. For those who are used to the old “Data” heading, the “Search” tab could perhaps be relabeled “Object and Data Searches”, to make it clear that the data are indeed available under that tab. Finally, the “Data” column in the old interface does have the advantage of reminding the user of the range of data available within NED. This should be communicated elsewhere on the website, e.g., a summary at the top of the page (where “News and Featured Updates” currently appears).

The front page should also convey any sense in which NED may be considered complete (e.g. all galaxies within some distance from us – as in the local volume survey, or all objects above some Galactic latitude as we’ve discussed earlier).

In all this restructuring, it is worth keeping in mind which of these options are actually *used*. For example, none of us has ever actually used the refcode search, and we suspect that it is very rarely used by the broader community. Do you track the usage of the various buttons on the current NED front page? We would be interested in seeing the statistics; it would be illuminating to know which are rarely used. For those that see a fair amount of usage, it would be valuable to identify the equivalent function in the new interface; one could imagine that clicking on a button in the old interface could bring up a little help window reminding the user how to do the equivalent with the new interface.

More broadly, it is worth thinking through the different types of tasks that users will come to NED to do. We can help you think this through, but it should be clear to the user how to proceed if they:

- need to know everything about a particular source;
- need to list all sources in a patch of sky;
- need to select sources according to particular criteria;
- need to cross correlate a list of sources with NED sources;
- need to search and filter specific catalogs (Nearby galaxies, SINGS etc).

We also reiterate our earlier suggestion that the NASA/ADS services be used to process reference lists obtained through NED searches. While NED is excellent for identifying references associated with objects or other searches, it really has no tools to cull, interpret, or analyze the list of papers further. Such capabilities are really the province of ADS. We strongly recommend that NED develop the capability to load reference lists into the ADS interface for further analysis by the user. Use of ADS to process the list might be offered as an option on the NED list, or to go even further, the NED list might be presented as an ADS page, *ab initio*.

Batch Queries by Parameter

We were happy to see this capability, and several of us tried it out. A simple query for all the quasars with redshift greater than 6 ran in a few seconds, while another query that asked for all objects within a small range (a few arcmin) in ra and dec around a given position failed after running for 1.5 hours.

We have a number of specific recommendations based on our experience:

- The query must be documented! That is, the user needs to see what the query they entered actually was. This could be given to them in SQL, and should be made available both on the status page (the one that says that the job is running) and in the final output. Also, the date and the version number of NED should also be included in this documentation; given that NED is constantly growing, the results of a query today will not be the same as that a year from now.
- It would be useful to have the option to get outputs in the form of a FITS binary table.
- There is currently no “Abort” button that allows an erroneous query to be cancelled.
- Magnitudes (both in the query form itself, and in the output tables) should be unambiguously labeled as “AB” or “Vega”. Indeed, this is also an issue with the current “by parameters” search tool within NED.
- There is currently no explicit cone search tool.
- It would be convenient to have the system e-mail the user when a query finishes. We understand that this would require an optional registration step. That’s fine; we anticipate that most regular users of NED will indeed register, and will not find this to be a burden.
- While we understand it is very difficult to estimate how long a given query might take, guidelines for the user on what sort of queries take longer than others would be quite useful.
- NED should make available the formal definitions (i.e., the filter curves) of all filters whose photometry it publishes. Compiling this information will not be trivial, of course, but it will be of great benefit for the community.

The “Best Practices for Publishing Data” Document

We were very happy to see this draft document. We discussed various ways to make this more “user-friendly”. In particular, it needs to emphasize more why it is in the *author’s* best interest to follow these guidelines (in the sense that it will lead to your paper being more useful to the community, and therefore more widely cited). In this spirit, the introduction to the Best Practices document, describing the advantages to this approach to NED, will be of less interest to most authors, and this perhaps belongs in a separate document.

To the extent that the advice given in this document is specific, perhaps tools could be written to allow checks (at least of machine-readable tables) automatically. Thus, we could imagine an on-line tool in which people upload their tables, and get back indications of what doesn’t meet specs.

We would also like to see the checklist expanded somewhat with examples. We imagine that most authors will read the detailed Best Practices document soon after it comes out, but will go over the Check List quickly as they are finishing a paper, to confirm that they are roughly in accordance with the rules. If the Check List includes examples of the most common and egregious mistakes, this will make it easier for the authors to quickly determine if their paper follows these best practices. That is, perhaps the Check List should be expanded to a “User’s Manual”, that would show examples of both good and bad practices, identify problems and solutions, and so on, while not bogging the user down with long discussions of intent, philosophy, motivations, and so on. We also think it would be valuable to “field-test” this manual, by asking some astronomers who are less familiar with NED to read it and give comments on clarity.

Ideally, these best practices should be endorsed and supported by the wider community. Few of the recommendations you give are specific to extragalactic astronomy, and we were happy to hear that this document is consistent with the best practices guidelines put together by the CDS and SIMBAD folks. The IAU does have guidelines for nomenclature, which we understood are in need of some update. An ideal situation would be to unify all these guidelines (NED, CDS, SIMBAD, IAU, etc), to clarify things for both authors of papers, and the editors of the journals. It is not NED’s sole responsibility to effect change at these high levels, but you are thinking about these things in more detail than anybody else, and we’re happy to hear that you are in communication with both the AAS Publication Board and the appropriate IAU Commissions.

We would like to see a model in which publishers remind referees to comment on how well the authors are presenting their data (and suggest reading all available best practice documents, or ideally a unified one). Then, the email that informs people that their paper has been accepted for publication should also include something like “We encourage you to consider archiving your images and spectra with at least one of NED/CDS/Simbad. This will increase the usefulness of your results to the community, and potentially increase the number of citations to your paper.” As the NUC, we are happy and eager to help advocate for this change in publication policy at the journals.

We had few comments on the specifics of the Best Practices document itself. One important detail is that photometry should always be explicit about the system on which it is reported (AB vs. Vega magnitudes). There are many ways to measure the brightness of extended objects, so quoted photometry should state unambiguously what flavor of magnitude (PSF/isophotal/Petrosian/model/aperture/Kron/etc) is used, and whether it has been corrected for Galactic or external extinction, evolution, or a K correction applied.

Merging Very Large Catalogues into NED

NED's database schema has served it well for two decades, but with the advent of new surveys with over 10^8 objects, we understand that it is time for a major overhaul of the way associations are done between "objects" (distinct astrophysical objects in the NED database) and "sources" (measurements in a catalog). NED has been very careful up until now to not include sources from a catalog until they can all be associated "unambiguously"¹ with NED objects, but this will become overwhelming as the number of very large surveys continues to grow. We therefore endorse the revamping of the schema, to make an explicit distinction in the database between objects and sources, allowing catalogs to be ingested much more speedily. Completeness in this regard is more important than reliability, especially if the appropriate caveats about the possibility of confusion are given. Tools to allow the user to know which catalogs and surveys cover any given coordinate on the sky would be very useful here. It will be a challenge to keep NED fully functional while this change is taking place in the bowels of NED, and we look forward to discussing the details of the effect of this change on users as the NED team starts to implement it.

We discussed the possibility of using crowd sourcing to help with the associations. We urge the NED team to look into this possibility. Finally, we understand that there is not support from the Senior Review for ingestion of the next generation of ground-based surveys (Pan-STARRS was given as a specific example). We do support the inclusion of such data (the user's community certainly expects this!), and we are happy to help the NED team articulate the case as to why NASA should support such an effort.

"Questions for Discussion"

You had given us a list of questions to guide our discussion. We didn't have time to go over all of them, but here are our responses to this. (We've edited your questions a bit for brevity):

1. Do you feel that NED's Work Plan for the next 3 years adequately addresses your primary recommendations from our 2011 meeting?

Response: Yes, we were quite pleased that our recommendations about the structure of the website, the need for a best practices document, and the importance of very big surveys, are driving NED in new and exciting directions.

¹ We put this word in quotes, because there are many situations in which this assignment is ambiguous at some level, even with expert human intervention.

2. Do you feel that NED captured your highest priority issues as reflected in the NASA Archive Senior Review Report and in NED's response? Do you feel that there is alignment between your priorities, NED's plans, and the directions from the NASA Senior Review?
Response: We didn't discuss the Senior Review Report at all in our meeting. They were less enthusiastic than we were in including quantification of survey selection functions. They are enthusiastic about Level5. We would like to discuss the nature and future of Level5 at our next meeting.
3. What are your impressions of NED's recent progress as highlighted in What's New and Coming Attractions?
Response: We didn't discuss this in any detail, but we were happy with what we saw, and have no specific concerns or criticisms.
4. Do the issues presented in the Web User Interface discussion capture and reflect your primary comments and concerns regarding the current NED website?
Response: We're not quite sure how this differs from the next question, but yes, as detailed above, we were very pleased that the restructuring of the web site is going in the direction we had hoped to see.
5. Does the User Interface Prototype website reflect a direction that you would like to see NED pursue?
Response: Yes, we very much like the much simpler top-level interface; we've given detailed suggestions above about further simplification.
6. Do you like the concept of collapsible elements (data tables)?
Response: We did not discuss this in the phonecon, but this concept seems fine.
7. What types of queries did you run using the prototype Long-Running By_Parameter query service (NBASQ)? Were they successful? Was the process (work-flow) clear, or did it become clearer as you used the system?
Response: We indeed tried a few queries; at least one (a position search) hung and timed out after 90 minutes. We give above a list of further recommendations for this service.
8. Do you feel that seeing early conceptual design/prototype versions of potential NED systems, services, and features as we have now on the Prototype website provides a useful vehicle for your participation in future NED developments?
Response: Yes. We are happy to work with you as you continue to develop this website, and would be open to give feedback on short timescales when you have added new features you'd like us to take a look at.
9. Do the Best Practices and Checklist Documents strike the right tone, and provide information that will be useful to you in better preparing the data in your articles for incorporation into NED?
Response: As we detail above, the emphasis has to be on why following these practices is good for the *author*, not just for NED.

10. Do you have any comments or suggestions regarding the draft NED Questionnaire for Authors?

Response: As we describe above, further examples of some of the most common errors would be very helpful to make this a useful quick guide for authors.

11. The effort to stream-line and improve speed and automation of incorporating Very Large Catalogs has the potential to be resource intensive. Do you feel that we have outlined a useful approach?

Response: We are indeed very supportive of this approach; given the increase in data volume, something like this *has* to be done.

To conclude, we continue to be impressed and amazed how much your small group is able to accomplish, and how you find inventive ways to stretch your productivity. Please continue to ask us for support and advice; we are delighted to be able to help, and are happy to express our enthusiasm to NASA headquarters or whomever else should hear it.

NED User's Committee:

Dr. Daniela Calzetti, University of Massachusetts
Dr. James Condon, National Radio Astronomy Observatory
Dr. David Hogg, New York University
Dr. Tod Lauer, National Optical Astronomy Observatories
Dr. Janice Lee, Space Telescope Science Institute
Dr. Andrea Prestwich, Center for Astrophysics, Harvard University
Dr. Anna Sajina, Tufts University
Dr. David Sanders, University of Hawaii
Dr. Michael Strauss, Princeton University (Chair)