SCHOOL OF ARTS AND SCIENCES

Department of Physics and Astronomy

Dr. Joseph Mazzarella, 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

January 7, 2020

Dear Joe, George and members of the NED team,

This is a report of the telecon held on September 26 2019 as well as the in-person meeting of the NED Users Committee (NUC) held at Caltech on December 17th 2019. The major agenda item both both meetings was the discussion of the NED proposal to be submitted as part of the Astrophysics Archives Programmatic Review 2020. We were also updated on some of the new functionalities within NED, both recently released and in development.

We were very impressed by the work of NED since our last meeting and as outlined in the NED proposal. A particular highlight was the growth in the rate of including new sources from 2 to 22 Src/s, thanks to a number of technical upgrades. This increased efficiency has allowed for the inclusion of over 700 million sources from the AllWISE catalog and led to NED surpassing 1 billion individual objects. We were also very happy to see the NED team has acted on prior recommendations of the NED Users Committee including continuing improvements in the new UI, and providing usage tutorials. We were particularly happy to see the new functionalities of overlaying survey footprints onto NED query results. A particularly key new functionality is the Gravitational Wave Follow-up (GWF) which allows for potential host galaxies of LIGO events to be identified quickly facilitating follow-up observations.

We discussed how critical for NED's continued key role in extragalactic research it is for it to plan on including the 10 billion+ galaxies expected from the LSST survey. This however involves a roughly 30X increase in the rate of incorporating new sources. This assumes keeping up with the current practices of including literature data roughly within a year and very large catalogs roughly a few years after their release. However, this significant lag is the strongest limitation in the use of NED at present. NED should therefore not only keep up with the large data volumes expected from upcoming NASA (e.g. Euclid) and complementary ground-based (e.g. LSST) surveys, but it should also strive to decrease the lag between when data are first published and when they are included in NED. Ensuring that NED holdings are both current and comprehensive is the key to the further growth of the use and impact of NED, and the success of planned future science-enabling functionality. Below we outline our specific recommendations for how to achieve these goals.

Top priority: Staying current and keeping up with upcoming very large catalogs:
— Continue with the upgrades to the database, upgrades in the hardware and continued parallelization
— Continue with automation efforts such as the machine learning classification of papers
— Need to substantially speed up including literature data by further outreach to authors as well as seeking the cooperation of arxiv.org and the astronomical Journals such that NED-ready data are prepared at the point of paper submission.
— Work directly with the upcoming large surveys (we suggest by having a NED Ambassador within each) to both understand their planned data contents and prepare cross-matches with NED ahead to their first data releases ensuring that NED can provide cross-matches close to the time of the data release
— Prioritize key data for NASA mission support such as spectroscopic redshifts (for example from surveys such as LAMOST, DESI, PFS) and photometric redshift (for example from surveys such as PanSTARRS and DES) while giving lower priority to other data such as those from surveys primarily at low Galactic latitudes.

Second priority: Improve use of NED for direct science discoveries:
— Expand recently setup efforts to disseminate information about the continuously changing NED holdings and functionality including through the use of NED ambassadors
— Further study and publicize the statistics and visualizations of the NED holdings such as present the redshift completeness as a function of redshift, magnitude and position on the sky.
— Improve support for transient events studies by growing the current GWF to make it the go-to service for probabilistic host galaxy association for any transient event (including supernovae and GRBs which are directly targeted by NASA missions). Increasing spectroscopic and photometric galaxy holdings, as well as improved completeness estimates (as described above), are key to the future success of this service.
— Enable unique time-domain studies by including the epoch of observation for galaxy photometry going forward and a flag for those sources that are known to be variable (e.g., AGN), with a link to full light curves (e.g., LSST).
— Further develop sample usage Python notebooks such as already developed for the use of APIs to query NED. Improve both NED query functionality and the ease of import of user-provided data into NED by seeking collaborations within astropy
— Improve the information about data available per object such as the availability of spectral data cubes (e.g. from surveys like MANGA). As discussed, NED shouldn't aim to store spectral data cubes, it should just provide the user with the information that said data exist, possibly a 2D preview image (e.g. a moment 0 map) and a link to where the full data cube can be viewed and downloaded.
— explore including homogenized derived quantities such as e.g. distances, diameters) to potentially include quantities such as stellar mass, star-formation rate, luminosity.

Third priority: allowing for large sample studies through new functionality and analysis close to the data:
— Incorporate machine learning tools for visualization and classification of NED objects
— Facilitate doing science closer to the data either by working with the IPAC Science Portal (ISP) if it is funded, or by facilitating NED queries from within other existing or upcoming science portals. This is particularly critical for any study that may want to involve billions of objects at once (although we recognize that small sample studies will continue to be critical even in the BigData era!)