



# Evolution of the NASA/IPAC Extragalactic Database (NED) into a data mining discovery engine

LISA VIII  
June 7, 2017  
Marion Schmitz



Caltech

NASA/IPAC Extragalactic Database

NED

3C 279 Photometry

$\log \nu_L$  (Wm<sup>-2</sup>)

$\log \nu$  (Hz)

Radio IR UV X-Ray Gamma Ray

<http://ned.ipac.caltech.edu>

The banner features a vertical title "NASA/IPAC Extragalactic Database" on the left, the "NED" logo at the top, and a central image of a galaxy with a spectral energy distribution (SED) plot overlaid. The plot shows flux density versus frequency for object 3C 279, with various energy bands labeled from Radio to Gamma Ray. The URL "http://ned.ipac.caltech.edu" is at the bottom right.

## Data fusion: methodology

- ◆ Key parameters are selected to distill, index, and fuse across the EM spectrum
- ◆ Data are reorganized into a unique and unified data model to enable panchromatic queries across all relevant articles/catalogs
- ◆ Data are provided as-published and converted to standard units, with uncertainties, when available
- ◆ Links are provided to complete content in mission archives and to the literature (via ADS)
- ◆ Derived quantities are computed
- ◆ Data from >100,000 journal articles and catalogs

## Basic Types of Data

### Photometry

- As-Published & in NED standard units [Hz, W m<sup>-2</sup> Hz<sup>-1</sup>]
- Enables SEDs

### Positions

- As-Published & Equatorial J2000

### Cross-IDs

- Normalized and exchanged with SIMBAD/IAU Dictionary

### Diameters

### Attributes

- Morphologies; activity types; etc.

### Redshifts

- As-Published & Heliocentric

- ◆ Catalogs are not duplicated nor served in their original form. Users should obtain those from Vizier or the publisher.

## Featured Holdings (May 2017)

- ◆ **301 million Multi-wavelength source XIDs**
- ◆ **253 million Distinct astrophysical objects**
- ◆ 40 million Object links to references
- ◆ **2.3 billion Photometric measurements**
- ◆ 609 million Diameters
- ◆ **11 million Redshift measurements**
- ◆ 5.2 million Objects with redshifts
- ◆ 2.5 million Images
- ◆ 578 thousand Spectra
- ◆ 502 thousand Object classifications
- ◆ 170 thousand Redshift-independent distances
- ◆ **107 thousand References**

## Discovery with NED - next few years

- ◆ Galaxy memberships in published pairs, groups, clusters (hierarchy)
- ◆ Searchable Telescope, Instrument, Filter parameters
- ◆ Information about survey coverage (footprints)
- ◆ K-corrections between observed and rest-frame photometric measurements
- ◆ Source list uploads and improvements to customized tabular output
- ◆ Complete transition to the new user interface
  - ❖ *<http://ned.ipac.caltech.edu/ui/>*
- ◆ New APIs with VO protocols:
  - ❖ *Simple Image Access*
  - ❖ *ADQL queries via Table Access Protocol (TAP)*
  - ❖ *Simple Spectral Access*

## Coordination – I

- ◆ Authors:
  - ❖ *Directly provide NED with published and peer-reviewed data ready for ingestion*
  - ❖ *Validation of published data*
  
- ◆ Librarians:
  - ❖ *Links to telescope bibliographies*
  - ❖ *Unified Astronomical Thesaurus*
  
- ◆ Publishers
  - ❖ *Links to original journal articles*
  - ❖ *Access to data tables and supplementary materials*
  - ❖ *Best Practices document*  
[http://ned.ipac.caltech.edu/ui/?q=docs/BPDP/NED\\_BPDP.pdf](http://ned.ipac.caltech.edu/ui/?q=docs/BPDP/NED_BPDP.pdf)

## Coordination – II

- ◆ ADS
  - ❖ *BIBCODE*
  
- ◆ CDS
  - ❖ *Continued working relationship*
  
- ◆ IVOA/NAVO
  - ❖ *Continued contributions and implementations*

## Summary

- ◆ NED usage is no longer dominated by researchers interactively looking up a few facts about their favorite galaxies one at a time; access is dominated by programmed queries.
- ◆ It is already a 'gold mine' for knowledge discovery based on an accumulation of panchromatic data fused from more than 100,000 catalogs and articles, growing daily.
- ◆ The team is continuously evolving the system with advances at the intersection of astronomy and informatics, in concert with NASA mission archives, ADS, and VOs.



## Successful NED Data Mining

- ◆ The discovery of a new class of super-luminous spiral galaxies based on data synthesized within NED demonstrates its power as a discovery engine - [2016ApJ...817..1090](#)

53 super spirals discovered out of a complete sample of 1616 SDSS galaxies at  $z < 0.3$  and  $L_r > 8L_{\text{sun}}$

2MASX J08542169+0449308

2MASX J16014061+2718161

2MASX J09470010+2540462



SDSS images

## NED Ambassador Program



### Ambassadors would help NED by:

- being a contact person between a major collaboration and NED
- training new students/team members about what NED is and how to use it
- discussing NED features with colleagues to ensure they use NED as it is intended
- recommending that images, spectra, data tables, etc. be supplied to NED as appropriate
- advertising new features as they arise (e.g. new user interface)
- reminding authors to acknowledge NED in new articles
- informing all authors about the Best Practices document
- encouraging suggestions and problems to be reported directly to the NED HelpDesk

### NASA/IPAC Extragalactic Database

<http://ned.ipac.caltech.edu/>



## NED is brought to you by...

