

NED Mission: Provide a comprehensive, reliable and easy-to-use synthesis of multi-wavelength data from NASA missions, published catalogs, and the refereed literature, to enhance and enable astrophysical research beyond the Milky Way.

> Rick Ebert & Joseph Mazzarella For the NED Team

Big Data Task Force NASA, Ames Research Center 2016 September 28-30

2016 California Institute of Technology







RADIO

NED Overview

VISIBLE

NED is more than an archive, it is a comprehensive and current census of the universe.

NED plays a key role in maximizing the science return from NASA missions by synthesizing NASA data with complementary data across the spectrum.

Distilling, vetting, and addition of derived information simplifies and accelerates research.

Keep current by continuously integrating new information from literature, surveys and mission archives.

MICROWAVE

INFRARED



ULTRAVIOLET

X-RAY

GAMMA RA



Improving NED with Technology

Our primary task is to run a robust operational system with continuous, distributed client connections, applying new technology where appropriate.

User Interface

- Using web application framework/ CMS Drupal
 - Leverages extensive base of community software and expertise
- NED Form Engine builds data driven user interface forms
- Simple Search for common queries

Machine Learning

- Initial objective:
- Classify e-journal articles for suitability to NED
- <u>Challenges</u>:
- Data are presented differently
- Capturing expertise
- <u>Current</u> focus:
- Evaluate various approaches
- Rule-base training
- <u>Future</u>:
- apply Machine Learning to data prospecting by NED users
- literature data extraction & object classification



Database

- <u>Ongoing:</u> refactoring and extensions to improve scalability, extensibility, and enable data prospecting, classification & mining
- <u>Future</u>: Leverage extensions for spatial indexing



3





Big Data Implications

Volume

- Over 2016-2019, NED is growing by > 10x to support NASA science with data joined for > 1 billion objects with > 10 billion attributes
- <u>Response</u>: Refactoring database schema to improve scalability, extensibility, and data driven design

Velocity

- Rising data rate from sky surveys & literature
- Keeping current is essential
- <u>Response</u>: Parallel processing of cross-IDs, including new pipeline, and small pilot project in machine learning

Variety

- 22+ NASA mission data sets
- γ rays through radio frequencies
- wide variety of data-types
- 100k+ articles & catalogs each with unique data presentation
- <u>Response</u>: advancing data capture & fusion techniques

Caltech



Veracity

Peer-reviewed literature is a primary data source. Data often not refereed to the same extent results. <u>Response</u>: Scientific expertise to make big data better

- Vetting, communicating with authors on issues
- Published document
 - "Best Practices for Data Publication"
 - $\checkmark\,$ by the NED Team for journals' Instructions for Authors





Science with NED-Today

Discovery of a new class of super-luminous spiral galaxies

This previously unknown class challenges theories of galaxy formation and evolution.

The discovery is based entirely on data joined in NED.

NASA press release 2016 March 17 Patrick Ogle et al. 2016, ApJ, 817, 109

NAC Big Data Task Force



"The Event Horizon of M87" by Avery E. Broderick et al. 2015, ApJ, 805, 179

"This work would not have been practical without extensive use of the NASA/IPAC Extragalactic Database (NED)."

NED is cited in 20% of Astronomical Telegrams published Jan-Jun 2016

Used extensively at the telescope. E.g., identification of host galaxies of newly discovered supernovae

2016-09-29

Caltech

5



Science with NED-Future



Enabling analysis of Spectral Energy Distributions with broad coverage for millions of galaxies. Vision: Evolve the database and user interfaces to facilitate statistical studies of panchromatic data fused for billions of galaxies and AGN observed across cosmic time

Enable powerful data prospecting queries with information joined across missions

- Build a Spectral Energy Distribution of WFIRST galaxies
- Find published galaxy clusters with > 30 members
- Find galaxies with Luminosity $> 3.2*10^{11} \, [L_{\odot}]$
- Find objects with multi-mission flux ratio f(GALEX FUV)/f(Spitzer 25 um) > 1E-04





Develop visualizations of content and completeness to support statistical studies, and better framed queries.

NAC Big Data Task Force

6





NED Literature citations

A record 730 peer-reviewed articles and 264 telegrams acknowledged using NED in 2015

Comparable to a NASA *Great Observatory*



2016 NASA Group Achievement Award

"for outstanding contributions to NASA's strategic goals in astrophysics and substantial impact on extragalactic research measured by citations in the peer-reviewed literature."

NAC Big Data Task Force

Every day NED serves over 70,000 gueries and is acknowledged by 2 peer-reviewed articles.

Usage

Trend of incorporating NED access in tools, services, and automated processes is reflected in usage.

26 million requests in 2015



2016-09-29

7



Question 1: Planning

What are the processes for planning for future (5-10 years) capabilities of your service?

How and from whom do you gather input for this planning process and where does input typically come from?

NASA Astrophysics Archive Senior Review Process Assessment of relevance to NASA Astrophysics Roadmap Recommendations from NED Users Committee (NUC)

User Surveys and Outreach at science conferences.

What new feature(s) do you really want to implement?

Statistical algorithms and Machine Learning analysis integrated and applied directly to the content of NED

Multi-dimensional exploration tools and interfaces

NAC Big Data Task Force

8





Question 2: Inertia

What feature(s) of your service would you like to stop performing?

Bibliographic searches by author name and serving journal abstracts

How do you gather input for making such decisions and where does input typically come from?

Consultation with our User Committee, User Surveys, Science outreach

NASA Astrophysics Roadmap

What is preventing you from stopping?

This is in progress, and requires:

Extending current interoperability between NED and Astrophysics Data System (ADS)

- NED providing additional metadata about journal article data to ADS
- ADS providing interface filters specific to extragalactic data tagged/categorized by NED

NAC Big Data Task Force

9





Question 3: Interoperability (1)

What steps you are taking to make your data interoperable with allied data sets from other data sites in and out of NASA?

NED provides an interoperability component used by other archives and tools.

- Object information and cross-identifications
- ▶ Name resolution
- Links in results to the originating and other allied archive data sets.

We maintain this by

- Participation in NASA Astrophysics Data-Center Executive Committee (ADEC)
- Participation in NASA Astronomy Virtual Observatory (NAVO)
- Use of IVOA query and data exchange recommendations where applicable
- Streamlining connectivity with open, community data analysis tools
- Development of NED specific APIs: Name Resolver; ObjectLookup; JSON

NAC Big Data Task Force

10





Question 3: Interoperability (2)

How do you find allied data sets and what criteria make data sets candidates for enabling interoperability?

Data sets are selected by scanning e-journal articles and monitoring NASA mission data product publication. Criteria:

- Contains published and peer reviewed information
- Contains newly identified extragalactic objects
- New or improved measured attributes for objects already in NED
- Contains well formed data and metadata

NAC Big Data Task Force

11







NED provides

- diverse and rich data synthesis on a large scale
- global connections to the astronomy community
- key interoperability services to other data centers
- tools and data to accelerate extragalactic research

NED is growing more than 10-fold in the next few years to facilitate panchromatic research on billions of objects.

NED is continuously seeking improvements for our next generation services and the technology to achieve them.

Data are continually integrated from the literature, mission archives, and surveys ...



