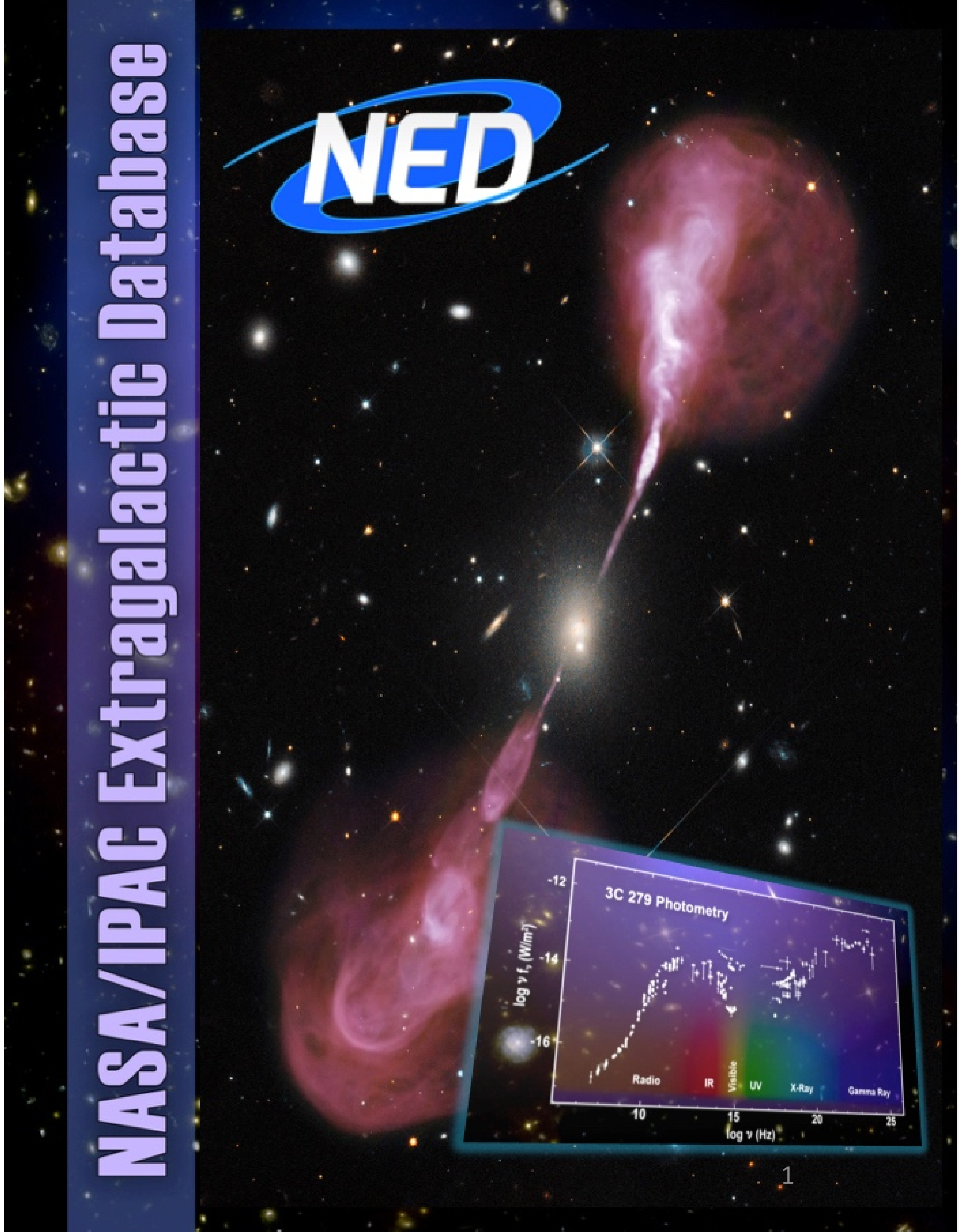
# Science Discovery with Diverse Multi-wavelength Data Fused in NED

Joseph Mazzarella
Caltech, IPAC/NED

*NED Team: Ben Chan, Tracy Chen, Cren Frayer, George Helou, Scott Terek, Rick Ebert, Tak Lo, Barry Madore, Joe Mazzarella, Olga Pevunova, Marion Schmitz, Ian Steer, Cindy Wang, Xiuqin Wu*

7 Oct 2019
ADASS XXIX
Groningen

NASA ipac Caltech



# Intro & outline

Astro data are growing at an unprecedented rate.

Ongoing expansion in volume, velocity and variety of data is creating great challenges, and exciting opportunities for discoveries from federated data.

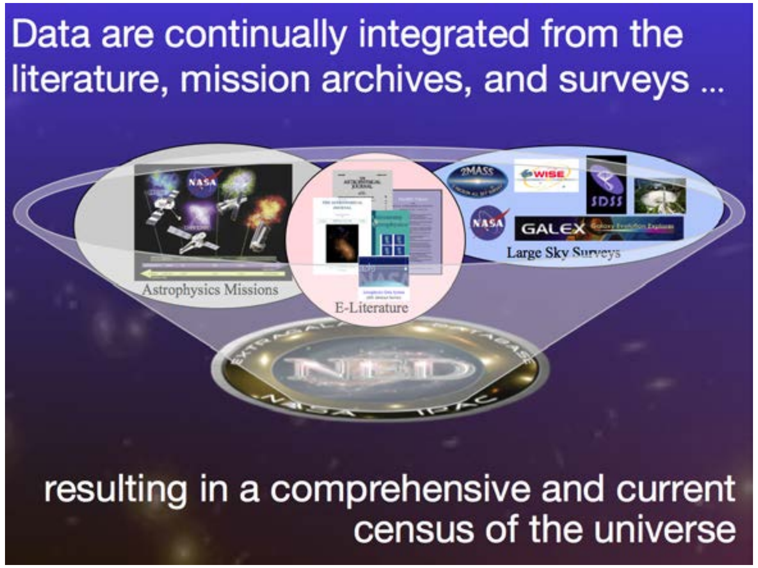
Advances in joining data across the spectrum from large sky surveys with > 100,000 smaller catalogs and journal articles, combined with new capabilities of the user interface, are helping astronomers make new discoveries directly from NED.

This will be demonstrated with a tour of exciting scientific results recently enabled or facilitated by NED.

Some challenges and limitations in joining heterogeneous datasets

Outlook for the future

# Overview: NED is maintaining a panchromatic census of the extragalactic Universe




***NED contains:***

- *773 million multiwavelength cross-IDs*
- *667 million distinct objects*
- *4.9* ***billion*** *photometric data points*
- *47 million object links to references*
- *7.9 million objects with redshifts*
- *2.5 million images*
- *And more …   +ALLWISE in October*

***NED is:***

- *A synthesis of multi-wavelength data*
- *Published data augmented with derived physical attributes*
- *Linked to literature and mission archives*
- *Accessible via VO protocols*
- *Easy-to-use*
- *Comprehensive*
- *Growing rapidly*

*NED simplifies and accelerates scientific research on extragalactic objects by distilling and synthesizing data across the spectrum, and providing value-added derived quantities and functionality.*

Published:

- Names
- $(\alpha,\delta)$
- Redshifts
- $D_{Mpc}$
- Fluxes
- Sizes
- Attributes
- References
- Notes

Contributed:

- Images
- Spectra

*Derived:*

- *Distances*
- *Metric sizes*
- *SEDs*
- *Luminosities*
- $A_\lambda$
- *Velocity corrections*
- *Cosmological corrections*


# Overview: Powerful science queries

**NED…**

Enables powerful science queries such as…

*What is the spectral energy distribution, spanning gamma ray energies through radio frequencies, for the quasar 3C 279?*

*What gravitational lenses are known surrounding galaxy cluster Abell 68?*

*Which objects have published redshifts z > 3.0?*

*What is the most precise redshift-independent distance measurement to M82?*

*Which spiral galaxies contain Type 1 AGNs?*

*What are the best candidate galaxies for hosting the latest LIGO/Virgo GW event?*

*NED has become an essential tool for many researchers and is woven into the fabric of modern astrophysics research.*

# Integration of diverse multi-λ data: process

Routinely joining data from two domains: astrophysics literature and large survey catalogs

*Very large catalogs (VLCs) with > 100 million sources are essential for studies of galaxy evolution, dark matter, and the cosmic web*

Catalogs are not ingested or served in original form.

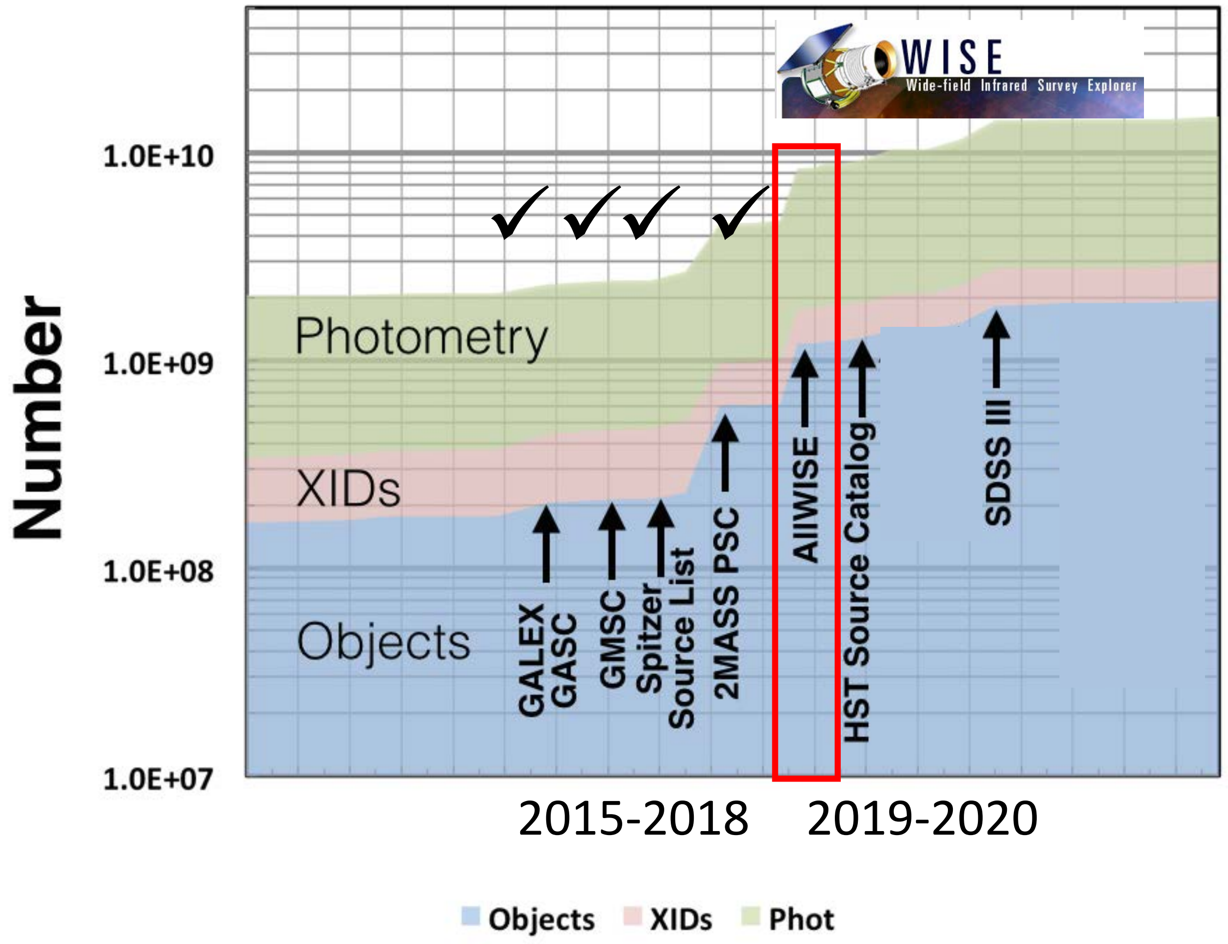
Selected data are reorganized and fused in a unified data model to enable panchromatic queries across all relevant articles and catalogs.

Probabilistic cross-matching process: *MatchEx*

Source density ➔ background contamination

Poisson ➔ balance completeness vs. reliability

***Fusion of GALEX, Spitzer Source List, and 2MASS PSC has vastly expanded NED the past few years.***




*In 2017/18, NED engineers achieved a 5-fold increase in database ingest speed through parallelization and other optimizations, enabling efficient integration of VLCs.*

# Integration of diverse multi-λ data: example public updates

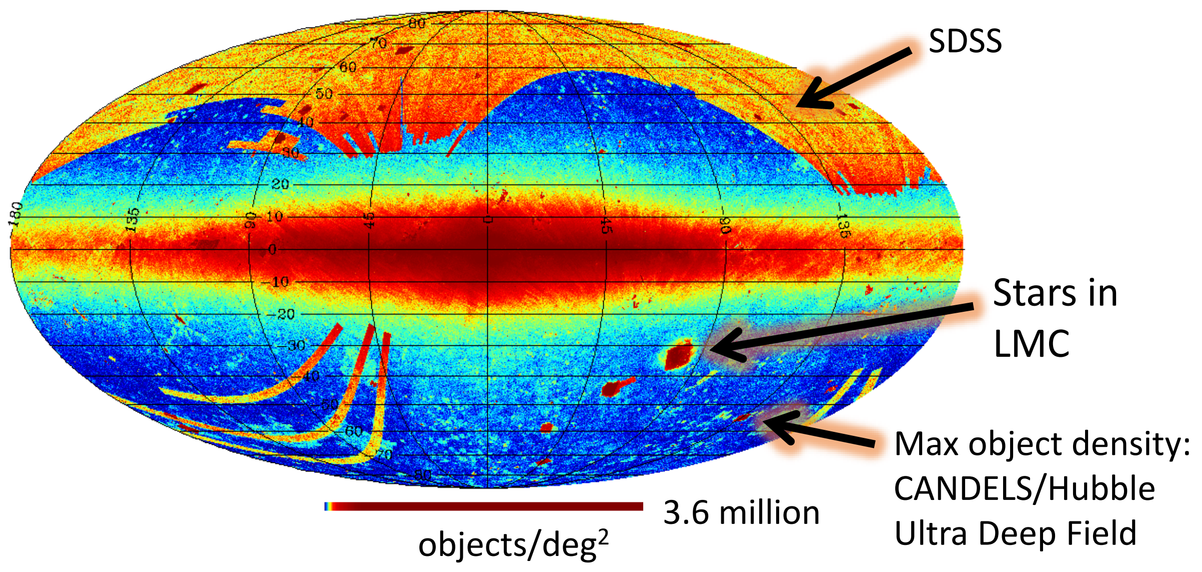
## July 2018 release:

***Results of cross-matching the entire all-sky 2MASS Point Source Catalog and fusing key data in NED.***

- 57M of 471M sources (12%) matched with prior NED objects, and their J, H, and K flux densities have been merged into spectral energy distributions.
- NED content more than doubled with this addition.

***+ Results of processing 47K sources from 3,248 references***

- 1.2M object links to 3,248 new references
- Over 1,000 spectra and 84 images
- 13 new articles in Level 5 Knowledgebase




## Dec 2018 release:

***Results of processing 842K sources from 1,663 references***

- 842K new object links to 1,663 references
- 677K new NED objects & 165K XIDs w/ prior NED objects
- 707K additional redshifts; 184K new objects w/ redshifts
- 267,819 spectra from Sloan Digital Sky Survey
- 7 new review articles in Level 5

## May 2019 release:

***Results of processing 1.6 million sources from 1,297 refs***

- 1.6 million new object links to 1,297 references
- 438K galaxies from Galaxy and Mass Assembly (GAMA, Davies, L. J. M. et al. 2015MNRAS.447.1014D)
- 293 HI images from The Local Volume HI Survey (LVHIS) (Koribalski et al. 2018MNRAS.478.1611K )
- 7 new review articles in Level 5

# ALLWISE Source Catalog - Overview

The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance (Wright+ 2010, AJ, 140, 1868)

All-sky survey in 4 infrared bands:
3.4, 4.6, 12, and 22 µm (W1, W2, W3, W4)

Angular resolution: 6.1, 6.4, 6.5, 12" FWHM

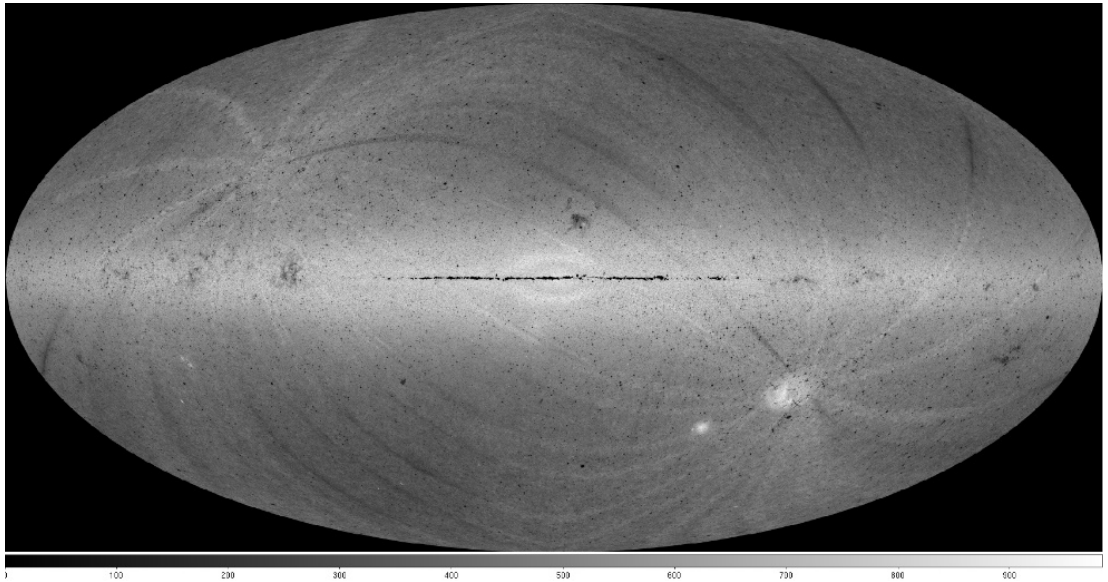
Astrometric precision: 0.15" (high S/N sources)

# sources: 747 million

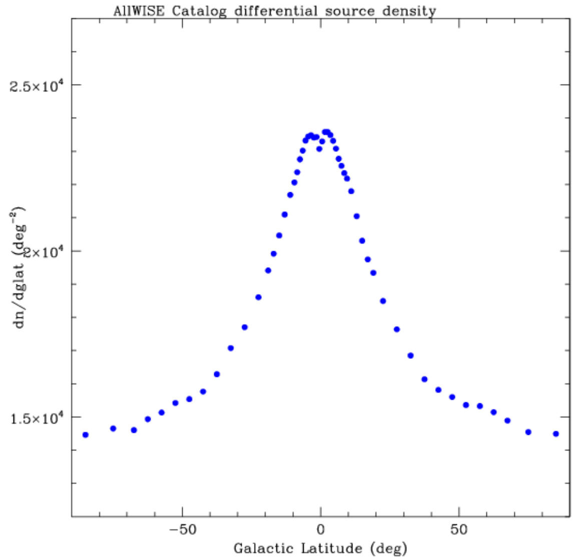
*Why do we process all sources in a VLC?*

- *Statistical cuts would omit some extragalactic objects*
- *Minimizes false matches in cross-identifications*
- *Provides SEDs spanning multiple surveys to enable classifications within NED: star-galaxy, galaxy type, AGN, etc.*
- *Supports discovery of galaxies in the Zone of Avoidance*
- *Recommended by NED Users Committee*

***NOTE: Stars and unclassified sources (e.g., IrS) can be omitted with filters on object type in the UI or via TAP.***



Source count map in Galactic coordinates




From Explanatory Supplement to the AllWISE Data Release Products (Cutri et al.)

# Data Integration: ALLWISE in NED

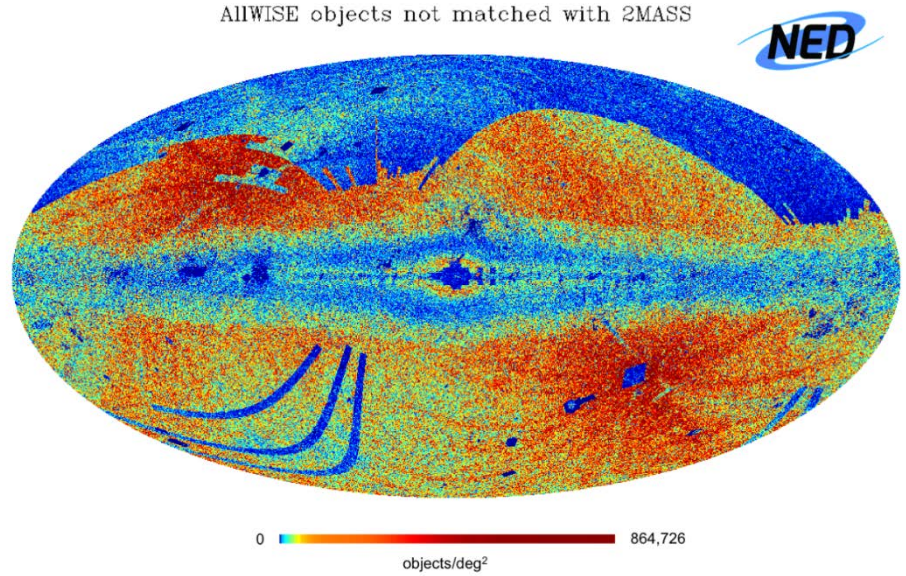
## October 2019 database release

***Results of cross-matching and fusing key data from 747 million entries in the all-sky ALLWISE Source Catalog – the largest catalog folded into NED to date.***

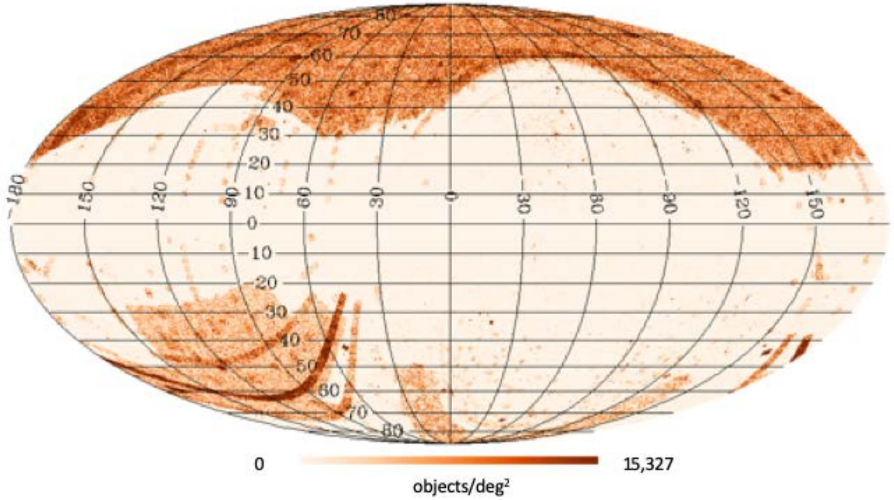
***Enabled by a ~2.5x increase in data ingest speed.***

- 307M sources (41%) matched with prior NED objects
- W1-W4 band flux densities merged into SEDs
- NED surpasses 1 billion distinct objects with XIDs
- NED now contains ~14 billion photometric measurements

- ***+ data for ~1 million objects from literature***




*AllWISE objects in NED that were not matched to 2MASS. Most are in the ecliptic pole areas where the AllWISE sensitivity is highest.*



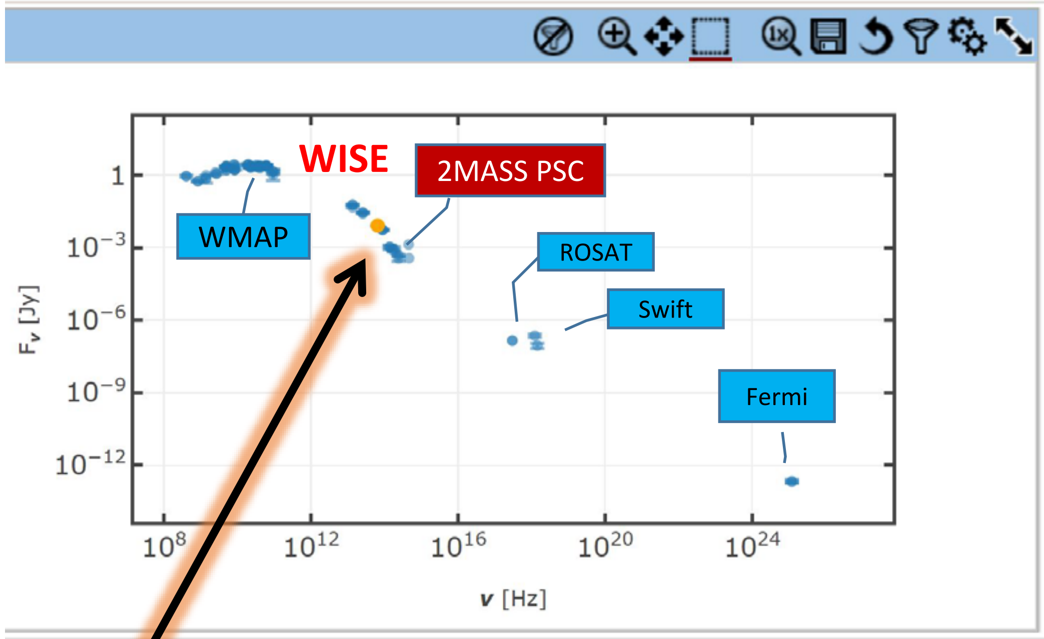

*Cross-IDs established in NED resulted in attaching redshifts to 4.2 million ALLWISE objects*

# Validation of XIDs using SEDs

QSO, Flat-Spectrum Radio Source
PKS 1057-79 = 2MASS J10584332-8003540
= WISEA J105843.25-800354.1
z = 0.58100

WISE W1,W2,W3,W4 phot merged with data across the spectrum

This example is shown directly from the NED UI (Production)




**Photometry for PKS 1057-79**

1 of 1 (1 - 77 of 77)

| No. | Observed Passband | Photometry Measurement | Uncertainty | Units | Frequency (Hz) | Flux Density (Jy) | imit of unc | imit of unc | mit of Flux | mit of Flux | NED Uncertainty | NED Units | Refcode | Significance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | = 'Jy' | | |
| 12 | H (2MASS/CTIO) | 1.530000e+01 | +/-0.112 | mag | 1.80E+14 | 7.480000e-04 | 7.7E-05 | 7.7E-05 | | | +/-7.72E-05 | Jy | Ref ✓ | uncertainty |
| 13 | K_s (2MASS/CTIO) | 1.450000e+01 | +/-0.076 | mag | 1.39E+14 | 1.080000e-03 | 7.6E-05 | 7.6E-05 | | | +/-7.56E-05 | Jy | Ref ✓ | uncertainty |
| 14 | K_s (2MASS/CTIO) | 1.450000e+01 | +/-0.217 | mag | 1.39E+14 | 1.040000e-03 | 2.1E-04 | 2.1E-04 | | | +/-2.08E-04 | Jy | Ref ✓ | uncertainty |
| 15 | W1 (WISE) | 1.190000e+01 | +/-0.023 | mag | 8.94E+13 | 5.490000e-03 | 1.2E-04 | 1.2E-04 | | | +/-1.16E-04 | Jy | Ref ✓ | uncertainty |
| 16 | W1 (WISE) | 1.190000e+01 | +/-0.007 | mag | 8.94E+13 | 5.460000e-03 | 3.5E-05 | 3.5E-05 | | | +/-3.52E-05 | Jy | Ref ✓ | uncertainty |
| 17 | W1 (WISE) | 1.190000e+01 | +/-0.011 | mag | 8.94E+13 | 5.290000e-03 | 5.4E-05 | 5.4E-05 | | | +/-5.36E-05 | Jy | Ref ✓ | uncertainty |
| 18 | W2 (WISE) | 1.080000e+01 | +/-0.021 | mag | 6.51E+13 | 8.180000e-03 | 1.6E-04 | 1.6E-04 | | | +/-1.58E-04 | Jy | Ref ✓ | uncertainty |
| 19 | W2 (WISE) | 1.080000e+01 | +/-0.006 | mag | 6.51E+13 | 8.280000e-03 | 4.6E-05 | 4.6E-05 | | | +/-4.57E-05 | Jy | Ref ✓ | uncertainty |
| 20 | W2 (WISE) | 1.080000e+01 | +/-0.008 | mag | 6.51E+13 | 7.950000e-03 | 5.9E-05 | 5.9E-05 | | | +/-5.86E-05 | Jy | Ref ✓ | uncertainty |
| 21 | W3 (WISE) | 7.620000e+00 | +/-0.017 | mag | 2.59E+13 | 2.830000e-02 | 4.4E-04 | 4.4E-04 | | | +/-4.42E-04 | Jy | Ref ✓ | uncertainty |
| 22 | W3 (WISE) | 7.610000e+00 | +/-0.007 | mag | 2.59E+13 | 2.860000e-02 | 1.8E-04 | 1.8E-04 | | | +/-1.84E-04 | Jy | Ref ✓ | uncertainty |

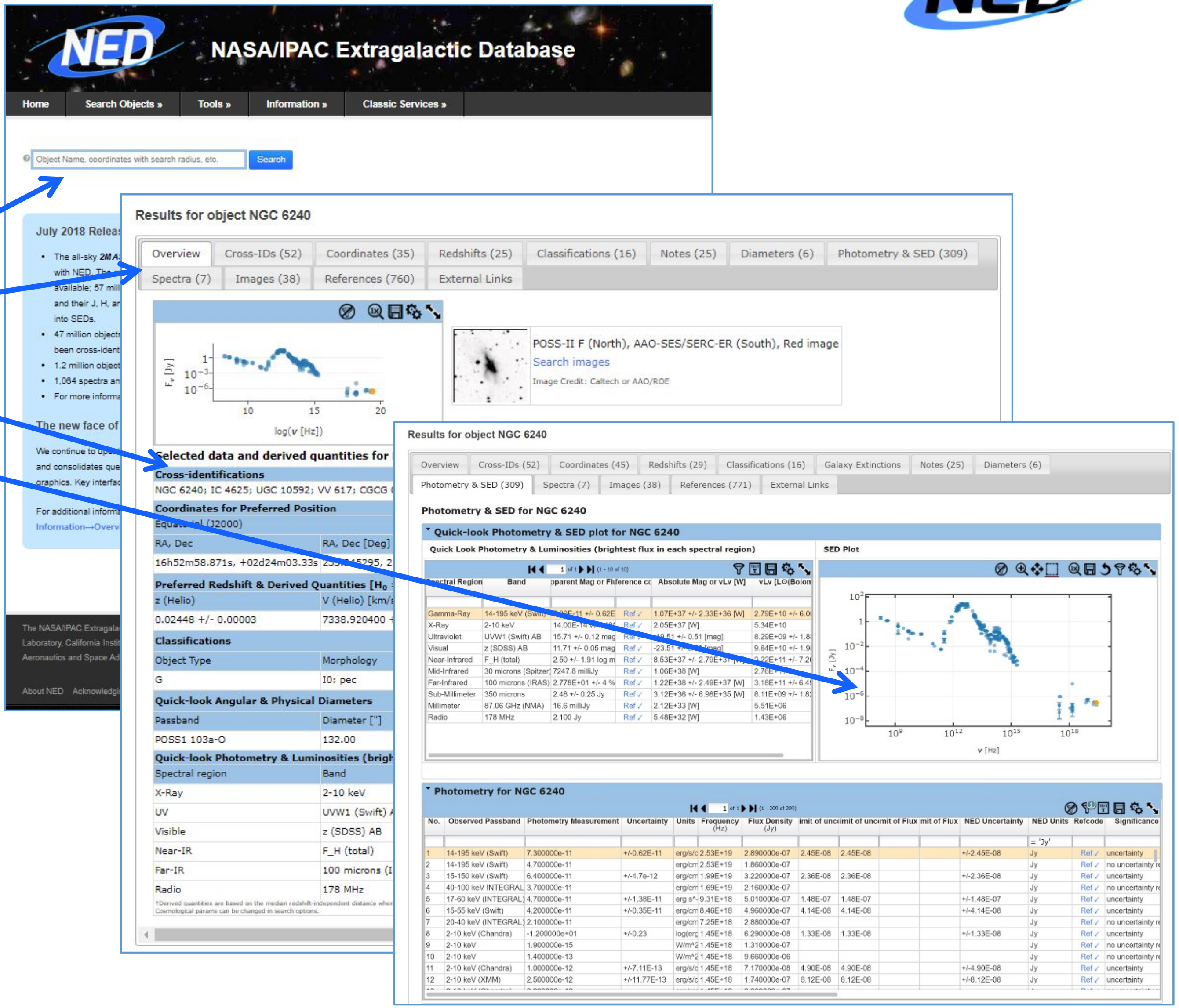
# New user interface

## Public release in June 2018

- Single input field supports common queries: object name, object name or coordinates w/ search radius, objects in reference, images
- Detailed data presented in tabs
- Concise Overview of key object properties
- Linked interactive visualizations via Firefly

## Updates in Nov & Dec 2018

- Near name, near coordinates & IAU style cone searches consolidated into single form
- Galactic Extinctions tab
- Multi-object query results with Firefly
- Pre-selected cosmological parameters updated



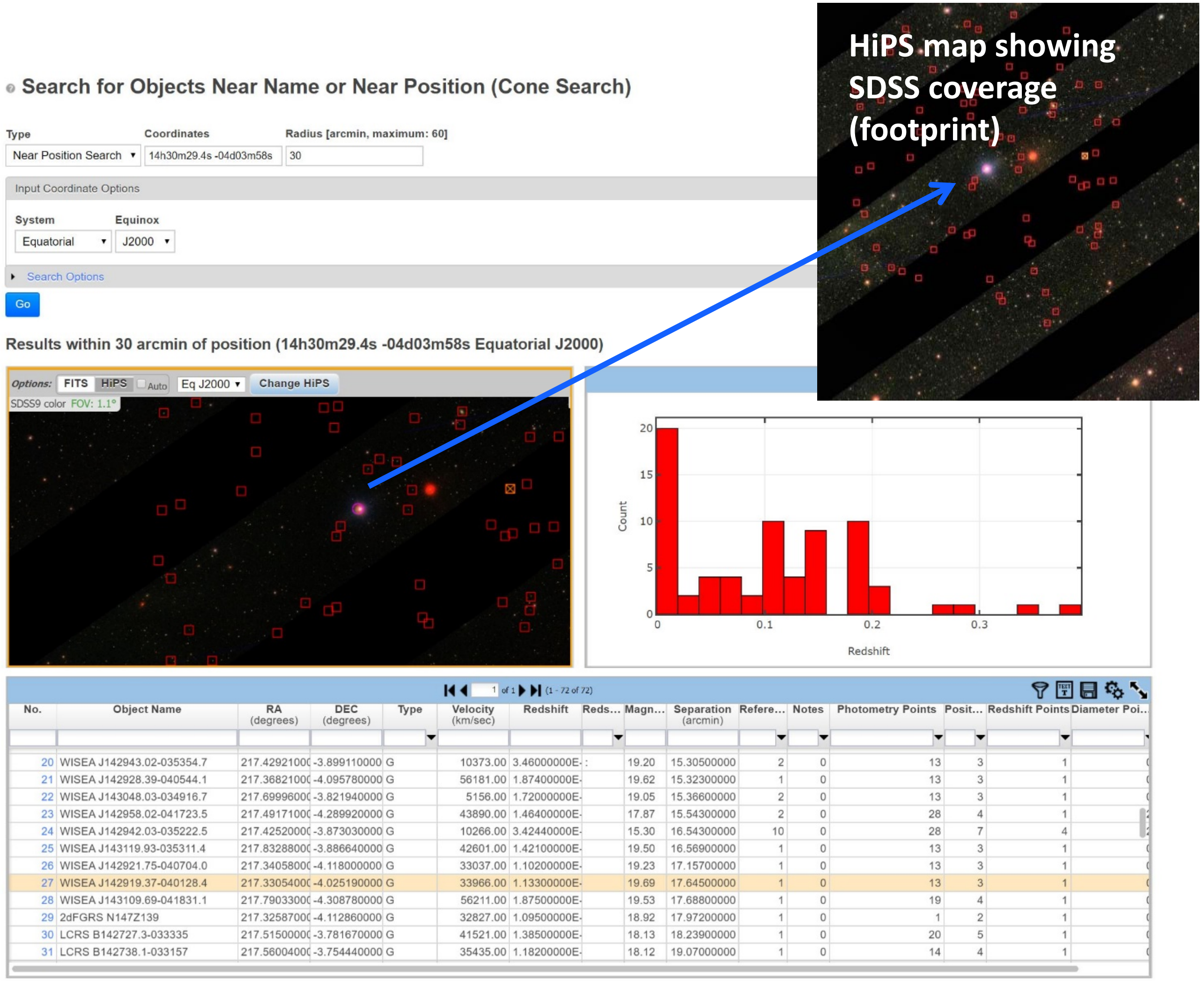
# UI (2)

## March 2019 UI release

- New presentation of bibliographic info
- Improved display of RA, Dec, Separation and other fields in object search results
- Improved input validation

## August 2019 UI release

- Full Refcode now displayed for links to ADS
- Multi-object search results now include object markers overlaid on sky image, a histogram of redshifts, and an interactive table.
- ByParameter search results include a link to preview VOTable results using Firefly
- Numeric values in interactive tables are now right-justified for better legibility
- Some known issues resolved

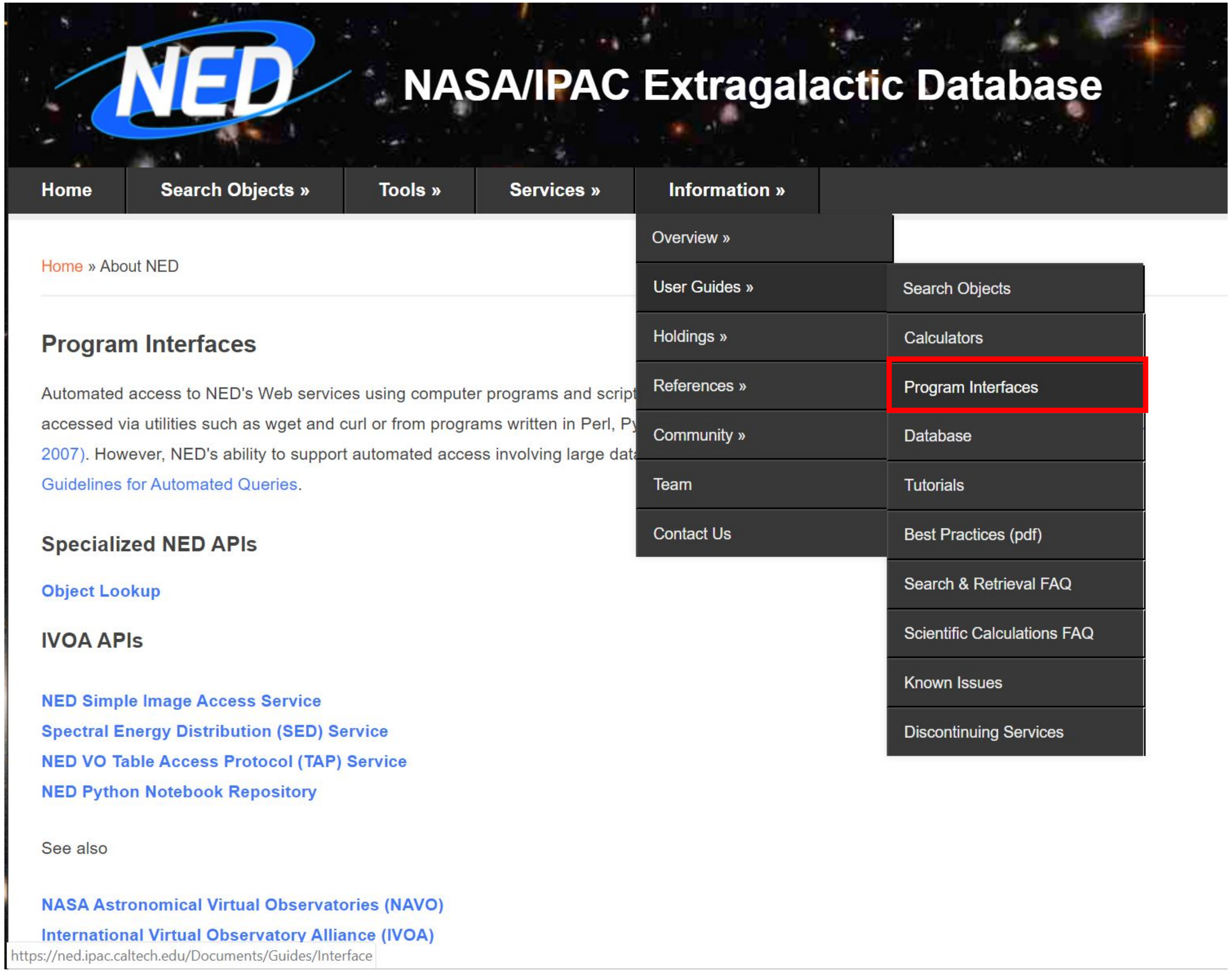



*Hierarchical Progressive Survey (HiPS) maps available in NED via Firefly allow users to visualize object locations with respect to survey coverage (footprints).*

# APIs with VO Standards

## VO Table Access Protocol (TAP) service

- Supports queries in Astronomical Data Query Language (ADQL)
- Currently, the service allows users to query the NED Object Directory ("objdir") from computer programs and scripts without the need to go through a browser.
- Ex: What galaxies in this sky region (cone, box, polygon) have no available redshift?
- ***Released in June 2018***
- ***Enhanced in August 2019***




## Jupyter Notebooks

- Illustrate access to NED via APIs
- Used in workshops at recent AAS meetings, teaching how to access NASA archives via Python

## DataLink

- In progress

# Science discoveries directly from NED's data synthesis

## *Super spiral galaxies*

*Ogle et al. 2016, ApJ, 817, 109*

SUPERLUMINOUS SPIRAL GALAXIES

PATRICK M. OGLE[1], LAURANNE LANZ[1], CYRIL NADER[1,2], AND GEORGE HELOU[1]
[1] IPAC, California Institute of Technology, Mail Code 220-6, Pasadena, CA 91125, USA; ogle@ipac.caltech.edu
[2] University of California, Los Angeles, California
*Received 2015 October 30; accepted 2015 December 17; published 2016 January 26*

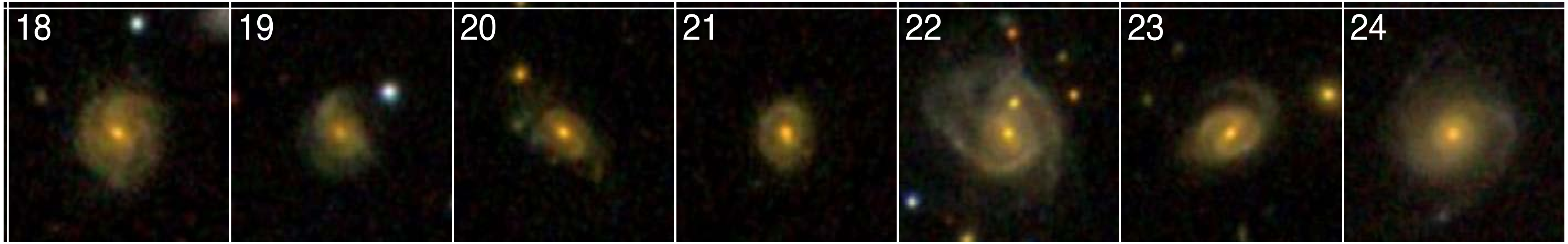
***Important new class discovered while characterizing completeness of redshift measurements in NED***

*Properties:*

$L_r > 8\ L^*$ — d = 60 -130 kpc

$M^* = 30 - 300$ billion $M_\odot$ — SFR = 5 - 65 $M_\odot$/yr

*Data joined within NED that enabled this work include redshifts, object types, diameters, SEDs and derived quantities: luminosity (SDSS), stellar mass (2MASS), star formation rate (GALEX, Herschel).*



# Science discoveries directly from NED's data synthesis

## *Super spiral galaxies*

***Fun fact: the most massive super spiral galaxies have 20 times more mass in stars and gas than our Milky Way galaxy.***

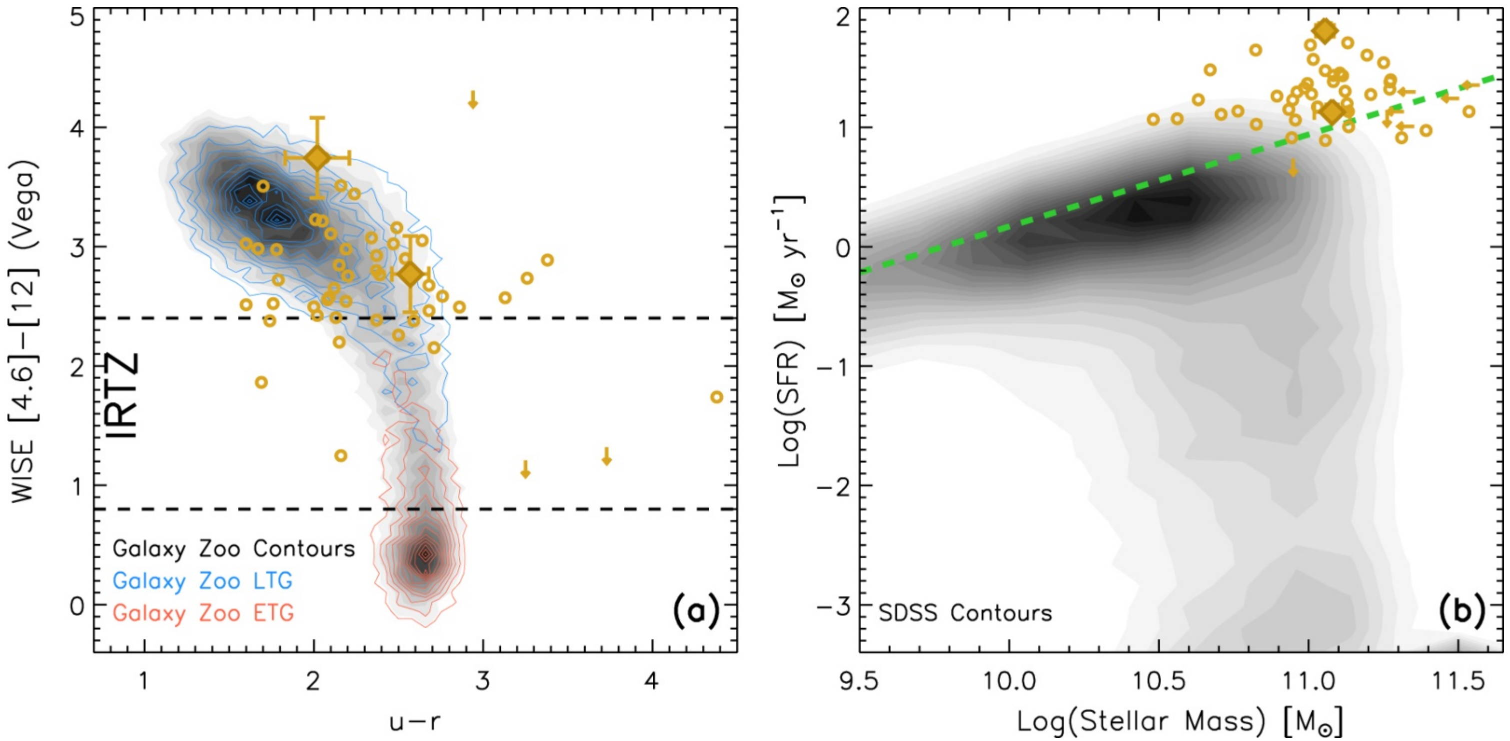



Figure 1 from Ogle et al. (2016).

(a) SDSS and WISE colors of super spirals (circles) compared to SDSS galaxies classified as late type galaxy (LTG) or early type galaxy (ETG) by Lintott et al. (2008). The infrared transition zone (IRTZ) is the mid-IR equivalent of the optical green valley (Alatalo et al. 2014).

(b) Star formation rates and stellar masses of super spirals compared to the SDSS-WISE sample of Chang et al. (2015). The dashed line indicates the star-forming main sequence at z  0 (Elbaz et al. 2007).

# Super spiral galaxies – follow up paper 1

A CATALOG OF THE MOST OPTICALLY LUMINOUS GALAXIES AT $Z < 0.3$:
SUPER SPIRALS, SUPER LENTICULARS, SUPER POST-MERGERS, AND GIANT ELLIPTICALS

Patrick M. Ogle[1], Lauranne Lanz[2], Philip N. Appleton[3], George Helou[3], Joseph Mazzarella[3]

2019, ApJS, 243, 14

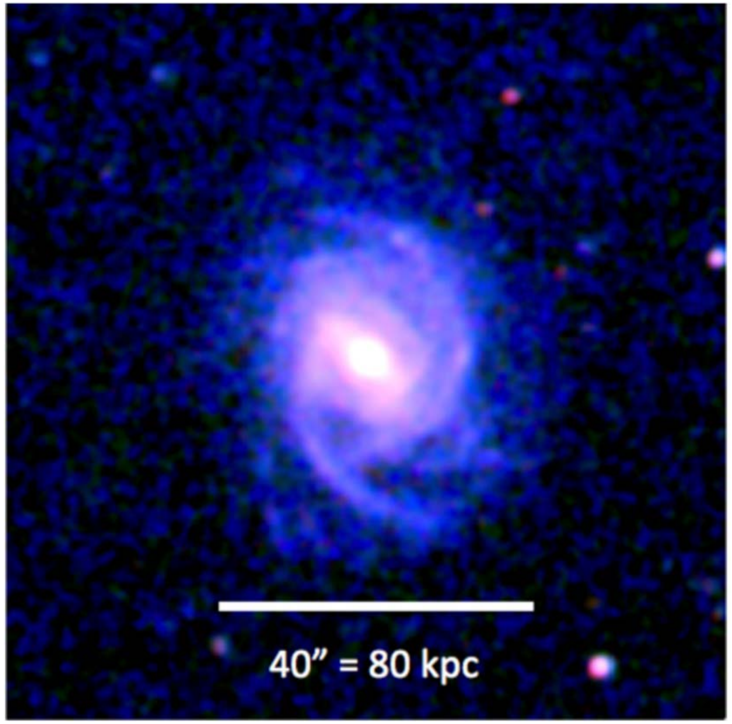
1525 most optically luminous galaxies from the Sloan Digital Sky Survey with r-band luminosity L r > 8L* and redshift z < 0.3

- **84 super spirals ➔ very rare at z < 0.3**
- 15 super lenticulars
- 14 super post-merger galaxies
- 1400 giant ellipticals

***With Mstars of $10^{11.3} – 10^{12}$ M$_\odot$ super spirals and lenticulars are the most massive disk galaxies currently known.***

***They follow* star-forming main sequence, with <SFR> ≈ *30* M$_\odot$/yr**

***Their disks are red on the inside and blue on the outside, consistent with inside-out growth.***

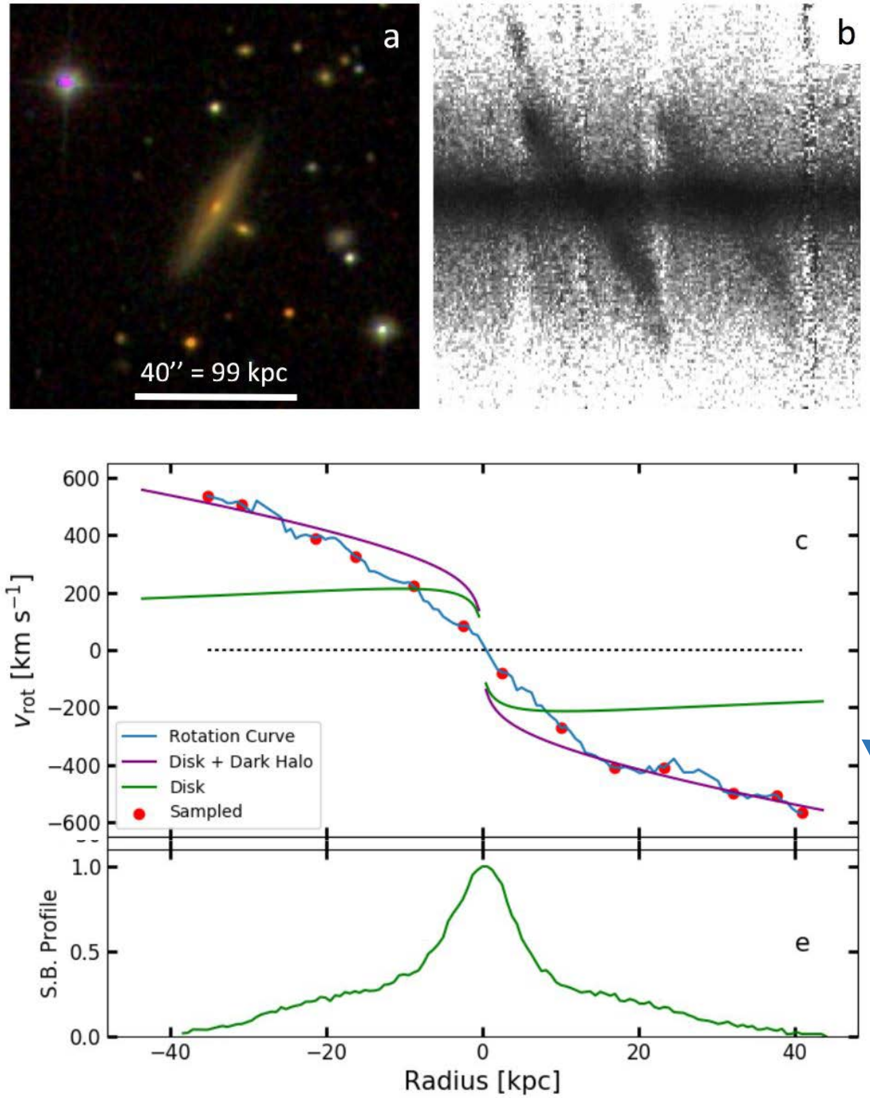



OGC 0543, SDSS gri image

# Super spiral galaxies – follow up paper 2

A BREAK IN SPIRAL GALAXY SCALING RELATIONS AT THE UPPER LIMIT OF GALAXY MASS

PATRICK M. OGLE[1], THOMAS JARRETT[2], LAURANNE LANZ[3], MICHELLE CLUVER[4,5], KATHERINE ALATALO[1], PHILIP N. APPLETON[6], JOSEPH M. MAZZARELLA[6]




2019 ApJL, in press
https://arxiv.org/abs/1909.09080

Rotation curves were measured for 23 super spiral galaxies

*They all rotate very fast, 240 – 570 km/s*

This is the fastest rotator found so far
2MFGC 12344 = 2MASX J15154614+0235564
= SDSS J151546.10+023556.5

# Super spiral galaxies – follow up paper 2

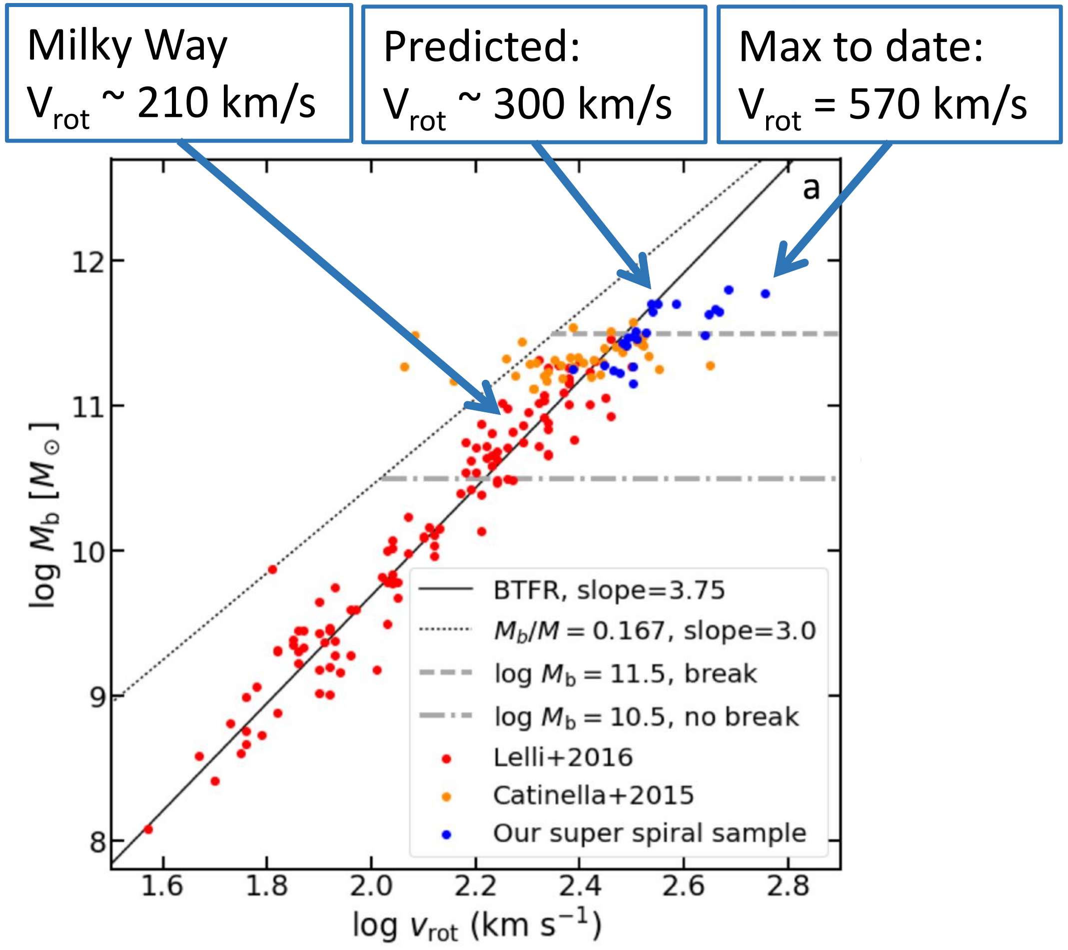
**The baryonic Tully-Fisher relation tells us that a spiral galaxy of given mass (stars+gas) rotates at a certain speed**

**Super spirals break this relation**

- Many rotate up to ~2x faster than expected
- They have low $M_b = M_{stars} + M_{gas}$ for their $V_{rot}$
- There appears to be a mass limit of spiral galaxies with $M_b \approx 6.3 \times 10^{11}\ M_{\odot}$ ($\log M_b/M_{\odot} = 11.8$).
- Agrees with theoretical prediction of the maximum dark matter halo mass where gas can cool and collapse within a dynamical time (White & Rees 1978)

*Super spirals' large rotation velocities indicate they reside in very massive dark matter halos, up to 3 trillion* $M_{\odot}$

***The deficiency in stellar mass wrt their halo mass suggests their huge DM halos inhibit star formation by accreting gas too fast to permit sufficient cooling to allow collapse, or their rapid spin rates make it harder for gas clouds to collapse.***




*Super spirals rotate faster than expected from the TF relation between Mass (stars plus gas) and Vrot for lower mass galaxies.*

# Science discoveries directly from NED’s data synthesis

NED featured in science news

*“**Researchers examined images of galaxies in the NASA/IPAC Extragalactic Database*** […] of different wavelengths of light, representing stars of different ages.

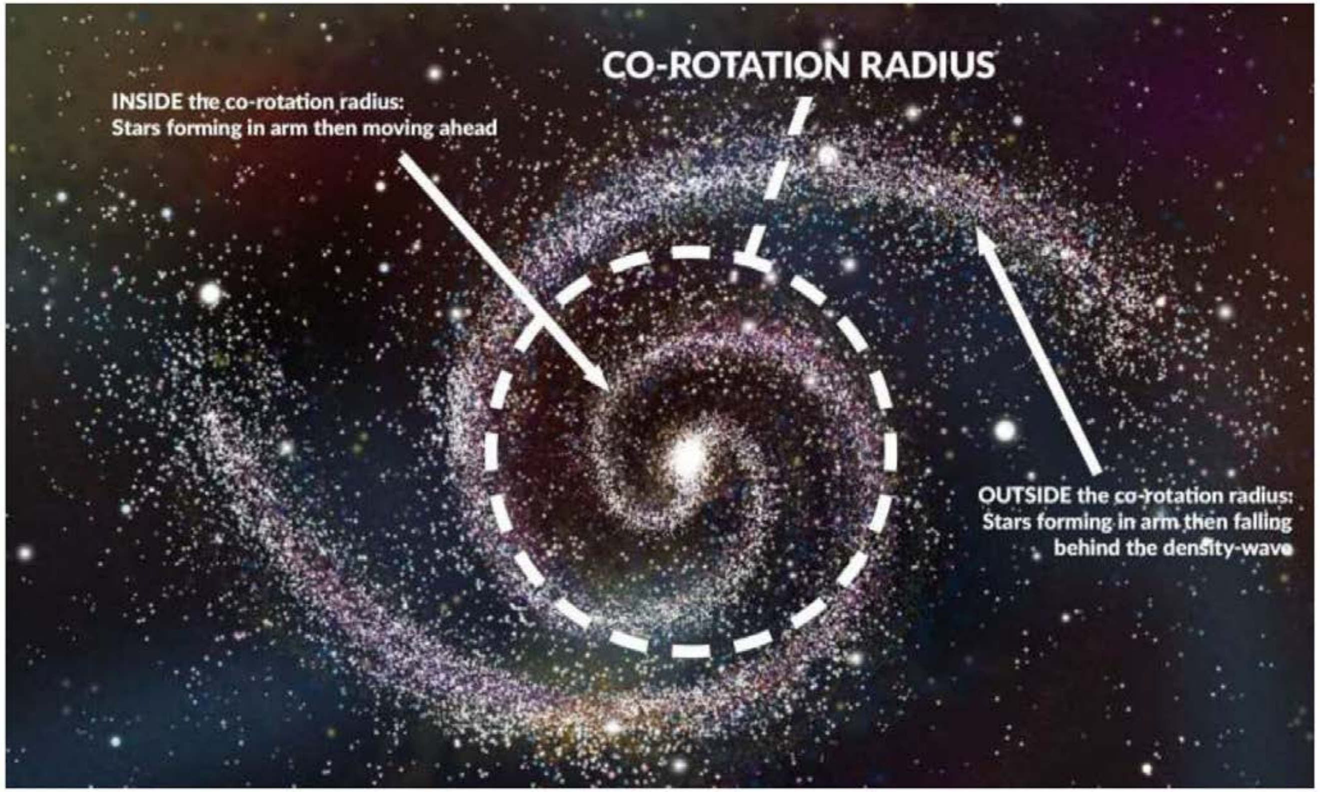
They found that each group of stars formed an arm with a slightly different "pitch angle," […].

By comparing […] to the angle formed by the center of the density wave, they showed that the location of these groups of stars matches […] density wave theory.”
-- [2019 April 23, Physics.org]

APRIL 23, 2019

## Research on disk galaxies sheds light on movement of stars

by Camilla Shumaker, University of Arkansas




Credit: University of Arkansas

University of Arkansas astrophysicists have taken an important step toward solving the mystery of how disk galaxies maintain the shape of their spiral arms. Their findings support the theory that these arms are created by a wave of denser matter that creates the spiral pattern as it travels across the galaxy.

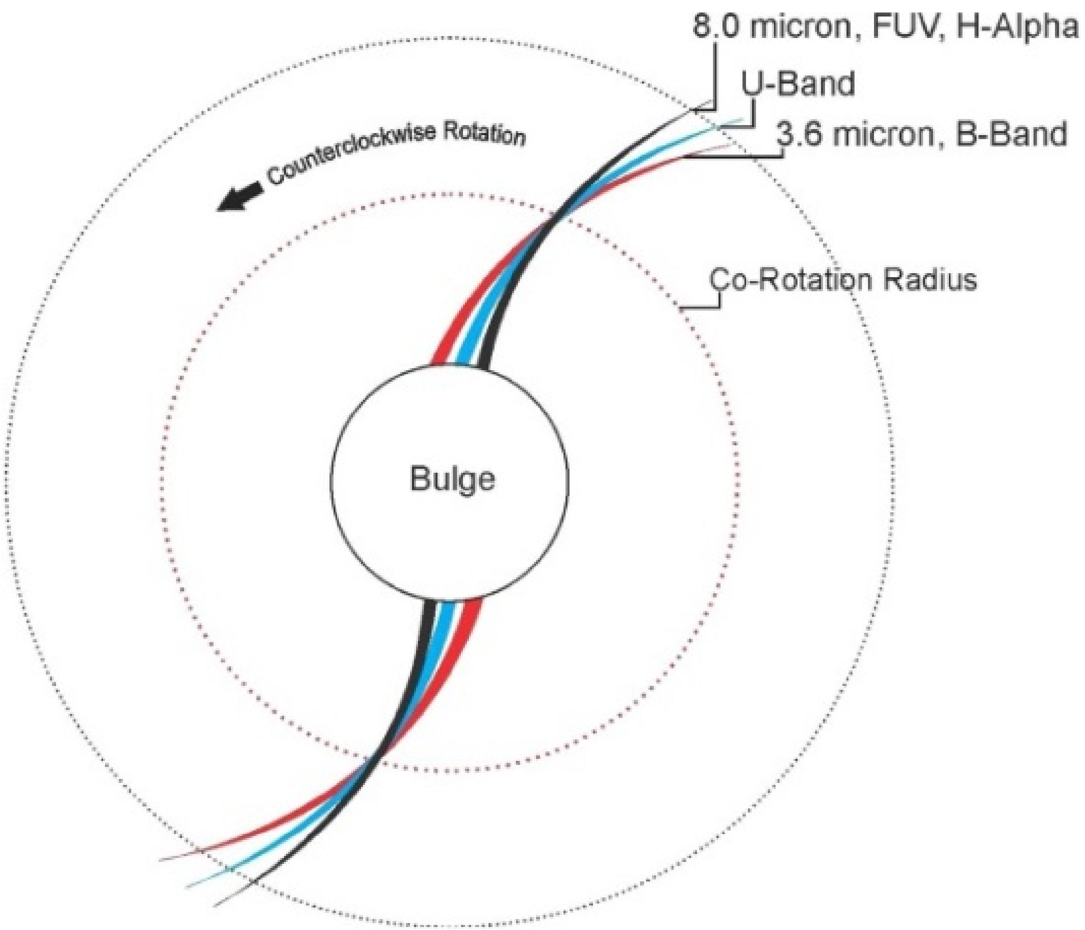
# Variation in spiral arm pitch angles test spiral density wave theory

"Investigating the Origins of Spiral Structure in Disk Galaxies through a Multiwavelength Study" (Ryan Miller et al. 2019, ApJ, 874, 177)

Illustration of spiral-arm structure based upon pitch angle measurements at multiple wavelengths measured from images in NED.

- Lower pitch angle (tighter arms) for 3.6 um and the B band
- Higher pitch angle (looser arms) for U band
- Even higher pitch angles (looser arms) for 8.0 um, FUV and Hα

***This means the star-forming arm is upstream from the blue arm, which is in turn upstream from the red arm. This conforms predictions of spiral density-wave theory.***



# NED is used heavily in studies of the Hubble Constant

## An independent determination of the local Hubble constant

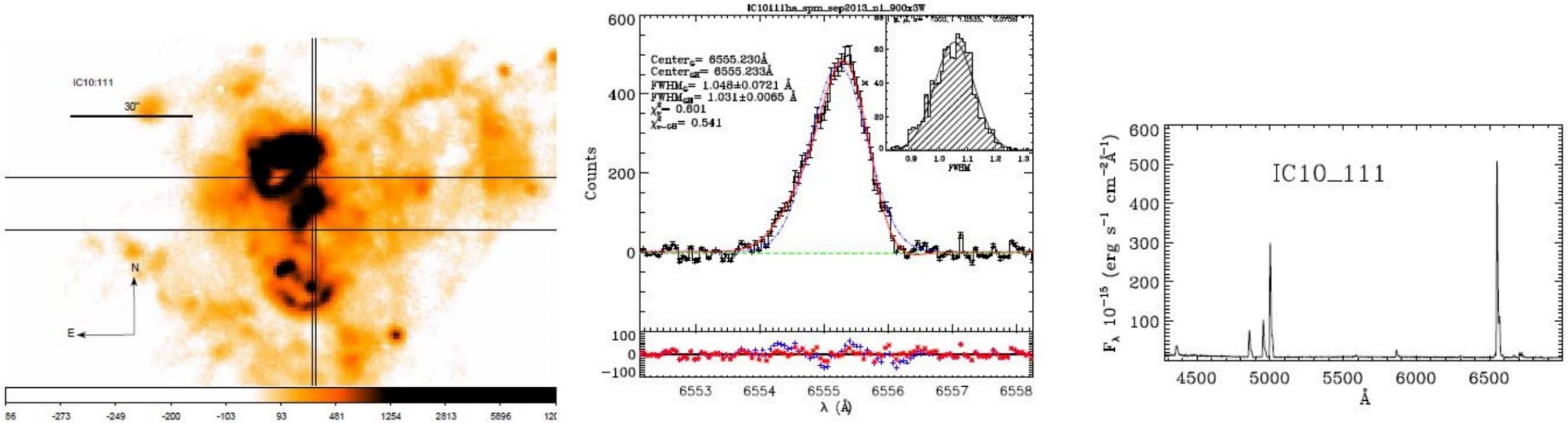
David Fernández Arenas, Elena Terlevich, Roberto Terlevich, Jorge Melnick, Ricardo Chávez, Fabio Bresolin, Eduardo Telles, Manolis Plionis, Spyros Basilakos

*Monthly Notices of the Royal Astronomical Society*, Volume 474, Issue 1, February 2018, Pages 1250–1276,

Fernandez-Arena et al. 2018, MNRAS, 474. 1250

Fernández Arenas et al. (2018) recently used a "standard candle" distance indicator provided by the relationship between integrated H line luminosity and the velocity dispersion of the ionized gas of H II galaxies (HIIGs) and giant H II regions (GHIIRs) to study new data for 36 GHIIRs in 13 galaxies.

***The authors made use of NED as a substantial source of data, including Cepheid distances and Hα images used to target their spectroscopic observations.***




Hα image from NED (left), high-resolution spectral profile for the GHIIR (center), and low resolution spectrum (right).

# An independent determination of the Hubble Constant

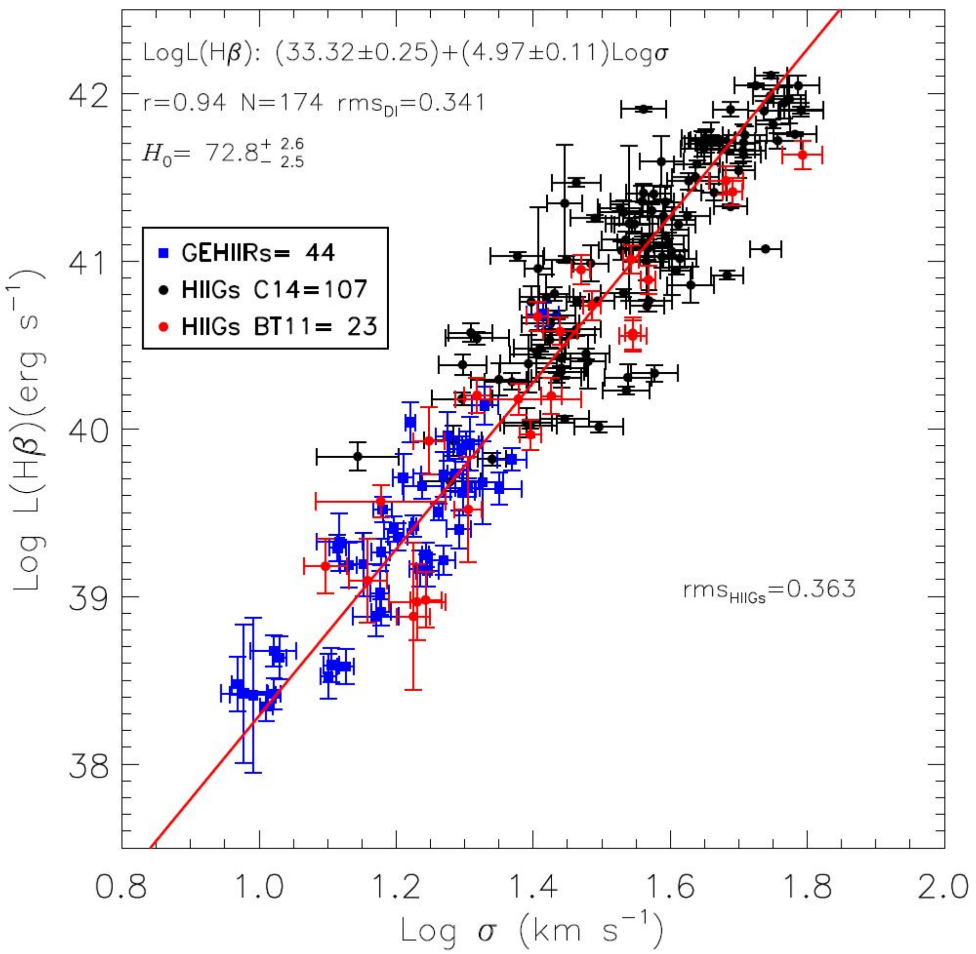
"An independent determination of the Hubble Constant" (Fernández Arenas et al. 2018).

Luminosity – velocity (L - σ) dispersion relation for Chávez et al. (2014) and Bordalo & Telles (2011) combined samples of H II galaxies.

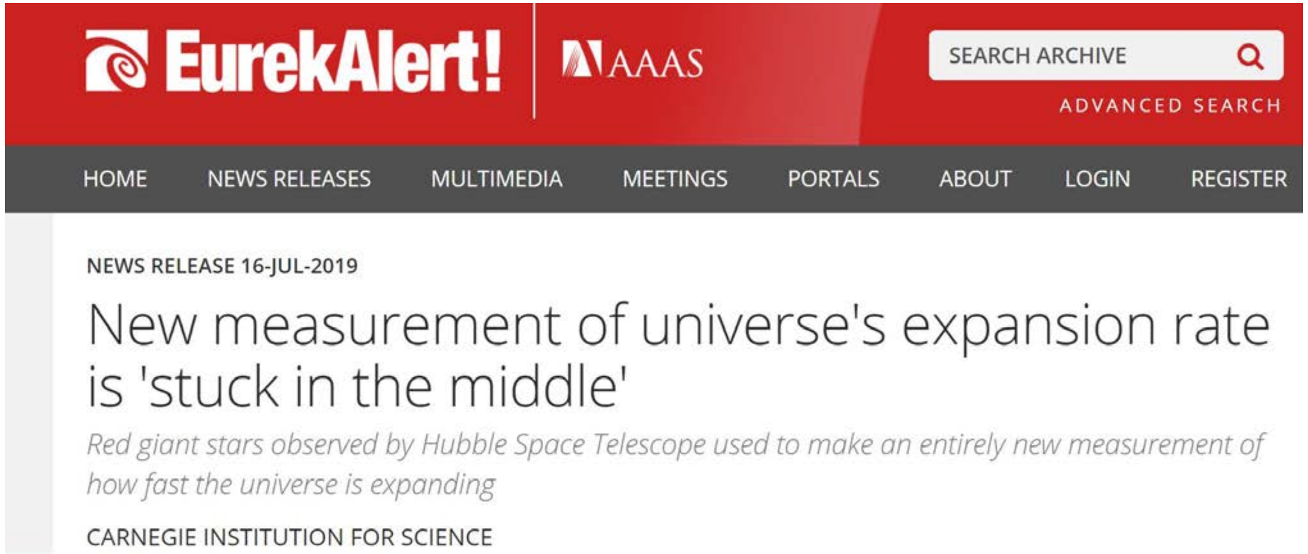
The solid line is the fit to the HII Galaxies (HIIGs).

The inset equation is the distance indicator where the slope is obtained from the fit to the HIIGs and the zero point determined.

***Their best estimate :***
***$H_0$ = 71.0 ± 2.8 (random) ± 2.1 (systematic) km $s^{-1}$ $Mpc^{-1}$, falls right between CMB Planck+ 2016 value and local value from Riess+2016.***



# New measurement of $H_0$ with TRGB method is “stuck in the middle”



EurekAlert! AAAS

SEARCH ARCHIVE

ADVANCED SEARCH

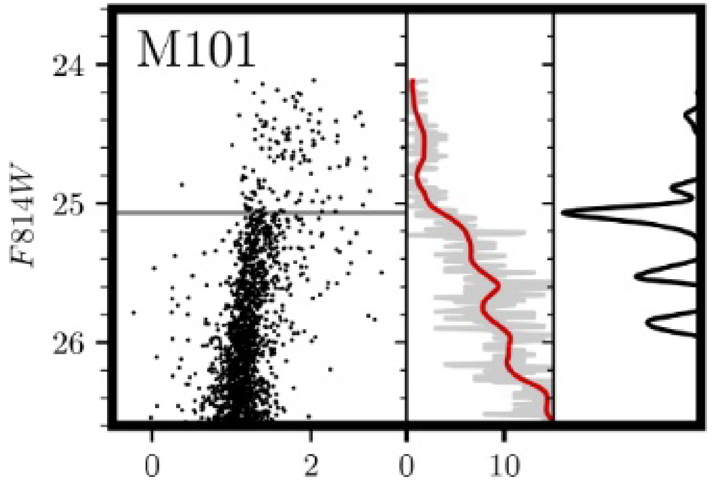
HOME NEWS RELEASES MULTIMEDIA MEETINGS PORTALS ABOUT LOGIN REGISTER

NEWS RELEASE 16-JUL-2019

New measurement of universe's expansion rate is 'stuck in the middle'

*Red giant stars observed by Hubble Space Telescope used to make an entirely new measurement of how fast the universe is expanding*

CARNEGIE INSTITUTION FOR SCIENCE




It has been claimed that cosmology is in a state of crisis given the wide divide between values of the Hubble constant inferred from the CMB as observed by the Planck mission and those measured locally using observations of Cepheid variables and distant SNe to measure $H_0$ today.

A new value measured by Freedman et al. (2019), using a high precision tip of the red giant branch (TRG) distance indicator, falls squarely between the two previous values, reducing the tension with Planck.

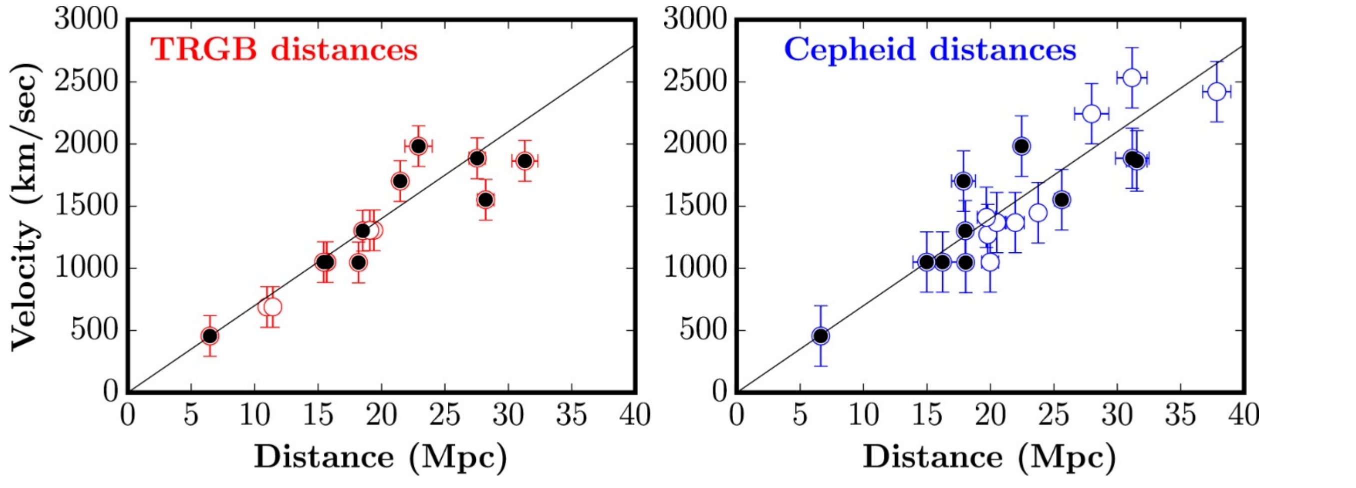
***The team based their results on a three-year study using HST, and they acknowledged being heavily reliant on NASA’s on-line archival databases, including the Astrophysics Data System (ADS), the Mikulski Archive for Space Telescopes (MAST) and NED in the planning, execution and interpretation of the observations.***

# New measurement of $H_0$ with TRGB method is "stuck in the middle"

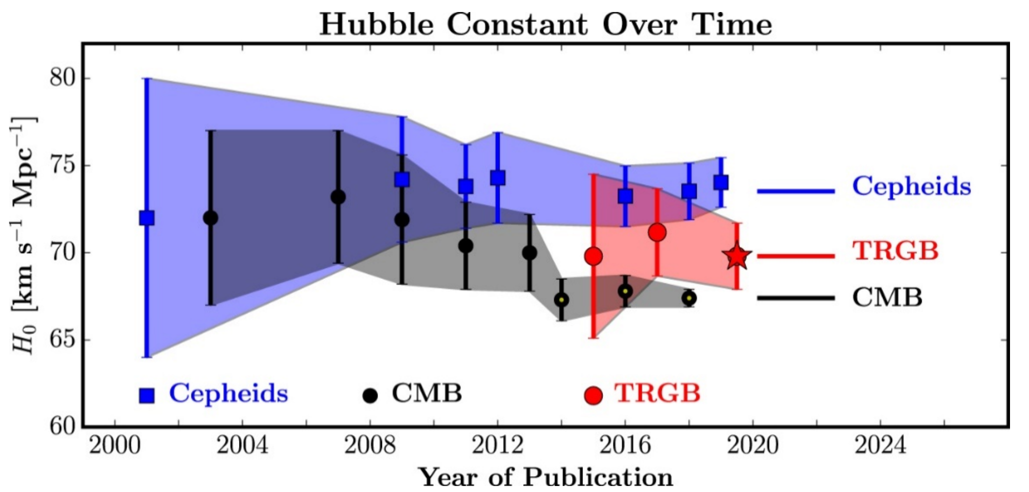
Freedman et al. 2019, ApJ, 882, 34

**The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch***

Wendy L. Freedman[1], Barry F. Madore[2], Dylan Hatt[1], Taylor J. Hoyt[1], In Sung Jang[3], Rachael L. Beaton[4], Christopher R. Burns[2], Myung Gyoon Lee[5], Andrew J. Monson[6], Jillian R. Neeley[7], M. M. Phillips[8], Jeffrey A. Rich[2], and Mark Seibert[2]




Hubble diagram for galaxies with TRGB (left) and Cepheid (right) distances. ***The velocities have been corrected for the presence of nearby mass concentrations using NED.*** A slope of $H_0$ = 70 km $s^{-1}$ $Mpc^{-1}$ is shown in black. (Freedman +2019, Fig 7).




$H_0$ estimates as a function of time.
**Black = CMB**. **Blue = Local Cepheids**.
**Red = Local TRGB; red star = best fit.**
(Freedman+2019, Fig 17).

# NED use in multi-messenger astronomy

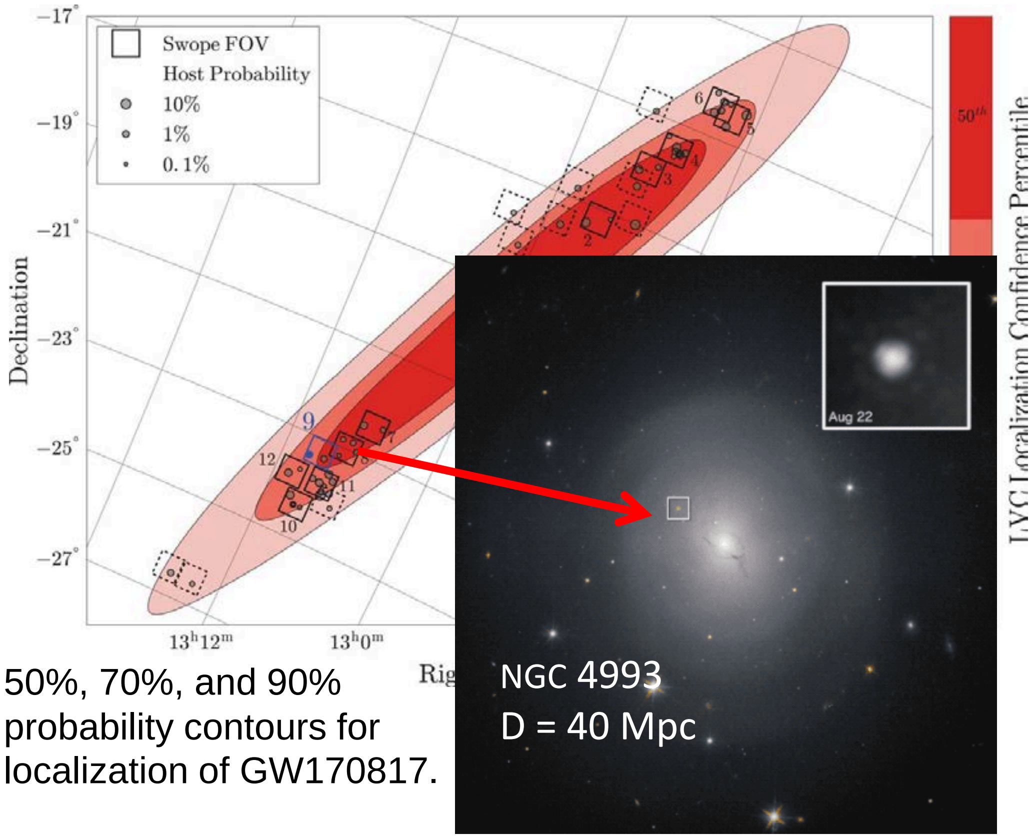
Beyond multi-λ comes multi-messenger astrophysics, combining heterogeneous measurements of EM radiation with other energy signals: gravitational waves (GWs), neutrinos and cosmic rays.

- 17 August 2017, the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo interferometer detected gravitational waves (GWs) from event GW170817
- Nearly simultaneously, Fermi and INTEGRAL telescopes in space detected a γ-ray burst GRB 170817A
- About 11 hours later, ***Coulter et al. (2017) discovered a fading optical source coincident with GW170817, Swope Supernova Survey 2017a, located in NGC 4993.***

***First detection of two neutron stars merging together both in gravitational waves and in EM radiation emitted in a cataclysmic event known as a kilonova.***

***Data in NED played an important role assisting the Coulter et al. team in narrowing down the candidate galaxies for a rapid identification of the EM counterpart.***

***"Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source" (Coulter et al. 2017, Science, 358, 1556).***




50%, 70%, and 90% probability contours for localization of GW170817.

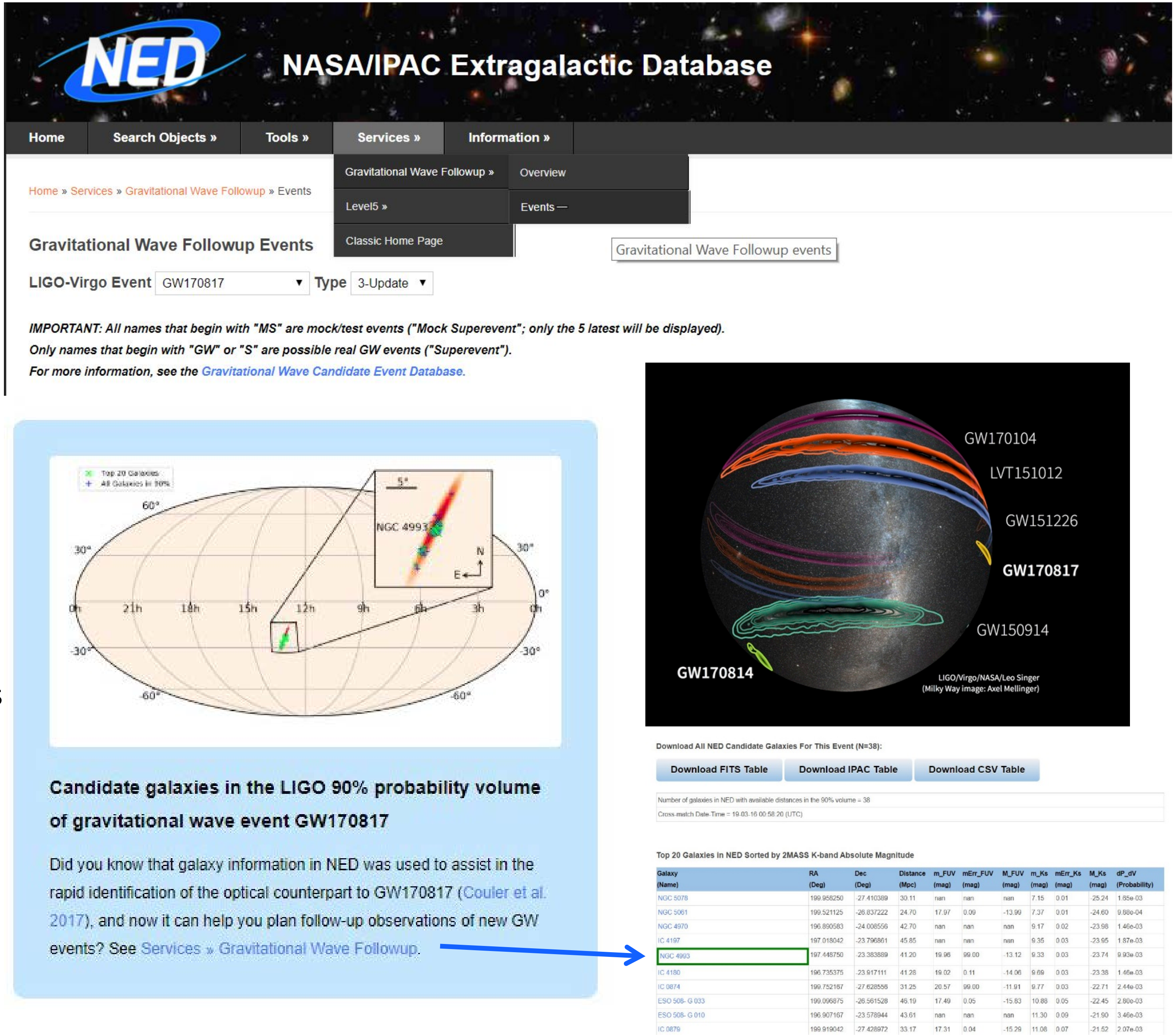
# Gravitational Wave Follow-up Service

**Inspired by Coulter et al. (2017) id of NGC 4993 as host of merging neutron stars in *GW170817***

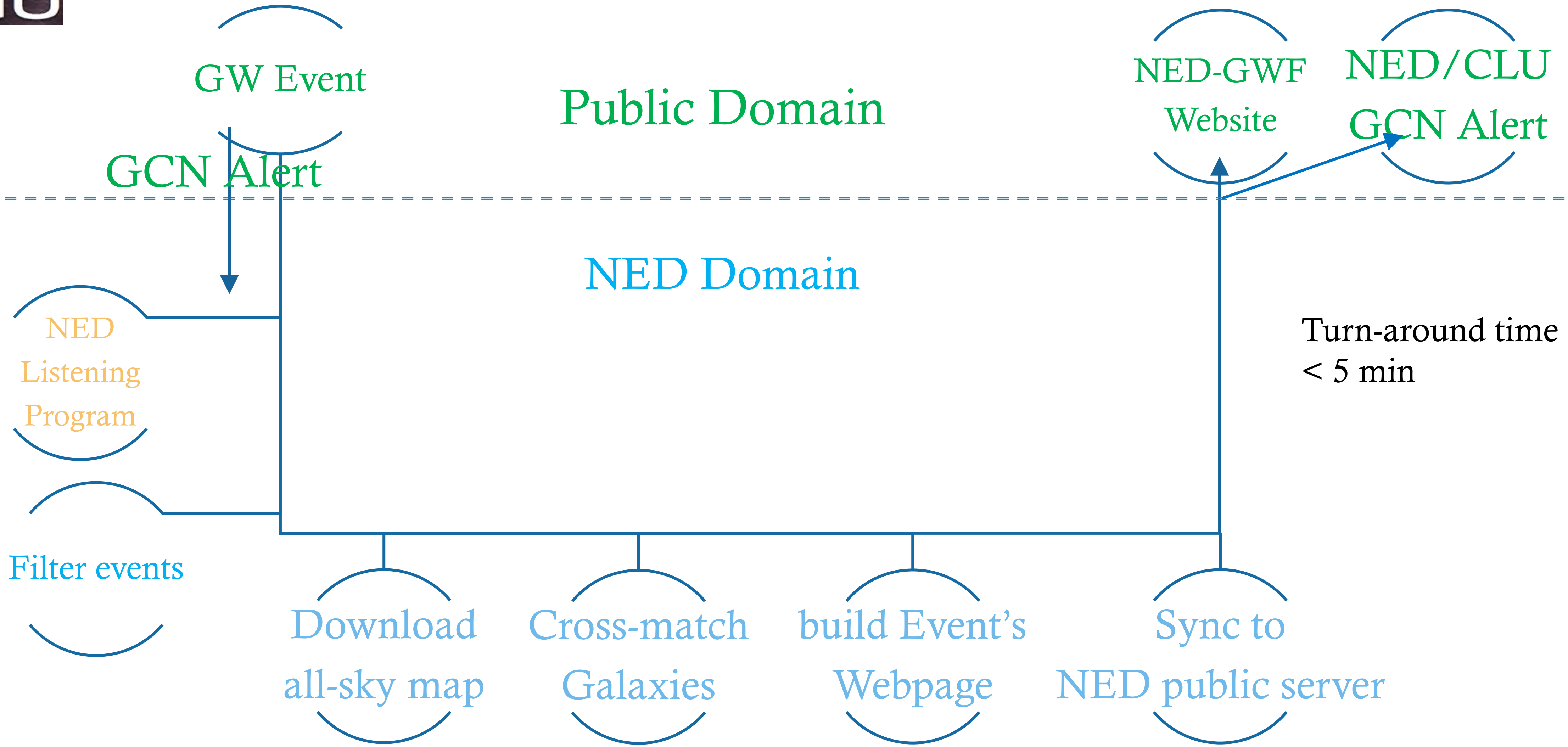
**Released in March 2019, in time for Observing Run 3 (O3) of LIGO/Virgo**

- Designed to leverage data in NED to facilitate searches for electromagnetic counterparts of GW events
- Within minutes after LIGO-Virgo collaboration issues a new GW alert on the NASA Gamma-ray Coordinates Network (GCN), NED responds by posting a list of candidate galaxies in the event's 90% probability volume.
- Menu bar: **Services »**

  **Gravitational Wave Followup » Overview**
- Currently supports galaxies with D < 200 Mpc
- Working on extending the sample



# How Does This Work?



# Into the 2020s: Challenges and complexities in data fusion

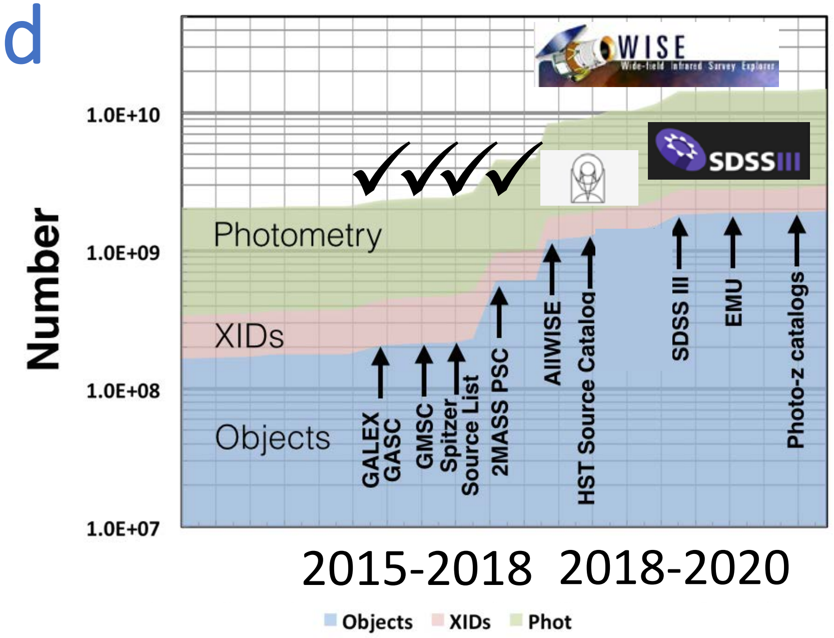
The 4 Vs of Big Data (volume, velocity, variety, and veracity) affect NED to differing degrees. But perhaps most challenging are …

**Variety:** Each mission source list, catalog, and journal article presents data differently. NED handles many types of measurements.

**Veracity:** In the literature, data are typically not vetted to the same degree as scientific results; thus there are often incomplete (meta)data, missing uncertainties, ambiguous identifiers, etc.

**Complexities** are highest in the literature, and coping requires evolving data location, capture, and integration techniques.

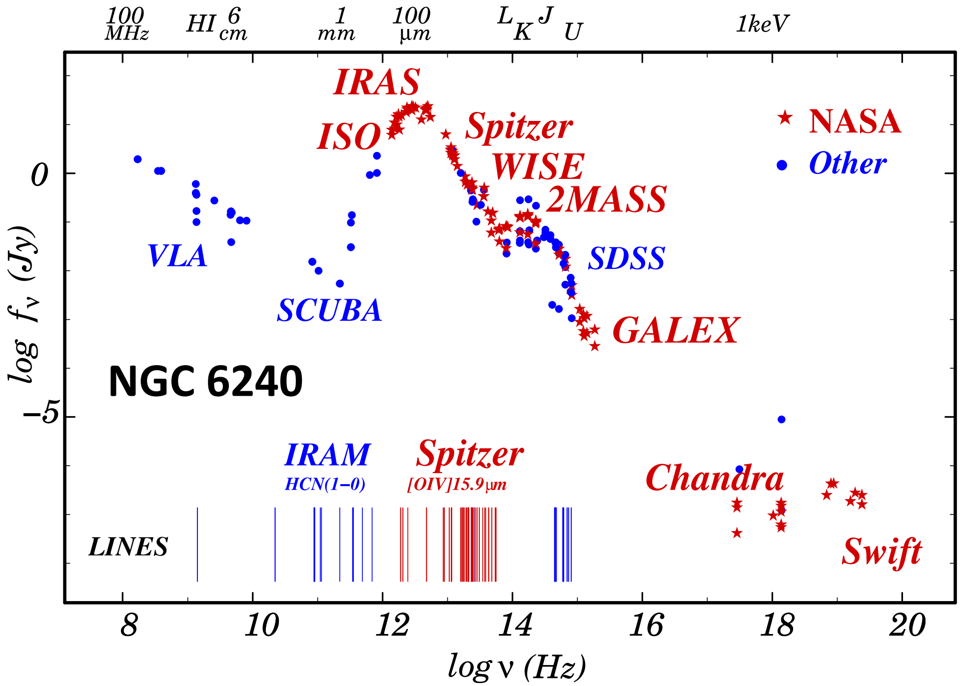
***Machine learning is a promising approach.***




JWST Euclid WFIRST LSST SKA Etc.

*And data from thousands of journal articles per year*

2020+



# Into the 2020s: Machine Learning

***Can a ML pipeline accurately and reliably classify journal article topics and their data content (e.g., astrometry, photometry, distances, classifications) for relevance to NED?***

***If so, we can focus human expertise on tasks that ML cannot do: interpret, transform, load, vet.***

***2016 Pilot project (Rick Ebert):*** supervised learning with Stanford Classifier (MaxECG) discriminative LogLinear classifier; trained using 410 papers from 2015 AJ classified by expert Cren Frayer.

***Outcome:*** 87% agreement with human expert. However, application of the same model to a different journal did not do well, due to the small size of the training set which compromised its applicability.

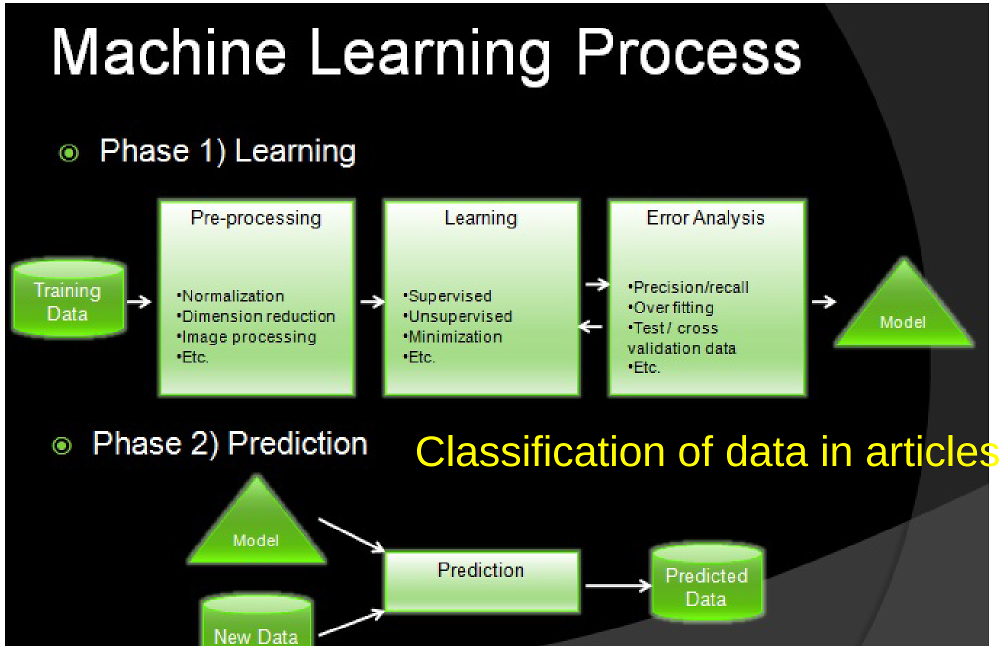



Figure credit: emaze.com

***In 2018/19, this work was resumed. Using a much larger training set increases accuracy to ≈ 97%.***

***In progress:*** First steps toward putting ML article classification into production.

***Next up:*** Apply ML to classify data types, and assist with tagging of relevant content for database entry.

***In the 2020s, we expect ML to play a substantial role in assisting data ingest and assisting science discoveries from heterogeneous, multi-wavelength data.***

# A Science Platform Network to Facilitate Astrophysics in the 2020s

A white paper submitted to the 2020 Decadal Survey

**Vandana Desai Caltech/IPAC desai@ipac.caltech.edu**

Mark Allen CDS mark.allen@astro.unistra.fr
Christophe Arviset ESA-ESAC Christophe.Arviset@esa.int
Bruce Berriman Caltech/IPAC gbb@ipac.caltech.edu
Ranga-Ram Chary Caltech/IPAC rchary@caltech.edu
David Cook Caltech/IPAC dcook@ipac.caltech.edu
Andreas Faisst Caltech/IPAC afaisst@ipac.caltech.edu
Gregory Dubois-Felsmann Caltech/IPAC gpdf@ipac.caltech.edu
Steve Groom Caltech/IPAC sgroom@ipac.caltech.edu
Leanne Guy AURA/LSST leanne.guy@lsst.org
George Helou Caltech/IPAC gxh@ipac.caltech.edu
David Imel Caltech/IPAC imel@ipac.caltech.edu
Stephanie Juneau NOAO juneau@noao.edu
Mark Lacy NRAO mlacy@nrao.edu
Gerard Lemson Johns Hopkins U. glemson1@jhu.edu
Brian Major National Research Council Canada
brian.major@nrc-cnrc.gc.ca
Joe Mazzarella Caltech/IPAC mazz@ipac.caltech.edu
Thomas Mcglynn NASA/GSFC
thomas.a.mcglynn@nasa.gov
Ivelina Momcheva STScI imomcheva@stsci.edu
Eric Murphy NRAO emurphy@nrao.edu
Knut Olsen NOAO kolsen@noao
Josh Peek STScI jegpeek@stsci.edu
Alexandra Pope UMass, Amherst pope@astro.umass.edu
David Shupe Caltech/IPAC shupe@ipac.caltech.edu
Alan Smale NASA/GSFC alan.p.smale@nasa.gov
Arfon Smith STScI arfon@stsci.edu
Nathaniel Stickley Caltech/IPAC nrstickley@ipac.caltech.edu
Harry Teplitz Caltech/IPAC hit@ipac.caltech.edu
Ani Thakar Johns Hopkins University thakar@jhu.edu
Xiuqin Wu Caltech/IPAC xiuqin@ipac.caltech.edu

# Into the 2020s: A Science Platform Network

***"In this white paper, we advocate for the adequate funding of data centers to develop and operate "science platforms", which will provide storage and computing resources for the astronomical community to run analyses near the data. Furthermore, these platforms should be connected to enable cross-center analysis and processing."***

## Big Data Challenges in the 2020s

- Searching through large data sets – maintaining reasonable response times as the data explode
- Visualizing and interpreting large data sets
- Downloading and storing large volumes of data
- Astroinformatics & Astrostatistics
- Multi-archive interoperability
- Reproducibility & Replicability
- Accessing adequate computing facilities
- Collaboration

## Vision: Science Platform Network for Astronomy

- Interactive and batch analysis environment co-located with data
- Application Program Interfaces (APIs)
- Web portals
- Collaborative Workspaces
- Cloud Computing
- Containers
- Networking and Data Transfer Tools
- Standardized data models and web APIs

***"Key Data Centers in the U.S. ground- and space-based astronomy community should be funded to deploy, maintain and extend science platforms by the end of the 2020s. The funding includes (a) development of software components, (b) investment in computing and storage resources for user computing, (c) operating expenses for the resulting systems and (d) resources for coordination."***

# Into the 2020s: Increasing the Discovery Space in Astrophysics

Six white papers submitted to the 2020 Decadal Survey, Fabbiano et al. 2019
https://arxiv.org/abs/1903.06634

***"The history of astronomy shows that paradigm changing discoveries are not driven by well formulated scientific questions, based on the knowledge of the time. They were instead the result of the increase in discovery space fostered by new telescopes and instruments. An additional tool for increasing the discovery space is provided by the analysis and mining of the increasingly larger amount of archival data available to astronomers. Revolutionary observing facilities, and the state of the art astronomy archives needed to support these facilities, will open up the universe to new discovery."***

***We should all look forward to many exciting discoveries in the 2020s enabled by the worldwide ecosystem of interconnected astronomical archives and analysis tool.***

## Increasing the Discovery Space in Astrophysics
## A Collation of Six Submitted White Papers

**Included Thematic Areas:**
Formation and Evolution of Compact Objects
Multi-Messenger Astronomy and Astrophysics
Stars and Stellar Evolution
Planetary Systems
Star and Planet Formation
Resolved Stellar Populations
Galaxy Evolution
Cosmology and Fundamental Physics

**Principal Author:**
Name: G. Fabbiano
Institution: Center for Astrophysics | Harvard & Smithsonian (CfA)
Email: gfabbiano@cfa.harvard.edu

**Co-authors:**
M. Elvis (CfA), A. Accomazzi (CfA), A. Avelino (CfA, Harvard), G. B. Berriman (Caltech/IPAC-NExScI), N. Brickhouse (CfA), S. Bose (CfA), D. Carrera (Penn State Univ.), I. Chilingarian (CfA), F. Civano (CfA), B. Czerny (Center for Theoretical Physics, Warsaw; CfA collaborator), R. D'Abrusco (CfA), B. Diemer (Cfa), J. Drake (CfA), R. Emami-Meibody (CfA), J. R. Farah (CfA), G. G. Fazio (CfA), E. Feigelson (Penn State Univ.), F. Fornasini (CfA), Jay Gallagher (Univ. Wisconsin), J. Grindlay (CfA), L. Hernquist (CfA), D. J. James (CfA), M. Karovska (CfA), V. Kashyap (CfA), D.-W. Kim (CfA), G. M. Lacy (NRAO), J. Lazio (JPL, Caltech), E. Lusso (Univ. of Florence, Italy; CfA collaborator), W. P. Maksym (CfA), R. Martinez Galarza (CfA), J. Mazzarella (Caltech/IPAC), M. Ntampaka (CfA), G. Risaliti (Univ. of Florence, Italy; CfA collaborator),D. Sanders (Institute for Astronomy, Hawaii), N. Scoville (Caltech), I. Shapiro (CfA), A. Siemiginowska (CfA), A. Smith (STScI), H. Smith (CfA), I. Stephens (CfA), A. Szentgyorgyi (CfA), S. Tacchella (CfA), A. Thakar (Johns Hopkins Univ.), V. Tolls (CfA), S. Vrtilek (CfA), B. Wilkes (CfA), D. Wilner (CfA), S. P. Willner (CfA), S. J. Wolk (CfA) Zhao, J.-H. (CfA)

# Summary

Ongoing advances in ***joining data across the spectrum from heterogeneous data sets***, combined with new UI capabilities, are helping astronomers make new discoveries directly from NED.

The ***evolution of NED from a reference database into an engine for science discovery*** was demonstrated with super spiral galaxies and analysis of spiral arm pitch angles to test spiral density wave theory. Some recent studies where info from NED played a less central, but important role, was also reviewed, including precision cosmology, and multi-messenger astronomy.

***Challenges in fusing heterogeneous, multi-wavelength data*** involve all four V's of Big Data, and are more difficult when combining data from the literature with surveys and missions.

Primary limitations of joining heterogeneous datasets across the spectrum are ***beam (resolution) differences from different telescopes/bands and lack of sufficient vetting of data in the literature.***

*Further opportunities for exciting new science from all astro archives will increase dramatically in the 2020s with **further expansion in discovery space** enabled not only by incorporating new data, but also **by applying new tech for managing and analyzing even bigger and more complex data**.*

These technologies include cloud, HPC & GPUs, (meta)data standards, advanced visualization, astroinformatics, machine learning, science platforms, etc. ➔ ***ADASS XXIX!***

# THANK YOU!