



NASA's Long-Term Astrophysics Data Archives

L. M. Rebull, IRSA
17 Oct 2016



HEASARC



With Input From:

- Vandana Desai (IRSA)
- Harry Teplitz (IRSA)
- Steve Groom (IRSA)
- Rachel Akeson (NExSci)
- Bruce Berriman (NExSci)
- George Helou (IPAC)
- David Imel (IPAC)
- Joe Mazzarella (NED)
- Alberto Accomazzi (ADS)
- Tom McGlynn (GSFC)
- Alan Smale (GSFC)
- Rick White (MAST)



Having said that...

- I'm an astronomer at IRSA, so that is what I know best, and most of my examples are IRSA-focused by necessity.

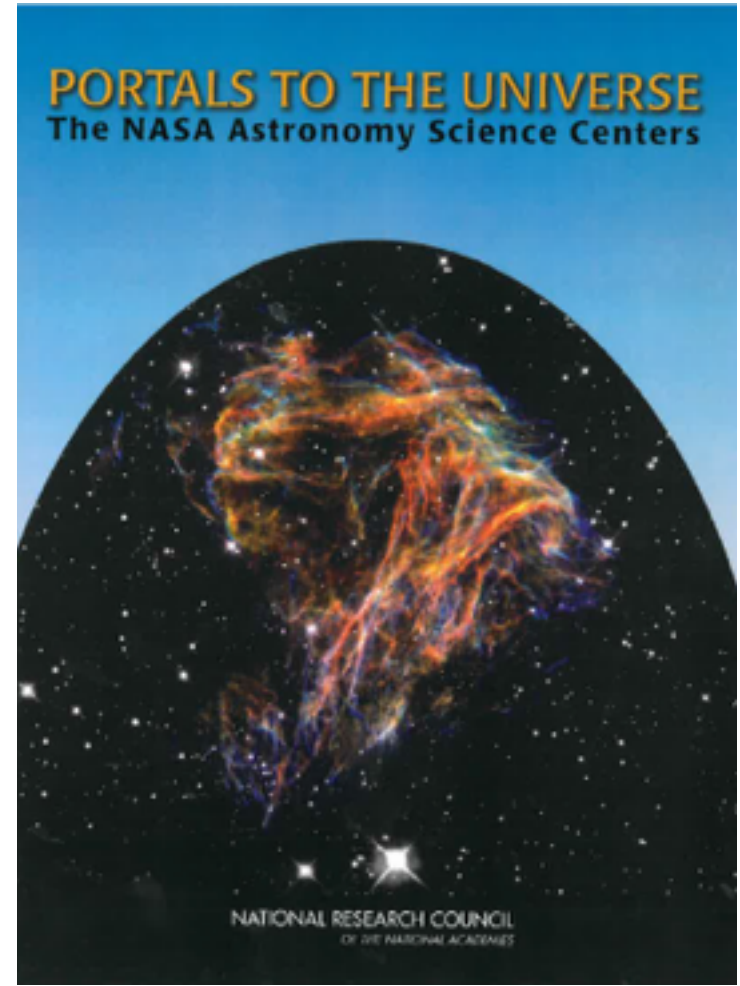
NASA's Commitment to Astrophysics Data Archives

- “NASA has regarded data handling and archiving as an integral part of space missions.”
- “This support now provides the major return on the considerable investment the agency made... over the past 20 years.”



“A Sustainable Archive”

- Provides data discovery and analysis tools. Facilitates new science.
- Contains high-quality, reliable data.
- Provides simple and useful tools to a broad community.
- Provides user support to the novice as well as to the power user.
- Adapts and evolves in response to community input.

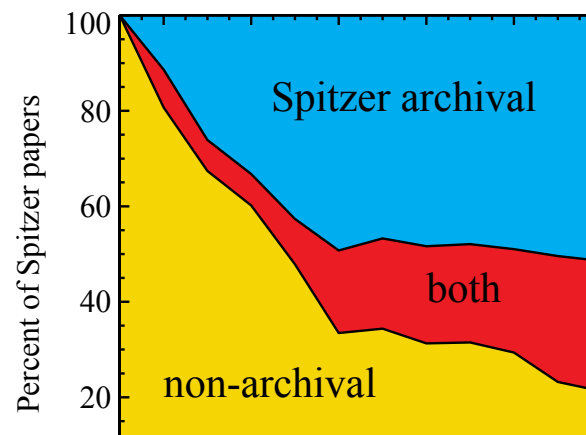
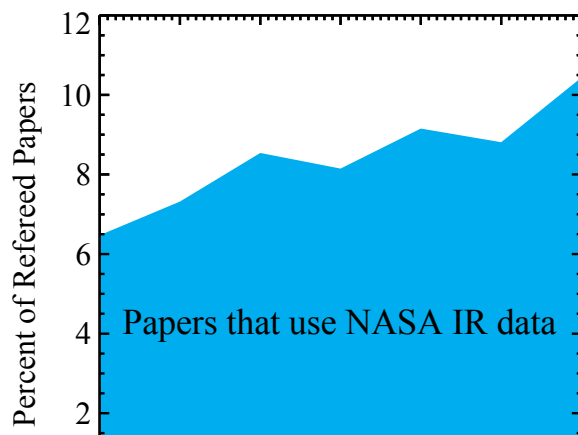


An Archive's Job

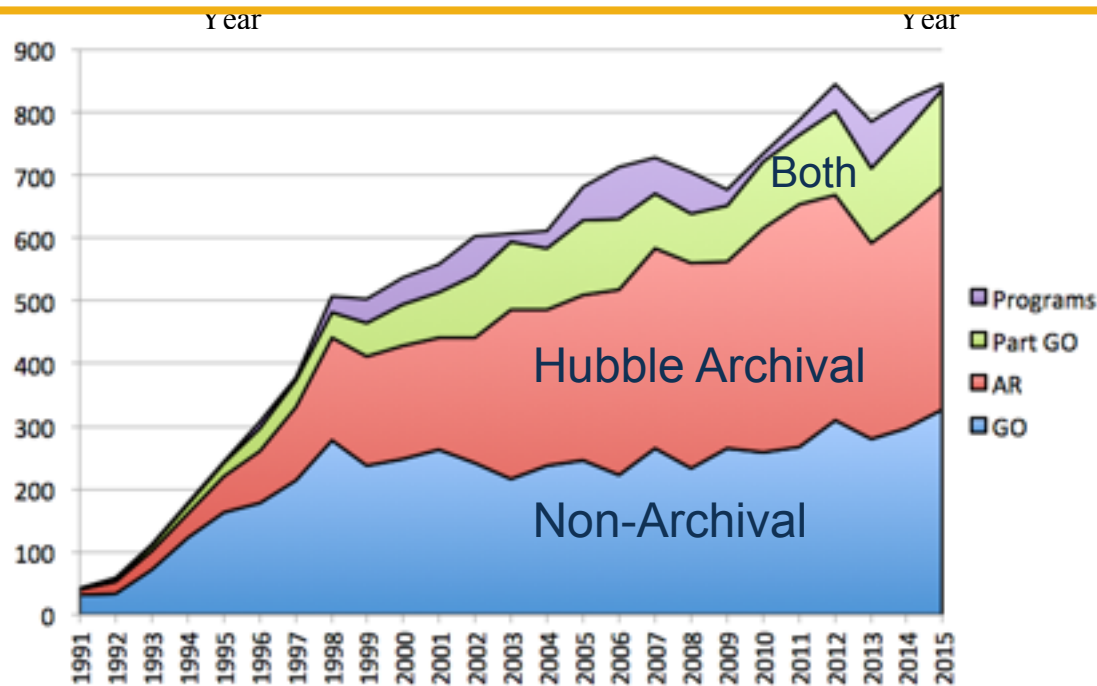
- Ingest new data (and reprocessing of old data).
- Maintain/serve vital repository of irreplaceable data:
 - Support for observation planning and mission planning.
 - Resource for original science.
 - High level science products.
- Enable cutting-edge research:
 - ~~API and Virtual Observatory.~~
 - User support by experts.
 - New/enhanced services.
 - Multi-wavelength projects.



Archival Papers Outnumber Non-Archival (GO/PI) Papers



Archives double an observatory's output!



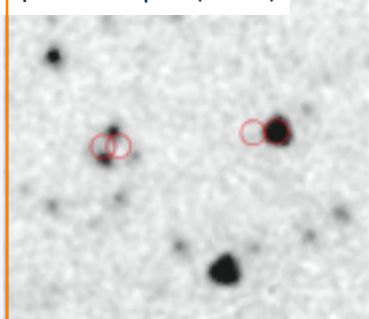
IRSA Science Highlights

Time domain

WISE 4.5 μ m (2010)

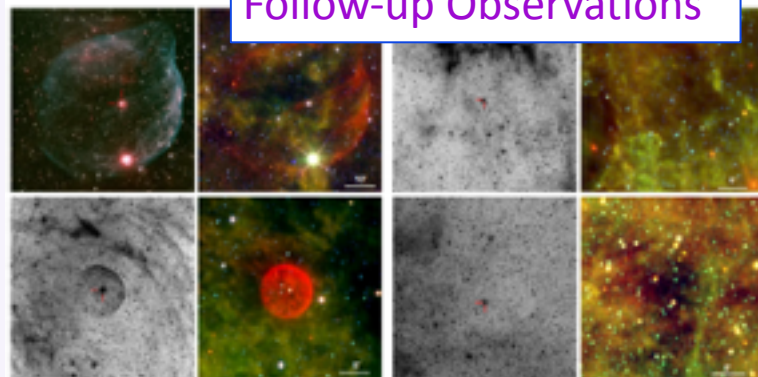


Spitzer 4.5 μ m (2014)

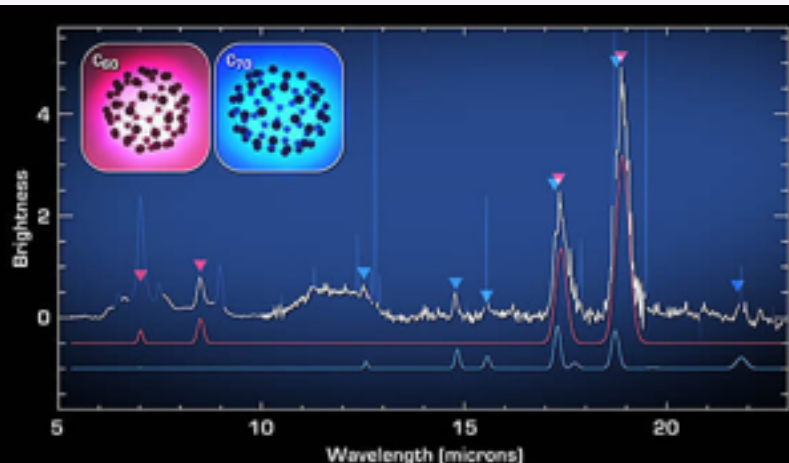


WISE+Spitzer discover the coldest brown dwarf (Luhman 2014, ApJL 786, L18)

Follow-up Observations



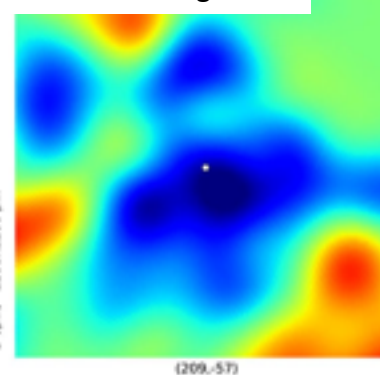
WISE morphological study of Wolf-Rayet nebulae, Toala et al. (2015, A&A 578, 66)



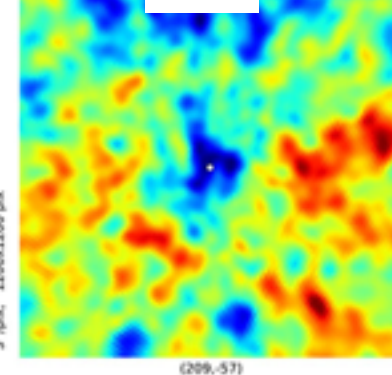
Buckyballs in a Young Planetary Nebula (Cami et al., 2010, Science, 329 1180)

Re-analysis of Spectra

WISE+2MASS galaxies



Planck



WISE+2MASS+PanSTARRS data may reveal super-void in CMB cold spot seen by Planck; (Szapudi et al. 2015, MNRAS, 450, 288)

Combination of Surveys

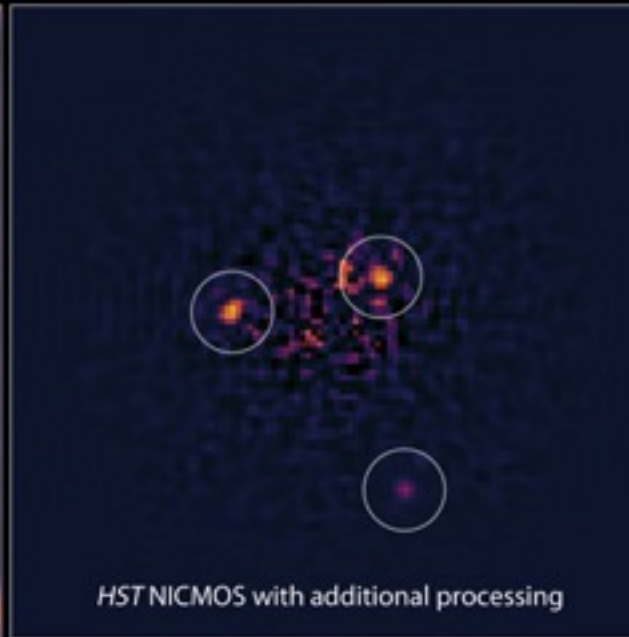


MAST Science Highlight: R8799 b,c,d imaged by HST in 1998

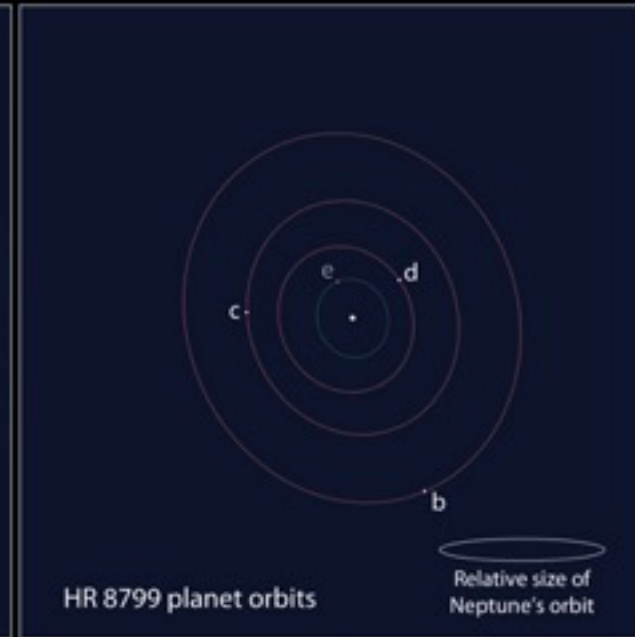
Exoplanet HR 8799 System



HST NICMOS



HST NICMOS with additional processing



HR 8799 planet orbits

Relative size of
Neptune's orbit

NASA, ESA, and R. Soummer (STScI)

STScI-PRC11-29

planet b:

83,000x fainter than star at 1.72 arcsec

planet c:

36,000x fainter than star at 0.96 arcsec

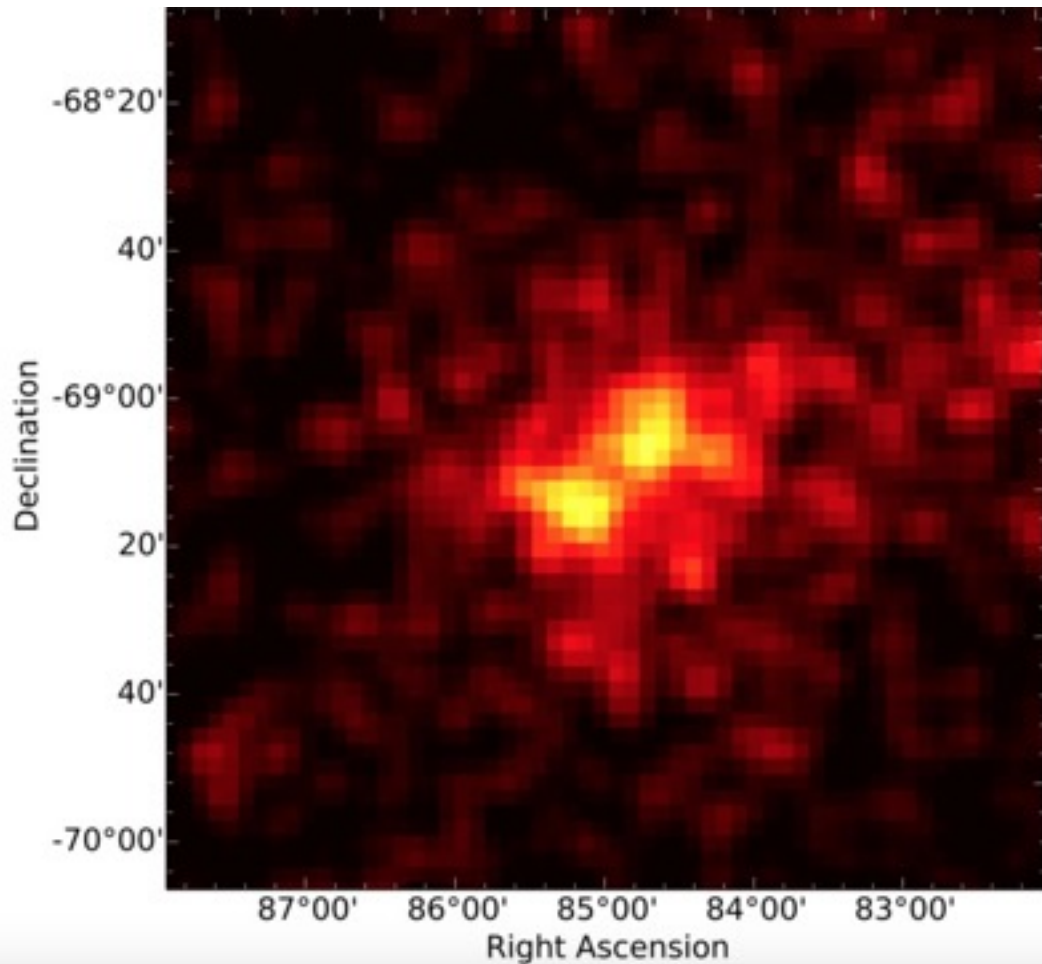
planet d:

33,000x fainter than star at 0.60 arcsec

Post-processing speckle subtraction, >an order of magnitude contrast improvement over "state of the art" when data taken in 1998.

Soummer et al. 2011, Pueyo et al. 2015

HEASARC Science Highlight



Tarantula
Nebula:
Combining 6
years of
Fermi data to
discover the
first
extragalactic
gamma-ray
pulsar!

Credit: NASA/DOE/Fermi LAT Collaboration, 2015, Science, 350, 801

Some NASA Archives by Center

- IPAC:
 - IRSA – IR, sub mm
 - NED – Extragalactic
 - NASA Exoplanet Archive
 - KOA (w/ WMKO) – Keck Observatory
- STScI : MAST – UV, optical, IR
- GSFC : HEASARC – high energy, CMB
- SAO/CfA:
 - ADS – literature
 - Chandra





IRSA

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

- IRSA = NASA/IPAC Infrared Science Archive, @ Caltech, @ IPAC.
- Charter is to provide interface to all NASA infrared and sub-mm data sets. ($\sim 1 \mu\text{m}$ $\rightarrow$ $\sim 1 \text{ cm}$).
- Founded 1993, original home to IRAS data (1983).
- IRSA ensures the legacy of NASA's "golden age" of IR:
 - Enable research that has not yet been envisioned.
 - Priorities set by missions and the community.
 - Support future flight missions.
- IRSA datasets are cited in about 10% of astronomical refereed journal articles.
- Total holdings over a petabyte ($>1000 \text{ TB}$); >120 billion rows in catalogs as of 9/2016.
- Total number queries: Over 33.7 million queries, 255 TB downloaded 1-9/2016.





NASA/IPAC INFRARED SCIENCE ARCHIVE

IRSA | DATA SETS | SEARCH | TOOLS | HELP

Login

Search for Source

Name or Coordinates

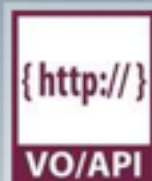
Search

Radius 10

arcsec

Search Catalog: WISE

Search



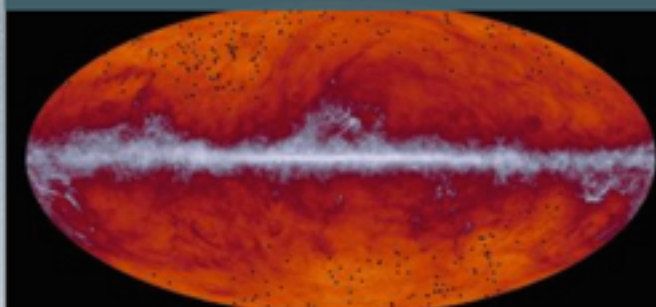
Knowledgebase

Documentation

Video Tutorials

Help Desk

Planck Public Data Release 2 Update



Planck Public Data Release 2 Update includes HFI polarization maps, time ordered data samples for variability analysis, cosmological likelihoods and source catalogs. See documentation for details of all updated data products.

Past News

Featured Images

- * Contact
- * Privacy Policy
- * Acknowledge IRSA

Search IRSA



Caltech JPL



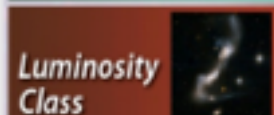
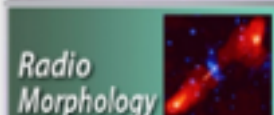
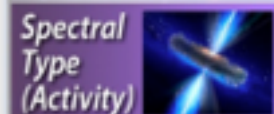
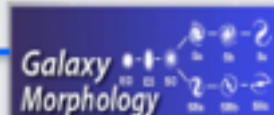


NED

- NED = NASA/IPAC Extragalactic Database (@ Caltech, @ IPAC).
- Primary hub for multi-wavelength research on extragalactic science.
- Merges data from catalogs and literature.
- 1000s of extragalactic papers per year, with unique measurements for millions of objects.
- 215 million objects with 256 million cross-IDs (from >102,000 articles/catalogs)!
- 2 billion photometric data points joined into spectral energy distributions.
- Myriad cross-links, notes, etc.
- Updates every few months.

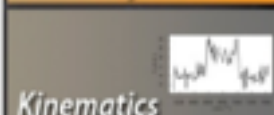
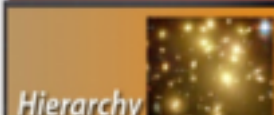
Published

- Object names
- Coordinates
- Redshifts
- D_{Mpc}
- Fluxes
- Sizes
- Attributes
- References
- Notes



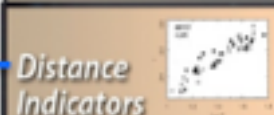
Contributed

- Images
- Spectra



Derived

- *Cross-IDs*
- *Distances*
- *Metric sizes*
- *Luminosities*
- *Velocity corrections*
- *Cosmological corrections*
- *SEDs*
- A_{λ}



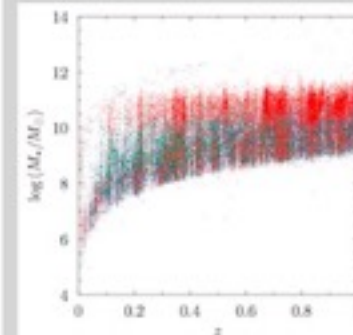


Go

News & Featured Updates - September 2016

- 2,049,873 new object links to 962 references
- 210,889 new literature sources cross-matched
- Hundreds of new images
- Thousands of new spectra
- Latest articles in Level 5

[Other news ...](#)



Recently added: NED contains 56,152 COSMOS, AEGIS, GEMS, and GOODS galaxies from the article "Bulgeless Galaxies at Intermediate Redshift", Bizzocchi, L. et al. The distribution of galaxy stellar masses (Fig. 2) is shown above.

[2014ApJ...782...22B](#).

The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

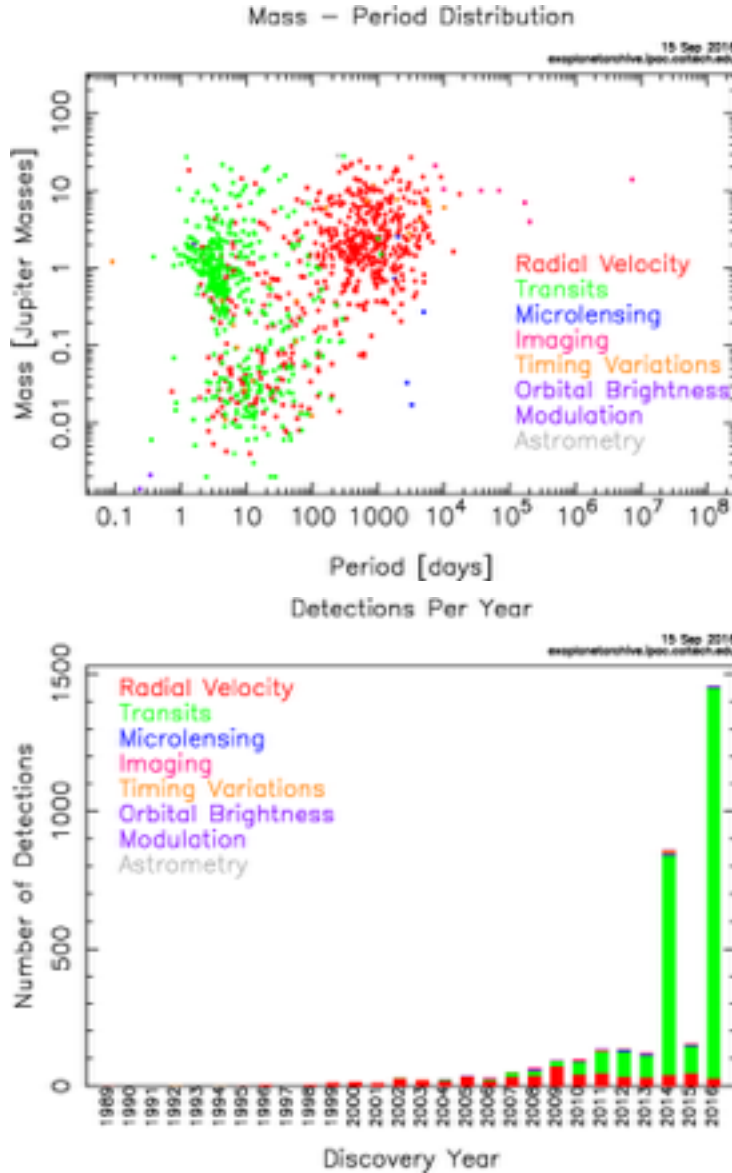


About NED Acknowledging NED Connect with NED:

NASA Exoplanet Archive

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

- (Also @ Caltech, @ IPAC)
- Focused on confirmed and candidate exoplanets and on data searching for exoplanets
- Includes Kepler data, and US portal to CoRoT data.
- Online tools to work with these data, like the periodogram service.
- Place for observers to upload/ share data (Exo-FOP).



NASA EXOPLANET ARCHIVE

A SERVICE OF NASA EXOPLANET SCIENCE INSTITUTE

FOR THE PUBLIC
PLANETQUEST


[Home](#)
[About the Archive](#)
[Data](#)
[Tools](#)
[User Guides & Help Desk](#)

3,375 Confirmed Planets
08/25/2016 →

570 Multi-Planet Systems
08/25/2016 →

4,696 Kepler Candidates
09/18/2015 →

View more Planet and
Candidate statistics →

Explore the Archive

[Advanced Search →](#)

Transit Surveys

21,853,310 Light Curves



The first space mission to search for Earth-sized and smaller planets in the habitable zone of other stars in our neighborhood of the galaxy.



Light Curves →



Objects of Interest (KOI) →



Threshold-Crossing Events →



Search Stellar Data →



Completeness and
Reliability Products →



Documentation →

Kepler

K2

KELT

SuperWASP

More

Meet the New Neighbor: Proxima Cen b

August 25, 2016 • New Data

Proxima Centauri b, the new Earth-sized planet that has the astronomy world buzzing this week, is in the Archive. To view its parameters, click to see its Planet Overview page.



News →

1

2

3

4

Plots →

1

2

3

4

Tools & Services



Periodogram →



Predicted Observables for
Exoplanets Service →



Transit and
Ephemeris Service →



Build a Query (API) →



Search Interactive Tables →



Search Extended
Planet Data →

Work with Data



Confirmed Planets
Plotting Tool →



Confirmed Planets Table →



Search K2 Targets →



Bulk Download Service →



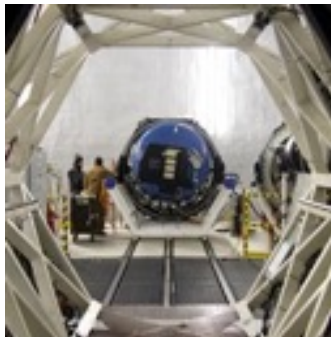
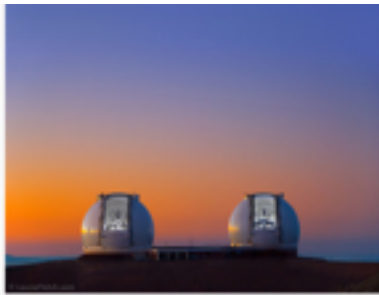
Transit Spectroscopy →



ExoFOP →

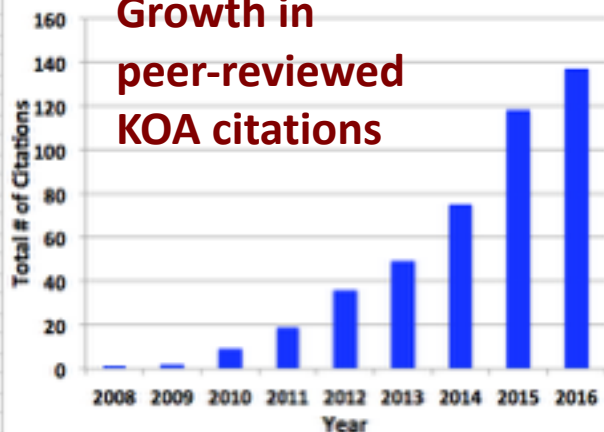
KOA

<https://koa.ipac.caltech.edu/>



- A collaboration between NExSci and the W. M. Keck Observatory.
- Public data for **all ten Keck instruments** since the Observatory saw first-light in 1994.
 - Browse-quality images of raw data.
 - Browse-quality, reduced data for HIRES, NIRC2, OSIRIS, and LWS, created by automating pipelines.
- Contributed data: Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ; N. Lehner, PI).
- Coming soon: NIRSPEC extracted spectra; moving target services.
- See Luca Rizzi's poster for more details.

**Growth in
peer-reviewed
KOA citations**





+ KOA Search Form

+ KI Search Form

+ Publicly Available Data

+ Contributed Datasets

+ KOA User Guide

+ Program Interface

+ Reducing Keck Data

+ FAQ

+ KOA Helpdesk

+ KOA News

+ KOA Bibliography

+ Related Archives

+ Login / Logout

KOA Data Access Service - v11.8.3

Currently not logged in. [Log in]

The W. M. Keck Observatory Archive (KOA) ingests and curates data from the following instruments: DEIMOS, ESI, HIRES, KI, LRIS, LWS, MOSFIRE, NIRC, NIRC2, NIRSPEC, and OSIRIS. The schedule for public access to non-proprietary data for these and future instrument releases can be found in the description of [User Access and Proprietary Periods](#). KOA serves data released by the [Keck Observatory Database of Ionized Absorbers towards Quasars \(KODIAQ\)](#), created by combining archival observations of quasars made with public HIRES data.



OSIRIS SPEC Upgrade - OSIRIS is back online for science after its SPEC detector upgrade. Interactive Visualizers are now available for calibrated NIRC2 images and reduced OSIRIS data cubes. A **FITS header viewer** is available from the quicklook column on the results pages.

The KOA web pages are supported under most common hardware platforms and operating systems, but on tablets such as iPads and Androids, features using mouse click-and-drag functionality are not available. Internet Explorer is not supported - we recommend Firefox or Chrome for Windows users.

Basic Search

More Search Options

Released Programs

1. Choose Instrument (all modes): ?

- | | | |
|--|---|---|
| ☒ HIRES | ☒ NIRC2 | ☒ NIRSPEC |
| ☒ DEIMOS | ☒ ESI | ☒ LRIS |
| ☒ LWS | ☒ MOSFIRE | ☒ NIRC |
| ☒ OSIRIS | | |

(To retrieve public Keck Interferometer data, use the dedicated [KI Search Form](#))

2. Search for Data by Night or Location: ?



MAST



<https://archive.stsci.edu/>

- Mikulski Archive for Space Telescopes @ STScI.
- Archive established with HST launch in 1990.
- Multi-mission since addition of IUE in 1998.
- Optical, UV, IR.
- Includes Hubble, Kepler, GALEX, IUE, FUSE, TESS, JWST, Pan-STARRS, DSS, GSC2, ...
- >700 TB of data (soon to jump to 2.5 PB with Pan-STARRS release), 2 million searches per month, 1200 refereed papers per year.



Select a collection...

All MAST Observations

[About Collections...](#)

and enter target:

Enter object name or RA and Dec to cone search

[Search](#)

[Show Examples...](#) [Random Search](#) [Advanced Search](#)



Upload Target List



My Download Basket: 0 files

[User Manual/Help](#)

[Leave Feedback](#)

[About This Site](#)

Home Page

AstroView

MAST: Barbara A. Mikulski Archive for Space Telescopes

The MAST Portal lets you search multiple collections of astronomical datasets all in one place. Use this tool to find astronomical data, publications, and images.

Quick Start:

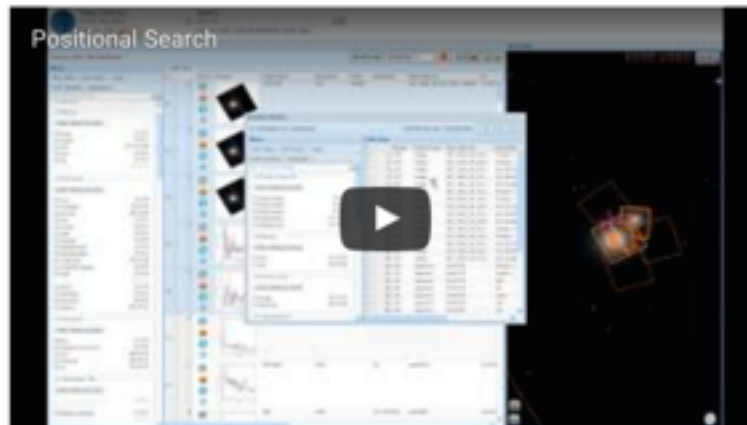
1. Select a collection and enter a new search target OR upload an existing list of targets.
2. Use the filters and analysis tools to find the exact data you're looking for.
3. Add files to the download basket to control your download options.

See the [User's Guide](#) for more detailed documentation and [video tutorials](#).

Currently available data collections:

- MAST Observations: Millions of observations from Hubble, Kepler, GALEX, IUE, FUSE, and more.
- Virtual Observatory: Search thousands of astronomical data archives from around the world for images, spectra, and catalogs.
- Hubble Source Catalog: A master catalog with a hundred million measurements of objects in Hubble images.

Featured tutorial: Conducting a search



See all video tutorials on [our YouTube channel](#).

MAST is managed by [Space Telescope Science Institute](#).

For more collections, visit [archive.stsci.edu](#).

Information about acknowledging the use of this resource may be found [here](#).

18:01:29.665 -39:28:01.10
17:44:00.000 -29:00:00.00

RA DEC
hhmmss.sss
deg





HEASARC

<http://heasarc.gsfc.nasa.gov/>

- High Energy Astrophysics Science Archive Research Center, @GSFC, since 1990.
- Extremely energetic cosmic phenomena ranging from black holes to the Big Bang.
- Chandra, XMM-Newton, Fermi, Suzaku, NuSTAR, INTEGRAL, ROSAT, Swift, & more than 20 others.
- Merged with Legacy Archive for Microwave Background Data Analysis (LAMBD A) in 2008 (CMBR): WMAP, COBE, ACT, etc.



National Aeronautics and Space Administration
Goddard Space Flight Center
Sciences and Exploration

GO

Search HEASARC website

[Advanced Search]

HEASARC Quick Links

—Quick Links—

HEASARC Home

Observatories

Archive

Calibration

Software

Tools

Students/Teachers/Public

NASA's HEASARC

High Energy Astrophysics Science Archive Research Center

About the HEASARC

Resources for Scientists

FAQ/Help

Site Map

Other Archives

Guest Observer Facilities & Science Centers

AGILE	ASCA
BeppoSAX	COBE
CGRO	Chandra
EUVE	Fermi
GALEX	HETE-2
Hitomi	INTEGRAL
MAXI	NICER
NuSTAR	ROSAT
RXTE	Suzaku
Swift	TESS
WMAP	XMM-Newton

NASA Archives

ADS	AstroGravS
EOSDIS	ExoArchive
HORIZONS	IRSA
KOA	LAMBDA
MAST	NEoScl
NED	NSSDC
PDS	SDAC
SPDF	SSC

The High Energy Astrophysics Science Archive Research Center (HEASARC) is the primary archive for NASA's (and other space agencies') missions studying electromagnetic radiation from extremely energetic cosmic phenomena ranging from black holes to the Big Bang. Since its merger with the Legacy Archive for Microwave Background Data Analysis (LAMBDA) in 2008, the HEASARC archive contains data obtained by high-energy astronomy missions observing in the extreme-ultraviolet (EUV), X-ray, and gamma-ray bands, as well as data from space missions, balloons, and ground-based facilities that have studied the relic cosmic microwave background (CMB) radiation in the sub-mm, mm and cm bands.

HEASARC Picture of the Week



APOD: Astronomy Picture of the Day



Latest News

- [Record-Breaking Galaxy Cluster Discovered](#) (30 Aug 2016)
A new record for the most distant galaxy cluster has been set using NASA's Chandra X-ray Observatory and other telescopes. This galaxy cluster called CL J1001+0220 may have been caught right after birth, a brief, but important stage of evolution never seen before. It is located about 11.1 billion light years from Earth. The discovery of this object pushes back the formation time of galaxy clusters by about 700 million years. A paper on these results has just appeared in *ApJ* and is available online [here](#).
- [Fifth IBIS/ISGRI Soft Gamma-Ray Survey Catalog](#) (30 Aug 2016)
The fifth version of the IBIS/ISGRI source catalog containing the soft gamma-ray properties of 939 sources found in the first 1000 orbits (~3000 days) of INTEGRAL observations (from [Bird et al. 2016, ApJS, 223, 15](#)) is now available in [Browse](#) and [Xamin](#).
- [Chandra Deep Field North 2-Megasecond Improved Point Source Catalog](#) (25 Aug 2016)
This catalog containing the X-ray and multiwavelength properties of 755 X-ray sources found in a new analysis of the CDF-N Field 2-Ms data (from [Xue et al. 2016, ApJS, 224, 15](#)) is now available in [Browse](#) and [Xamin](#).
- [XMM-Newton AG-16](#)



astrophysics data system



- ADS = Astrophysics Data System (@SAO/CfA).
- Indexes 12 million publications in astronomy, physics, arXiv.
- Complete coverage of astronomy and refereed physics literature.
- Tracks citations, institutional and telescope bibliographies, links to data products (back to the other archives).
- New interface and API integrating ORCID, full-text search, analytics.





astrophysics data system

[Classic Form](#)
[Modern Form](#)
[Paper Form](#)

QUICK FIELD:

[Author](#)
[First Author](#)
[Abstract](#)
[Year](#)
[Fulltext](#)
[All Search Terms](#)

Advanced ▾



author author:"huchra, john"

citations citations(author:"huchra, j") ?

first author author:"^huchra, john"

references references(author:"huchra, j") ?

abstract + title abs:"dark energy"

reviews reviews("gamma-ray bursts") ?

year year:2000

year range year:2000-2005

refereed property:refereed ?

full text full:"gravitational waves"

astronomy database:astronomy ?

publication bibstem:ApJ ?

OR abs:(planet OR star) ?



Use a classic ADS-style
form



Learn more about searching
the ADS



Access ADS data with our
API

More ...

- There are other archives (based at these centers), not necessarily NASA-funded, that follow this model.
 - Ex: Pan-STARRS, VLA-FIRST @STScI
 - Ex: Palomar Oschin wide-field survey @ IRSA:
 - Zwicky Transient Facility (2017+)
 - intermediate Palomar Transient Factory (iPTF; 2013-2016)
 - Palomar Transient Factory (2009-2012)
- There are (of course) many other non-NASA archives (SDSS, NRAO,...) and non-US archives (Simbad, ESA, ESO,...).
- Also, observers can deliver data back to these centers for distribution (which may include data beyond original program).





Lessons Learned ...

Ease of Access

- Researchers at all levels (team members, emeriti, summer students) need to be able to get and use data.
- Intuitive, web-based interface.
 - No extra software installation.
 - Visualization of resources, data, tools...
 - Easy choices to “just give me the table”, etc.
- Help needs to be there when users need it, easily found or promptly answered.

Support

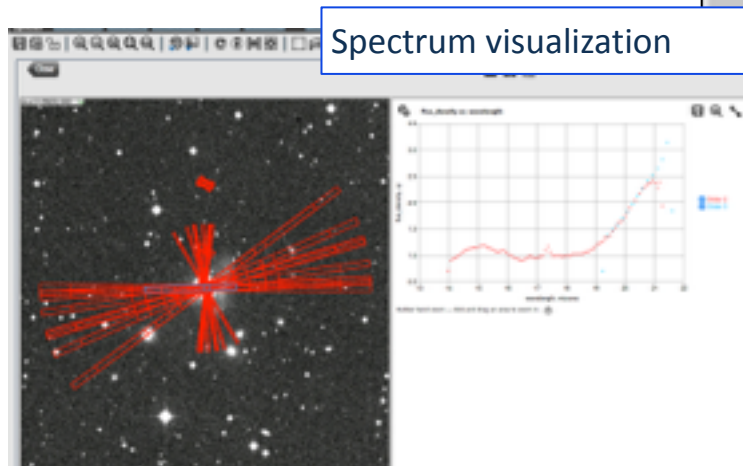
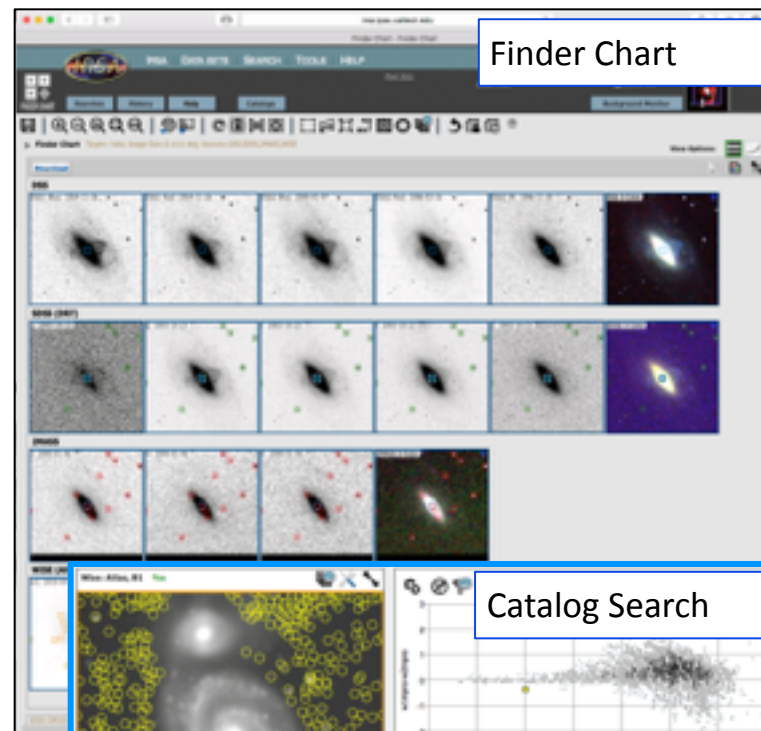


- Need to have knowledgeable staff, *who have done science with the data products*, who can (a) find problems; (b) pass on valuable experience to new users.
- Helpdesk:
 - Speed and accuracy matters!
 - Questions can be complex.
- Documentation:
 - Tools/data releases.
 - Documentation updates in response to tickets.
- Demos:
 - Live (AAS, ADASS, DPS, etc.).
 - Video tutorials (IRSA has >60 videos; >4500 views total).
- **The complexity of Science User needs increases with time.**



Visualization

- Data, catalogs, plots
- What do you have?
- What do I need?
 - What I know I need...
 - What did I just find?

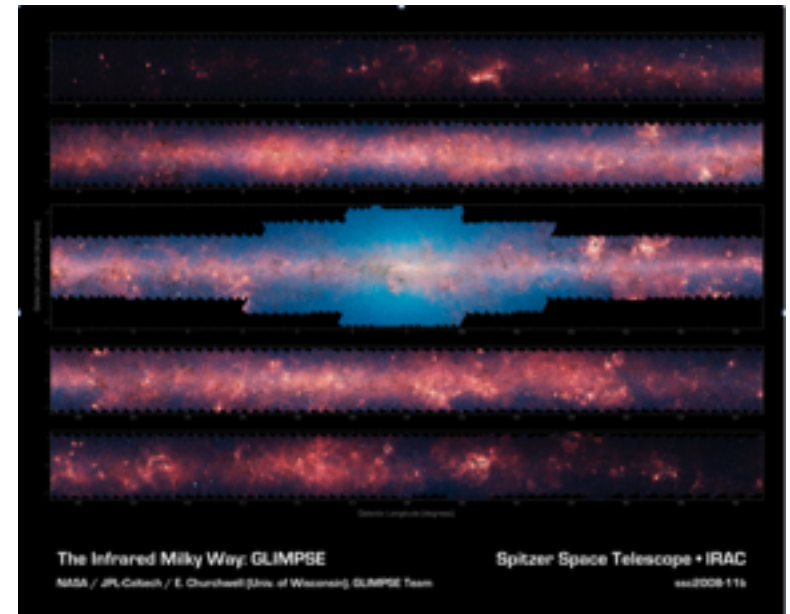


The image shows a software interface for catalog search. It features a scatter plot of 'redshift' vs. 'redshift' with a label 'Catalog Search' overlaid on the top right. Below the plot, there is a table of data.

id	name	ra	dec	z	mag	type	comment
100	NGC 100	10.0	10.0	0.0	10.0	galaxy	
101	NGC 101	10.1	10.1	0.0	10.1	galaxy	
102	NGC 102	10.2	10.2	0.0	10.2	galaxy	
103	NGC 103	10.3	10.3	0.0	10.3	galaxy	
104	NGC 104	10.4	10.4	0.0	10.4	galaxy	
105	NGC 105	10.5	10.5	0.0	10.5	galaxy	
106	NGC 106	10.6	10.6	0.0	10.6	galaxy	
107	NGC 107	10.7	10.7	0.0	10.7	galaxy	
108	NGC 108	10.8	10.8	0.0	10.8	galaxy	
109	NGC 109	10.9	10.9	0.0	10.9	galaxy	
110	NGC 110	11.0	11.0	0.0	11.0	galaxy	
111	NGC 111	11.1	11.1	0.0	11.1	galaxy	
112	NGC 112	11.2	11.2	0.0	11.2	galaxy	
113	NGC 113	11.3	11.3	0.0	11.3	galaxy	
114	NGC 114	11.4	11.4	0.0	11.4	galaxy	
115	NGC 115	11.5	11.5	0.0	11.5	galaxy	
116	NGC 116	11.6	11.6	0.0	11.6	galaxy	
117	NGC 117	11.7	11.7	0.0	11.7	galaxy	
118	NGC 118	11.8	11.8	0.0	11.8	galaxy	
119	NGC 119	11.9	11.9	0.0	11.9	galaxy	
120	NGC 120	12.0	12.0	0.0	12.0	galaxy	

Ease of Use: High Level Science Products

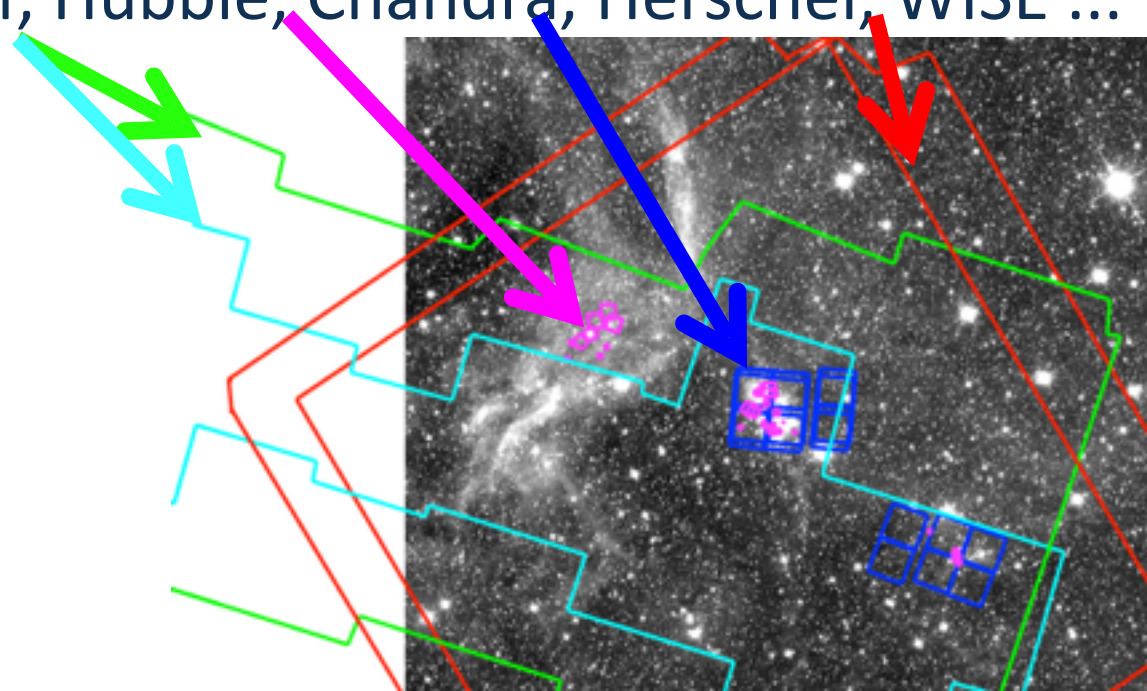
- *Greatly* enhance the science return of the archives.
- Hubble Legacy high-level science products (HLSP) are used 10x as much as typical pipeline products.
- Make complex data sets accessible to a wider audience of researchers.
- Expand the use of large, coherent projects:
 - Hubble Treasury
 - Spitzer Legacy, Exploration Sci
- Generated by the community or by the center.



Spitzer/GLIMPSE

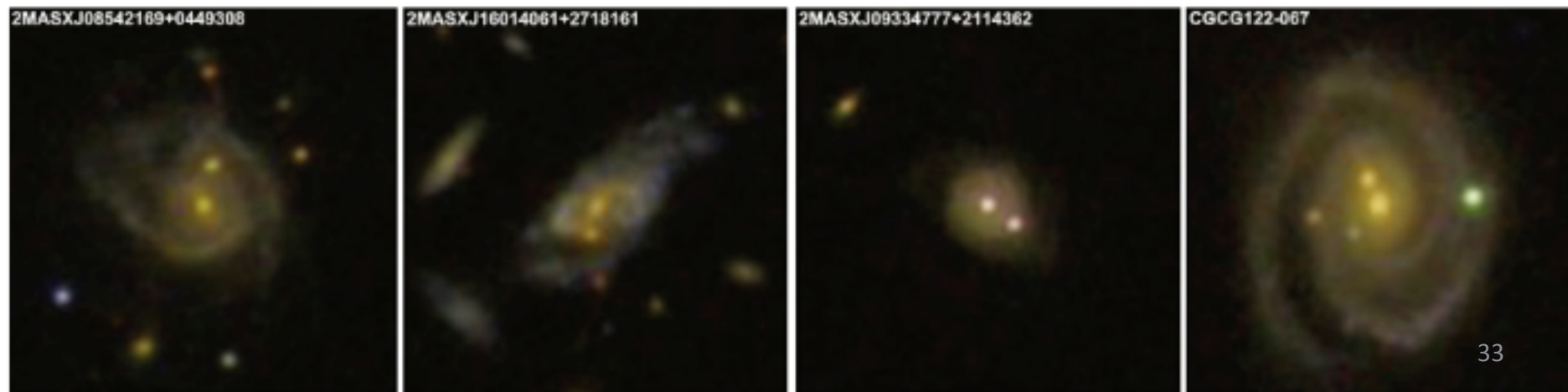
Ease of Use: Multi-Survey

- Combining information across wavelengths, surveys, missions...
- Source lists from entire missions:
 - Spitzer, Hubble, Chandra, Herschel, WISE ...



NED Science Example

- In context of assessing NED completeness, looking at fusion of GALEX, SDSS, 2MASS, WISE, ...
- Found super-luminous spiral galaxies!
- Ogle et al., 2016, ApJ, 817, 109



Changing Mission

- This is a result that came from looking at what was in the archive already.
- As data get bigger and bigger, won't be able to pull data out of the archive to work with it.
- Mission evolving from “search-and-retrieve” to “do [some] analysis in situ.”
- Science discoveries waiting in the archives that were never imagined or expected by the mission or even program PIs.

Community Feedback

- Need to talk to community to find out what is needed, wanted, wished for.
- Mission members, user committees, surveys, helpdesk, talking at conferences, and review cycles all feed into setting priorities.



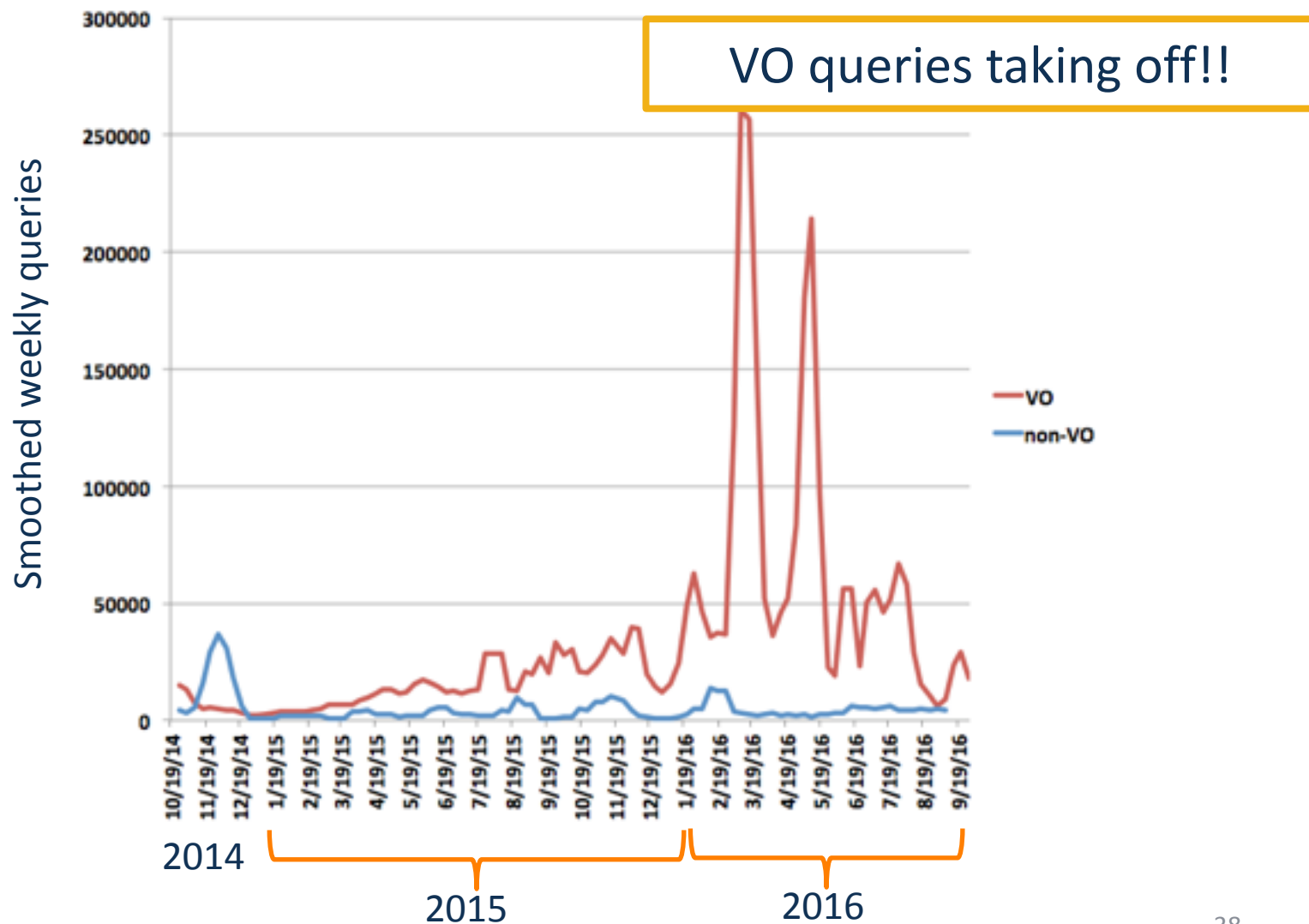
Commitment to Archival Research

- NASA as a whole (+ sometimes missions) explicitly funds archives, **and archival research**: NASA ADAP (Astrophysics Data Analysis Program).
- Well-designed archive & products can greatly enhance research value of the dataset.
 - Reducing barriers: finding data, making data accessible (reliability, units, file format, artifacts, documentation).
- NASA enables new ideas of things to do with older data.
- NASA has strong tradition of active collaboration between missions and archives.
 - Thinking about archive during the mission!

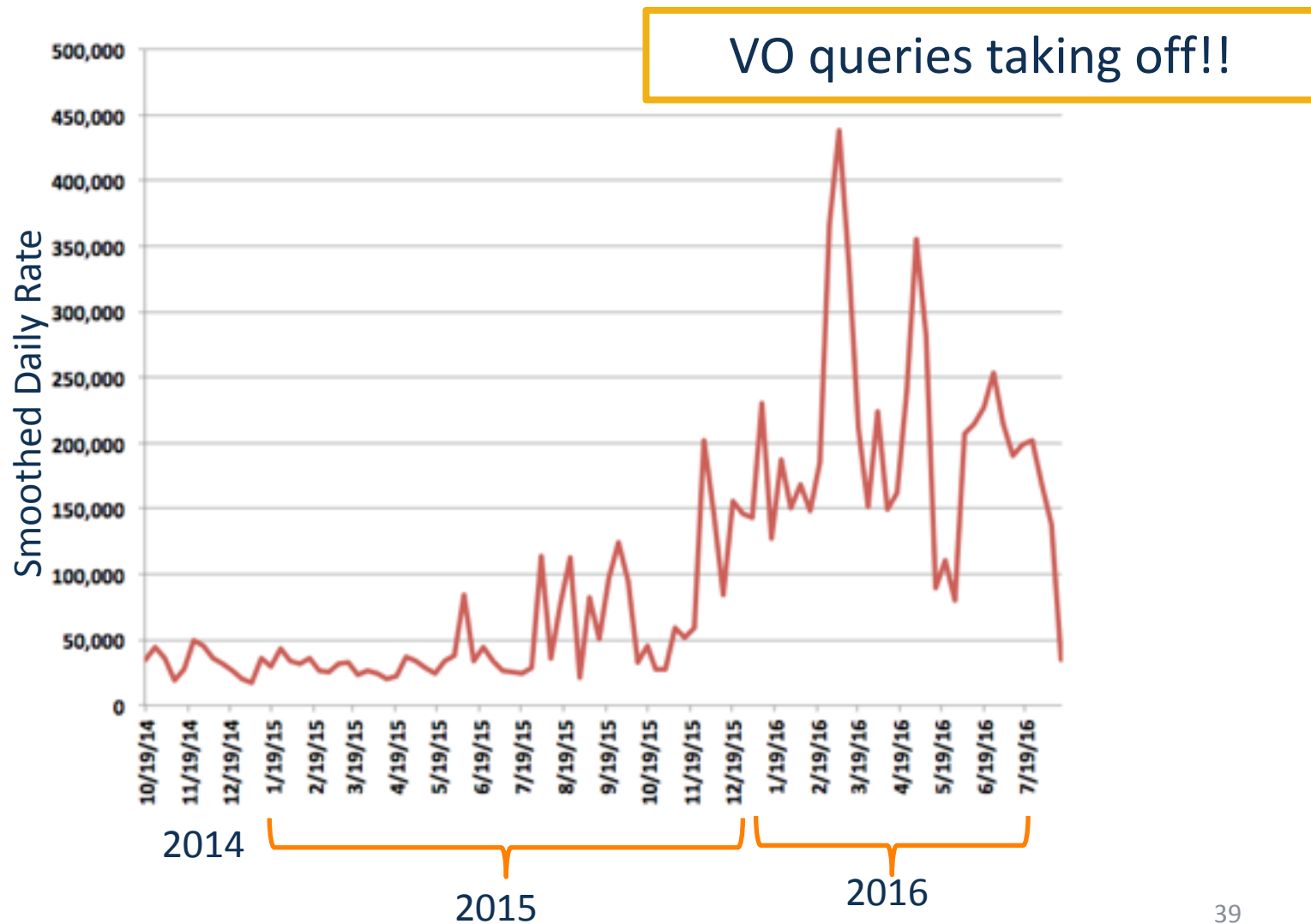
Ease of Use: VO

- VO= Virtual Observatory, IVOA=: International Virtual Observatory Alliance
 - Standardized *protocols* for interoperability between archives (i.e., NOT the applications that use the protocols).
 - Data discovery.
- Use interface you know, to get to data elsewhere.
- Interoperability of tools, within archives and across archives.
- (People) communication and collaboration across archives:
 - Astronomy Data Centers Executive Committee (ADEC).
 - US Virtual Observatory Alliance (USVOA).
 - NASA Astronomical Virtual Observatories (NAVO).
- NAVO:
 - Provide comprehensive and consistent access to all NASA data through VO protocols.
 - Coordinate NASA interactions with international and national VO communities.

IRSA VO Weekly Queries



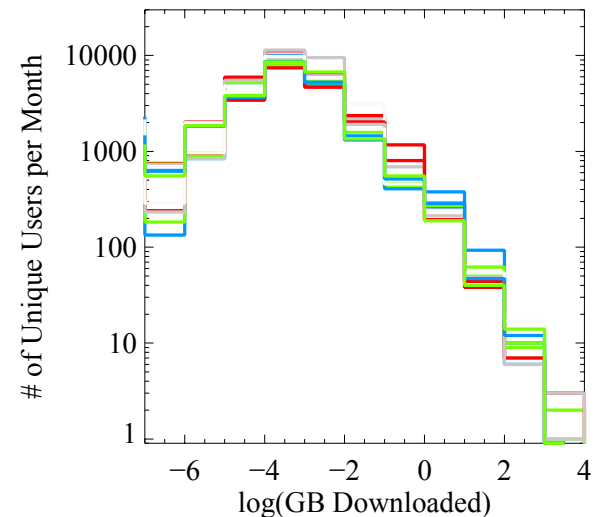
HEASARC+MAST+IRSA+NED



Ease of Use: API

- Application Program Interface.
 - (e.g., call by programs, scripts, command line)
- Allows scripted access to archive data.
- Enables complex projects.
- Enables rapid queries ...
- (also inadvertent DoS so need to watch and throttle!)

Lots of users → small requests
 Few users → enormous requests!



Operational Issues

- Keep archive running while making it better.
 - Assembling the plane while in flight!
- Growing audience, usage within existing resources.
 - Be efficient in how use resources.
 - Use the same software across multiple data sets (X. Wu, earlier).

Innovations

- Interactive UI, using pieces developed by others.
- Machine learning.
 - NED: data embedded in free-form text, tables not standard (e.g., RA, Ra, ra, R.A., ...); pilot project to apply ML to classify data and facilitate extraction.
- Improving scalability, extensibility, data prospecting...
- Greater integration of functionality and content across systems.
 - ADS: ORCID claiming; search by object via SIMBAD TAP service; embedding of publisher images via APIs.

Technical: Data Ingest

- Spitzer Legacy programs changed culture by requesting products be delivered back to the community.
- Now a common feature of Spitzer proposals.
- Brings these data to larger audience via central archive.
- IRSA has to have resources to ingest these products.

Technical: Data Ingest

- *Expense is not necessarily in TB but in education of the people delivering the products.*
 - Need to have delivery well-organized and documented.
 - People who have done it a lot: easy.
 - People new at this: not necessarily easy.
 - Tools to help people new at this.
 - Complexity is not just about size!
- Can end up with, e.g., optical and UV data available through the Spitzer archive (SINGS, LVL).

What's Next: Big Data

- “Big data” in some missions, certainly “big data” across all NASA Archives.
- To some extent, have already been working with “big data”!
- IRSA has already invested in data visualization services to help people identify and experiment with data quickly.
- Planning: identifying the most critical needs of users, including increased analysis at the archive facilitated by user workspaces.
- Richer services for in situ analysis.
- All archives thinking about this in some way.

Summary

- Long-term, stable archives greatly increase the return on observatory investment. (Doubles papers!)
- Robust support for both expert and novice users pays off.
- User support by instrument experts is crucial.
- Standardization of tools within an archive increases efficiency.
- Interoperability between archives benefits everyone.
- High level data products can expand the reach of large data sets.
- Shift in approach from “search and retrieve” to “analyze in situ”.