ARCHIVING GROUND-BASED DATA: PERSPECTIVE FROM SPACE

George Helou, Caltech

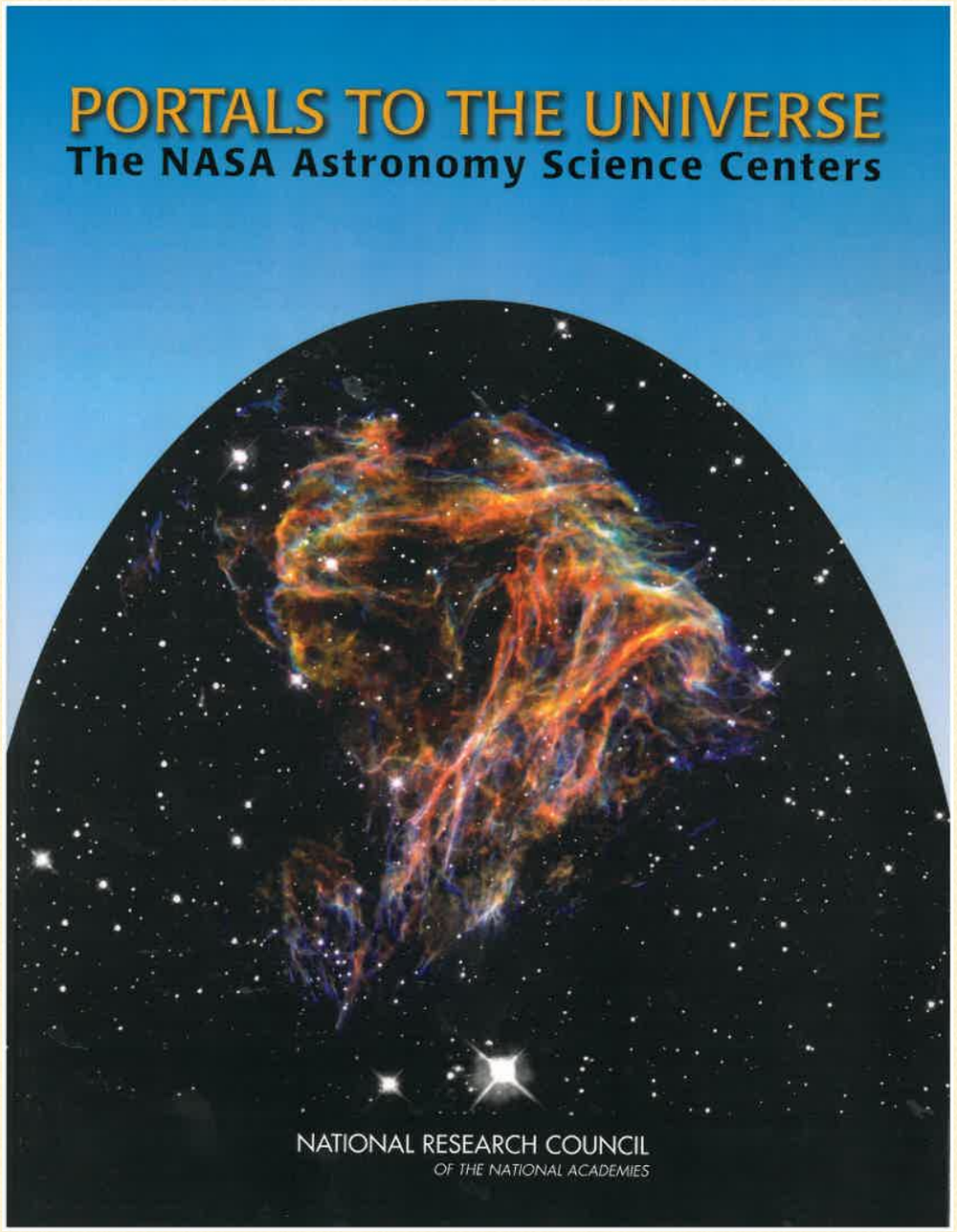
NRC OIR System Committee
Irvine, 12 October, 2014

PORTALS TO THE UNIVERSE

- *Portals to the Universe: The NASA Astronomy Science Centers*, NRC Report (2007):

- Science Centers can best process, store and disseminate their data if they maintain mission expertise at the archive centers for the long-term support of active users

- Successful research using archival data sets is dependent on the resident expertise and corporate memory that reside at the science centers.



QUESTIONS & ANSWERS-1

- What is the appropriate ratio of software/hardware cost? What's the key expense?

  - These numbers vary with project and phase

  - The key expense is science expertise & user support (>1/3 of total), then software (<1/2) and hardware (5-15%)

QUESTIONS & ANSWERS-2

- Discuss raw data vs. pipeline data, and the need for pipelines to be done by experts on the particular instruments.

- Without science-ready data, an archive will eventually become unusable and therefore useless. Therefore pipelines are essential

- Extensive scientific familiarity with the instrument is critical in guiding the development of science-ready data

- Expertise needs to be captured in software, reduced data (multiple products!), and documentation to ensure long-term usefulness

QUESTIONS & ANSWERS-3

- There's a lack of standardization for ground-based data because of the need to reprocess, specific to different instruments, different conditions, different science goals. How will this intermediate-level processing be possible? Where will it occur? (that is, data will be too large to download on individual computers for reprocessing). Is a coordinated effort needed – data centers, etc.?

  - The problem becomes tractable by limiting archive-supported instruments, modes, conditions. Surveys are the obvious extreme case to preserve, but support is not viable at the extreme case of idiosyncratic observations

  - The experience gained from archiving space data can contribute critical insights and lessons to ground-based archiving, e.g. instrument operations. Effective data archiving starts with instrument design

QUESTIONS & ANSWERS-4

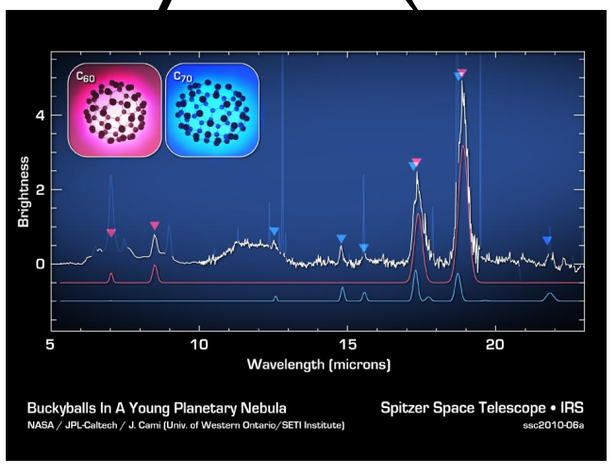
- How/where will archiving take place?
  - Policy addresses the goals and requirements; the answer to how/where depends more on practicalities than on policy.
  - The requirements are for data integrity, usability, longevity, etc. Implementing those may be best done by a distributed team, as for the Keck Observatory Archive, leveraging diverse expertise

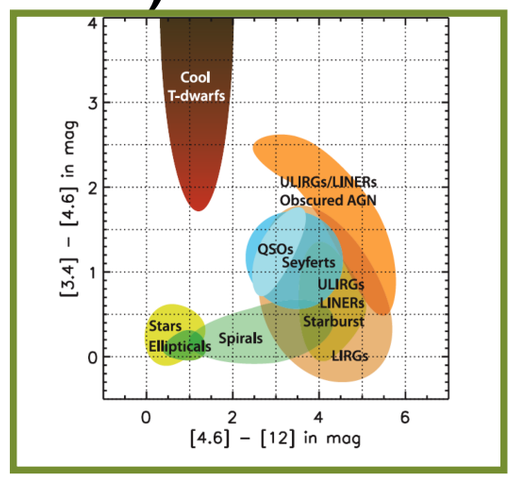
BACKGROUND: IPAC ARCHIVES

- IPAC is a multi-mission science and archive center, interfacing projects to community to enable science beyond the expected
  - The NASA center of science, data and operations expertise for IR-submm astrophysics, and of Exoplanet Science, especially **IRSA** (InfraRed Science Archive), **NED** (NASA Extragalactic Database), **Exoplanet** Archive, and **Keck Observatory** Archive
  - Also supporting **LCOGT** (archive), **PTF/ZTF** (processing and archive), **LSST** (Science User Tools)

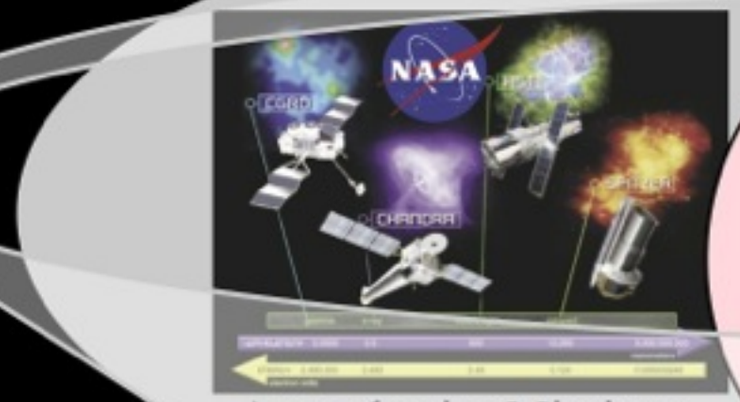
InfraRed Science Archive at IPAC

- Provides efficient access to NASA mission data
  - Began with IRAS and 2MASS, leading into the decade of IR missions
- IRSA serves "Heritage Archives" from Spitzer and other observatories, plus a unique array of all-sky IR surveys covering 20 bands from 1 μm to 10 mm
- Supports many research communities from Solar System (NEOWISE) to Cosmology (Planck)





Spitzer WISE Planck Herschel 2MASS IRAS AKARI MSX SWAS IRTS BLAST SOFIA



Astrophysics Missions

E-Literature

Large Sky Surveys

EXTRAGALACTIC DATABASE
NED
NASA IPAC

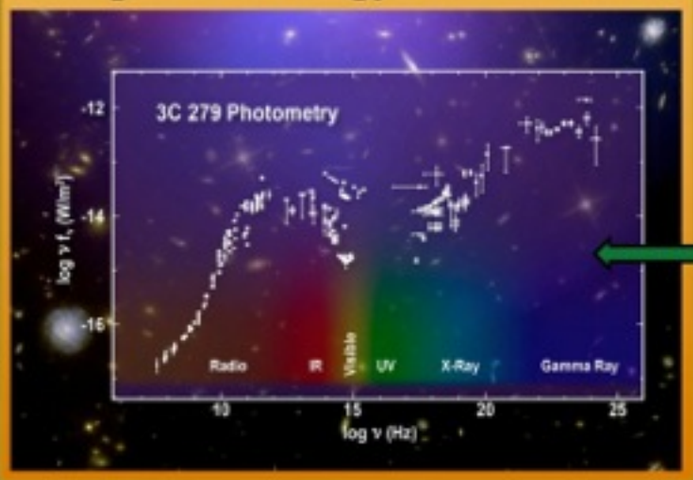
Overview - What is NED?

THE hub for multi-wavelength research on extragalactic objects

Unified multi-λ DB and Knowledgebase

- Standard Formats & Units
- Velocity Corrections
- Cosmological Corrections
- Distances, Luminosities, etc.

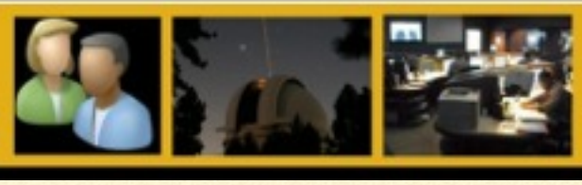
- Best available coordinates via any alias: Name Resolver
- Spectral Energy Distributions



f(λ):

- Name
- (α,δ)
- Redshift
- D_{Mpc}
- Flux
- Diameter
- Attributes
- References
- Notes
- Images
- Spectra
- A_λ

- Research scientists and educators
- Observation and mission planning



- Access via Web, email & computer
- VO program interfaces

NEW! PI Access to LWS Data and Public Access to ESI Data Now Available: Data Access

Funded by NASA, the *Keck Observatory Archive (KOA)* is a joint development between the W. M. Keck Observatory (WMKO) located in Waimea, Hawaii and the NASA Exoplanet Science Institute (NExScI) located in Pasadena, California. Currently, KOA archives data taken with HIRES, NIRSPEC, NIRC2, LRIS, KI (Keck Interferometer), MOSFIRE, DEIMOS, ESI, OSIRIS, NIRC, and LWS.

The process by which data are archived is as follows:

- Locate all FITS files taken in the previous 24 hour period
- Determine the validity and data integrity of each file
- Add meta-keywords to the FITS header
- Archive the metadata (native + meta-keywords) for each FITS header
- Archive the data for each FITS file
- Archive ancillary data associated with each night's data
  Ancillary data include Keck nightly plots of weather and image quality trends and, when available, all-sky images and photometric data

The physical location of the archive is at NExScI. Both metadata and raw FITS files are transferred via the internet from WMKO to NExScI, and are available through a User Interface within hours of the end of observations. Both the data files and the metadata are subject to a proprietary period policy. During the proprietary period, program PIs will be able to access their data via a user login/password protected site. The data are released to the public upon expiration of the proprietary period.