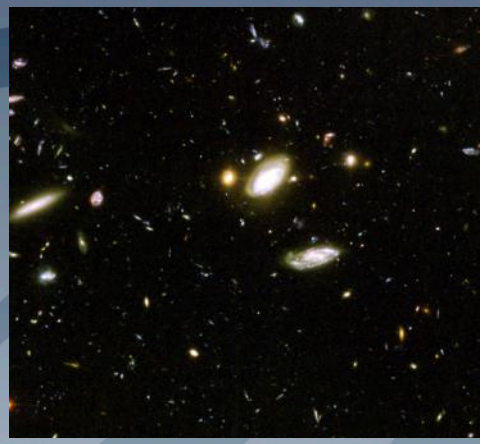
National Aeronautics and Space Administration

Astrophysics



Big Data Task Force of the NAC Science Committee
Washington DC
June 29, 2016

Hashima Hasan
Program Scientist,
Astrophysics Data Archives
Astrophysics Division
Science Mission Directorate

Astrophysics: Science and Data Archives

NASA Astrophysics supports an integrated system of science data archives for data from all recent and future Astrophysics missions:

Mission data with relevant tools are held at

- High Energy Astrophysics Science Archive Research Center (HEASARC)/GSFC (extreme UV/X-ray/gamma-ray data)
- Mikulski Archive for Space Telescopes (MAST)/Space Telescope Science Institute (primarily UV/visible/near-IR wavelengths)
- Infrared Science Archive (IRSA) at the Infrared Processing and Analysis Center (IPAC)/Caltech (infrared and sub-millimeter observations)

Specialist archives integrating published data for individual objects:

- NASA Extragalactic Database (NED) at IPAC/Caltech: fusion of multi-wavelength data and bibliography for objects beyond the Milky Way
- NASA Exoplanet Archive (NEA) at NASA Exoplanet Science Institute (NExScI)/Caltech: collating data on exoplanets and their host stars

Cross-archive activity:

- NASA Astrophysics Virtual Observatory (NAVO): consistent access to NASA data through international Virtual Observatory standards

The Astrophysics Data System (ADS)/Smithsonian Astrophysical Observatory: a digital library with bibliographic databases

http://science.nasa.gov/astrophysics/astrophysics-data-centers/

Reviews of Astrophysics Data Archives

- External Reviews of the Astrophysics Data Archives help NASA to optimize the scientific productivity that the Astrophysics archives enable, within the available budget
  - Reviews held in 2008, 2011, 2015
  - Next Review planned in 2019
- Findings from the 2011 Senior Review resulted in a major restructure of the exoplanet data holdings
  - Astrophysics division at NASA directed the NASA Exoplanet Science Institute (NExScI) to transition the Star and Extrasolar Planet Database (NStED) to a more focused service called the NASA Exoplanet Archive (NEA) and provided specific areas for this new archive to support. In addition, the NEA supports datasets of interest to future missions as directed by the Exoplanet Program.
- Challenges identified in the 2015 Review report:
  - The infrastructure and the technological approaches that are being used will certainly be obsolete at the end of the next 4-5 year review cycle.
  - Network bandwidths available to the data centers will soon be two generations behind the current standard for research internet.
  - Data centers need to raise concerns about sustainability where they exist, regardless of budgetary constraint

2015 Review of Astrophysics Data Archives (1)

Archives Reviewed: ADS, MAST, HEASARC, IRSA, NED, NEA

Motivation: As we enter an era in which Big Data management is undergoing rapid evolution, and the volume of data from NASA astronomy missions is increasing, it is important for the Archives to:

- keep abreast of new technologies and techniques;
- develop innovative strategies to manage in a cost effective manner;
- intelligently prioritize tasks and services;
- and continue to provide efficient community service.

Interoperability and coordination between the Archives is becoming increasingly more important as multiwavelength scientific investigations grow in number, scope, and complexity, as reflected by the NAVO infrastructure that the HEASARC, MAST, IRSA and NED Archives are now jointly managing.

2015 Review of Astrophysics Data Archives(2)

Review Committee Expertise: An eight member committee was appointed that covered a broad area of expertise which representation from science as well as archive management from the Earth Science and Planetary Science communities.

Measuring Progress: The Archives were asked to include in their proposals Prioritized Archive Objectives, against which their progress could be assessed annually.

Overall Evaluation:

- ADS – EXCELLENT
- MAST – EXCELLENT
- NED – EXCELLENT
- HEASARC – VERY GOOD
- IRSA – VERY GOOD
- NEA – GOOD

2015 Review of Astrophysics Data Archives(3)

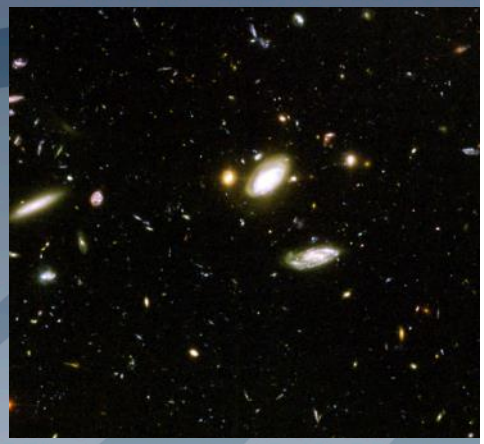
Response to Concerns Identified by Review:

- NASA augmented ADS budget to enable transition to new system with improved services and new functionalities
- MAST is giving a high priority to upgrading its network to increase the speed
- NASA given a modest augmentation to the NED budget to enable the proposed Physics within Galaxies program and the Machine Learning experiment.
- HEASARC and IPAC are cognizant of the issue of slow network speed and obsolescence of infrastructure and regularly refresh their infrastructure.
  - For HEASARC, the network speed will increase by an order of magnitude to 10 Gb/s when the Goddard Center Networking Environment is upgraded.
  - Key IPAC servers have been using 10 Gb/s for several years and this is being extended through IPAC's networking core and beyond.
  - IPAC is keeping up with technology and is experimenting with cloud computing

National Aeronautics and Space Administration

Astrophysics



BACKUP

ADS Objectives

- Maintaining a comprehensive, timely and complete database of the scholarly literature in Astronomy & Astrophysics
- Providing discovery services to support research in Astrophysics and related fields
- Promoting the use of NASA Astrophysics data by integrating bibliographies and links to data products generated by NASA missions, and hosted by NASA archives
- Providing services for curators and librarians involved in maintaining bibliographies, linking literature and data products, measuring bibliographic impact
- Interfacing with publishers and the community to facilitate the implementation of agency policy and government mandates related to Open Access publishing
- Making its efforts in software development freely available under an open-source software license

Overview of HEASARC

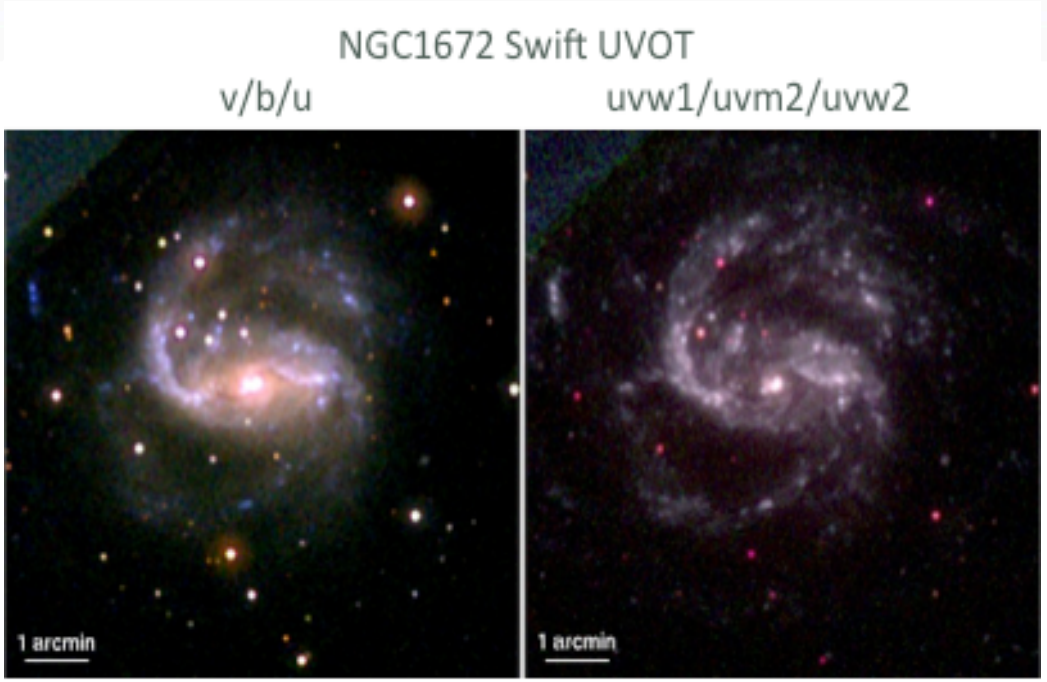
- Founded in 1990, the HEASARC is NASA's primary archive for High Energy Astrophysics and Cosmic Microwave Background data, supporting the broad science goals of the Physics of the Cosmos theme.
- Curates data from NASA, ESA, and JAXA space missions and associated suborbital experiments dealing with electromagnetic radiation from extremely energetic phenomena, ranging from black holes to the Big Bang.
- The HEASARC archive is now in excess of 85 TB, and is increasing in size by 10 TB/yr; in the coming 4 years will ingest data from 19 operating HEA and CMB missions/experiments, while serving data from 39 missions which are no longer operational.

Maintains mission datasets. Keeps them secure and accessible

Much more than just an archive, the HEASARC enables world-class science

Overview of MAST

- MAST is NASA's data center for optical/UV/near-IR data
- Active missions
  - Hubble, Kepler/K2, Swift UVOT, XMM-OM
- Legacy missions
  - GALEX, FUSE, IUE, EUVE, EPOCh, Astro, ORFEUS, Copernicus
- Future missions
  - TESS, JWST
- High-Level Science Products, DSS, GSC2
- MAST holds roughly 820TB of data



Overview of IRSA/NED

The Infrared Processing and Analysis Center (IPAC) is NASA's multi-mission center of science, data, and operations expertise for long-wavelength astrophysics. IPAC carries out challenging data processing tasks essential to the science return from large astronomy programs such as all-sky surveys and Great Observatories. IPAC is NASA's designated Science Archive Center for infrared-to-mm astrophysics.

The IPAC Archives, funded under the Astrophysics Data Curation and Archival Research program, consist of:

IRSA: The NASA/IPAC Infrared Science Archive

The archive for NASA Infrared-Submm Missions: 1600TB of data holdings

- **IRSA Charter:** Serve calibrated science products from NASA's IR-to-mm missions; provide tools for efficient access; and facilitate archival research

NED: The NASA/IPAC Extragalactic Database

NASA's "Google for Galaxies": 4TB of data

- **NED Charter:** Provide a comprehensive, reliable and easy-to-use synthesis of multi-wavelength data from NASA missions and the refereed literature, augmented with derived physical attributes, to enhance and enable astrophysical research beyond the Milky Way.

Overview of NEA

- The NASA Exoplanet Archive, consisting of a database of confirmed and candidate planets, numerous project and contributed data
- sets, and integrated analysis tools.
- The database currently includes interactive tables containing properties of all published exoplanets; numerous data products from the Kepler mission, including Kepler planet candidates, threshold-crossing events, data validation reports and target stellar parameters, and multi-quarter light curves; data from the CNES CoRoT mission and from
- several ground-based surveys; and spectra and radial velocity measurements from the literature.
- Tools provided to work with these data include a transit ephemeris predictor, light curve viewing utilities, and a periodogram service.
- The Exoplanet Archive directly supports present and future NASA missions (Kepler and K2 now, and in the future, TESS, JWST and WFIRST-AFTA) in line with NASA's long-term Strategic Goal to search for life on planets around other stars.

NAVO Objectives

- The NAVO mission is to facilitate the maximum science return for NASA astronomy data and resources using internationally agreed-upon standards through the International Virtual Observatory Alliance
- Develop comprehensive and consistent access to NASA data through Virtual Observatory standards
- Represent NASA and US interests in the development of astronomy data standards
- Maintain the infrastructure needed to discover and access Virtual Observatory services in the US and world-wide

NASA Astrophysics archives have maintained a common FITS standard since the 1990s: see W.A. Pence et al. A&A (2010) 525, 42 and fits.gsfc.nasa.gov/fits_standard.html