



The NASA Exoplanet Science Institute Archives: KOA and NStED

<http://nexsci.caltech.edu/>

G. Bruce Berriman

October 5, 2009
ADASS XIX, Sapporo, Japan



Exoplanet astrophysics is upon us ...

- NASA Exoplanet Science Institute (NExSci)
 - Science operations and analysis center for NASA's Exoplanet Exploration Program
 - Part of Greater Infrared Processing and Analysis Center (IPAC), Caltech
- This talk will describe two archives with large exoplanet content housed at NExSci
 - KOA – Keck Observatory Archive (Open July 2006)
 - <https://koa.ipac.caltech.edu>
 - Collaboration between NExSci and W. M. Keck Observatory
 - Archives data from the High Resolution Echelle Spectrometer
 - NStED – NASA Star and Exoplanet Database (Open July 2007)
 - <http://nsted.ipac.caltech.edu>
 - Serves validated parameters and observations of exoplanets and nearby stars, light curves from transit surveys
 - U.S. Data Center for *Convection, Rotation and planetary Transits* mission (CoRoT)



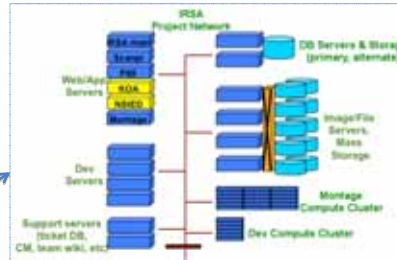
Building KOA and NStED on the IPAC Heritage



NASA Exoplanet Science Institute

- The NASA/IPAC Infrared Science Archive (IRSA)
 - Archive for NASA's IR and submm datasets since 1999
- IRSA architecture and infrastructure is highly re-usable by design
- NStED inherits from IRSA

- Science information system for processing user queries
- Configuration Management System and Code Repository
- Testing Methodology
- User Support Tools
- Mass-storage and server architecture



NASA Extragalactic Database (NED)

- Knowledge base for Extragalactic Astronomy since 1991
- NStED is a close stellar analog of NED
- Relational database design was used in the development of NStED
- Name resolver software was adapted
- Catalog and cross-identification expertise used to provide object name cross-ids
- Reference coding was used for tracking the literature (also used by ADS)

3



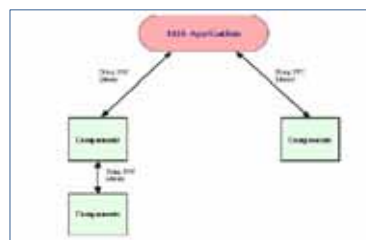
A Common Software Architecture with IRSA



NASA Exoplanet Science Institute


- IRSA uses a component based architecture which enables strong re-use and adaptation
 - All services are integrated into the *Infrared Science Information System (ISIS)*
 - Optimized for astronomical spatial searches and complex, general queries regardless of wavelength and type of mission
 - Over 25 million queries, >100 TB downloaded
 - NStED and KOA inherit functionality such as coordinate transformation and visualization tools
 - Design supports efficient development and controls maintenance costs
- Extended to support KOA & NStED
 - Complex proprietary data protection
 - Automated ingestion of data
 - Interfaces and return built dynamically from configuration files (150+)

Anatomy of an Application





Application is usually a CGI program

- Each component is a module with a standard interface that communicates with other components and fulfills **one** general function
- Modules are stand-alone portable ANSI-C tools
- Components plugged together & controlled by an executive library
- Executive starts components as child services & parses return values



Keck Observatory Archive


NASA Exoplanet Science Institute

- Raw data from the High Resolution Echelle Spectrograph (HIRES) from 1994 to date.
 - Raw data from 1994 – 2004 (single-chip CCD) recovered from magnetic tape at WMKO
 - Raw data from 2004 onwards (three-chip CCD) prepared for archive at WMKO and transferred to NExSci by afternoon following observations
- PI's have proprietary access to their data (including metadata) for at least 18 months after date of observations.
 - Data protected per CCD
 - "Row-based security"
- Data from 678 programs over 1,943 nights are public (200,000+ science CCD's)

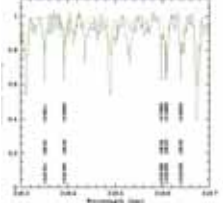
KOA HIRES Raw Data Holdings (August 2009)

Raw Spectra	Total
# Nights	2,218
# Science files	87,685
# Calibration files	135,136
Volume	3.1 TB

KOA Archive Queries (Cumulative)




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





Example of archival research with KOA

Bonifacio et al. (2009) A & A accepted. Abundances of extremely metal poor stars in a study of the chemical history of the Galaxy

Figure: Spectrum of CS 31085-024 with model fits superposed




Extracted HIRES Spectra at KOA

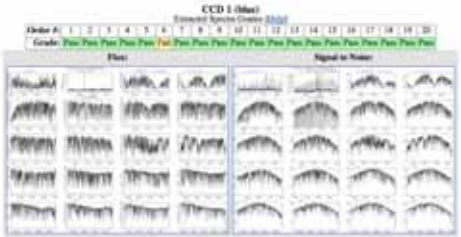
NASA Exoplanet Science Institute

- KOA developed an automated pipeline for the extraction of 1-D browse spectra from the raw science and calibration files
 - Upgrade to existing package (*makee* v5.2.x)
 - Wavelength calibration extended to 9800 Å
 - Improvements to cosmic-ray masking, trace-finding and slit-finding routines.
- Product contains:
 - 1-D spectrum *per order per CCD*
 - Automated extraction grade
 - Manual validation of grade
 - Diagnostic data including:
 - S/N
 - Traces and profiles
 - Preview images
 - Reduction log



5% of QA on Extractions are incorrect
 5% of Extractions are Bad
 93% of Extractions are good, with correct QA.

- Good Extractions, correct QA
- Good Extractions, incorrect QA
- Bad Extractions, incorrect QA
- Bad Extractions, correct QA



Extracted Spectra	Total
# Observations	64,343
# Files	31,074,591
Volume	1.84 TB

Summary of Extracted Spectra Holdings (Aug 2009) for single chip and three-chip data, 1994-2009.

NASA Star and Exoplanet Database (NStED)

NASA Exoplanet Science Institute

- Database of nearby stars and exoplanets
 - Data validated and traced to source by science staff
 - Stellar Services
 - Data related to relatively bright nearby stars
 - All known planet-hosting stars
 - Query for individual stars or by stellar/planetary parameters
 - Images and spectra
 - Exoplanet Services
 - Data related to known exoplanets
 - Time series data of exoplanets
 - Radial velocity and photometric light curves
 - Dedicated interface for exoplanet transit surveys
- Enables complex and detailed searches on potential and known exoplanet stellar hosts
- *Data visualizers use already existing IRSA tools.*
- *Pages built dynamically to accommodate changes and additions*

NStED Usage

Distribution of Sources by Type

- Small dots - dwarfs
- Large dots - exoplanet hosting stars
- Large plus signs - stars with radial velocity curves or photometric lightcurves;
- Open squares/diamonds - stars with images/spectra.
- Large circles - CoRoT "eyes".

NASA Star and Exoplanet Database (NStED)

NASA Exoplanet Science Institute


- Stellar Content for ~145,000 stars
- Exoplanets content for all known systems
- 250,000 light curves from transit surveys

Published Parameters	Derived Parameters	Associated Data
<ul style="list-style-type: none"> • Position, Distances • Kinematics • Photometry, Colors • Spectral type • Luminosity Class • Metallicity • Rotation • Activity Indicators • Variability • Multiplicity • Number of Planets • Planetary Mass • Orbital Parameters 	<ul style="list-style-type: none"> • Temperature • Luminosity • Radius • Mass • LSR Space Motion • Habitable Zone Sizes • Astrometric and Radial Velocity Wobbles 	<ul style="list-style-type: none"> • Images • Spectra • Radial velocity curves • Lightcurves of transiting exoplanetary systems • Jupiter and Earth sized transit depths



Exoplanet Transit Survey Content

Survey Region	Number of Stars	Time Series Filter	Additional Status Filters	Time Span (Days)	Number of Spectra
CoRoT Exoplanet	29278	CCD Wide light and 3 color Filter	B, V, R, A, I	20-750 days	1000 - 270000
Kepler 1	95	CCD Wide light	V	20-750 days	10000 - 420000
TRISight	27047	V, R	B, V	75	~10000
WELL-Precess	48837	R _c		75	~3000
NSIC 2001	1000	B	B, A	10	~100
NSIC 2002	10000	B, V	B, V	100	~100
NSIC 10	40000	B, V	B, V	300	~50
NSIC 2010	20000	B, V	B, V	300	~50

N2K spectrum of HD 10015



U. S. CoRoT Data Center

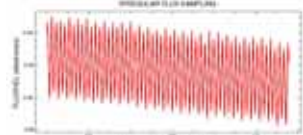
NASA Exoplanet Science Institute

NStED is the U.S. portal to the science data from the CONvection, ROTation & planetary Transits (CoRoT) mission.

- Agreement between NASA and CNES
- NStED provides an interface to the CoRoT data for U. S . Astronomers and will develop tools to support analysis of light curves; tools will have generally applicability

CoRoT Data at NStED

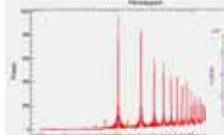
- NStED serves data from three releases – 30 asteroseismology stars, 28,079 exoplanet targets
- Complex spatial, temporal, photometric and stellar property searches
- Visualization of light curves



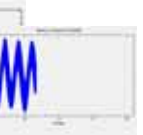
HD 50773

Light Curve Analysis Tools


- Web-accessible periodogram analysis tool
- Supports three algorithms: Lomb Scargle, Box Least Squares, binless phase dispersion minimization
- Parameters under user control





Periodogram



Phased light curve

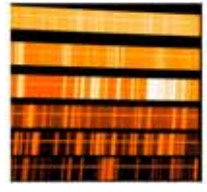


Near Future for KOA and NStED

NASA Exoplanet Science Institute

- KOA -
 - KOA will serve raw (level 0) data from the Near Infrared Echelle Spectrograph (NIRSPEC) in Spring 2010
 - Available with or without adaptive optics (AO).
 - High- (R ~ 25,000) and low- (R ~ 2,500) resolution modes.
 - Over 1,000 nights of observations from 1999-2009. (~250,000 raw science and calibration files)
 - Evaluating modification of *makee* pipeline, designed for HIRES, to extract NIRSPEC spectra.
- NStED -
 - Support new CoRoT data sets as released
 - Release light curve analysis tools, beginning with periodogram
 - Release exoplanet transit ephemeris service (based on Greg Laughlin’s “Transit Search.org”)
 - New data sets: M2K spectra (spectroscopic survey of late-type stars for planets), Spitzer Space Telescope exoplanet light curves, ...
 - New transit light curve data sets – Amateur Exoplanet Archive (AXA), Hungarian Automated Telescope Network (HATNet), EXtrasolar Planet Occultation Research of Open Clusters (Explore O/C), Wide Angle Search for Planets Prototype (WASP0), Vulcan, Burrell-Optical Kepler Survey (BOKS), ...
 - Augment archive by 300,000 light curves



Sample NIRSPEC L-band
high dispersion spectrum

10



Contributors to KOA and NStED



NASA Exoplanet Science Institute

KOA:

Tom Barlow *Caltech*
Bruce Berriman *Caltech*
Bob Goodrich *WMKO*
Stephen Kane *Caltech*
Mihseh Kong *Caltech*
Anastasia Laity *Caltech*
Christina Lau *Raytheon*
Jeff Mader *WMKO*
Melanie Swain *Raytheon*
Hien Tran *WMKO*
Myrna Tsubota *WMKO*

NStED:

Mark Abajian *Caltech*
Babar Ali *Caltech*
Bruce Berriman *Caltech*
Steve Berukoff *Raytheon*
Amy Beekley *Raytheon*

NStED cont.

Amy Beekley *Raytheon*
Geoff Bryden *JPL*
Kaspar von Braun *Caltech*
Benny Chan *Raytheon*
David Ciardi *Caltech*
John Good *Caltech*
Stephen Kane *Caltech*
Anastasia Laity *Caltech*
Christina Lau *Raytheon*
Meca Lynn *Caltech*
Doug McElroy *Caltech*
Peter Plavchan *Caltech*
Solange Ramirez *Caltech*
Moiria Regelson *Raytheon*
Michael Saucedo *Raytheon*
Marion Schmitz *Caltech*
John Stauffer *Caltech*
Pamela Wyatt *Caltech*
Angela Zhang *Caltech*