EPDS ARCHIVE AND DATA MANAGEMENT

Rachel Akeson
Deputy Director
NASA Exoplanet Science Institute (NExScI)
NExScI Role in EPDS

• NExScI is the NASA Exoplanet Science Institute and is the science operations and analysis center for NASA's Exoplanet Exploration Program

• As stated in the call, NExScI is expected to have responsibility for data processing and archiving after commissioning

• We have developed and are operating the following archives:
  • KOA – Keck Observatory Archive
  • LBTI Archive (no public data yet)
  • NASA Exoplanet Archive
EPDS archive

- All science data will be in the archive and will go public after a 12 month proprietary period
  - Includes GTO time
  - Engineering and commissioning data will be evaluated for format compliance and usability before being placed in archive
- Selected team will produce raw (Level 0) data
- Selected team will provide a pipeline to generate internally calibrated (Level 1) data
  - Includes bad pixel removal, dark and background subtraction, wavelength calibration, and flat fielding
- Archive will also include ancillary data (e.g. weather)
  - Coordinated with NOAO
Proposal requirements on data issues

- Step 1 and 2 proposals do not need to describe interface in detail (per call)
- NExScI will coordinate with selected teams during Instrument Concept Study to ensure submitted data pipeline plans meet requirements and work is properly scoped
- NExScI and NOAO will work with team selected for full implementation to develop data management plan (DMP) covering
  - Data transfer mechanisms and responsibilities
  - Level 0 format (FITS) and contents
  - Level 1 format (FITS) and contents
  - Pipeline requirements
Why discuss data content now?

- Although Step 1/2 proposals do not require pipeline and data details, our experience is that the final product is better when the data issues are considered early.
  - If keywords have to be added during commissioning, early data is less useful or even impossible to properly understand and archive.
Level 0 data header contents

- Header contents must include all information necessary to understand what data was taken and to drive pipeline processing
- Note that some of this information will have to be gathered from the telescope
- Telescope description
  - Location, date, telescope, image rotators
- Instrument state (anything set by observer should be recorded)
  - Filters, focus
- Observation parameters
  - Integration time, source name, sky coordinates, airmass
Level 0 (continued)

- Calibration type
  - dark, flat
- Calibration source information
  - Lamp type, settings
- Detector information
  - Gain, bias, readout mode

- KOA Examples:
  - HIRES
  - NIRSPEC [http://www2.keck.hawaii.edu/koa/public/keywords/NirspecKeywords.php](http://www2.keck.hawaii.edu/koa/public/keywords/NirspecKeywords.php)
Archive-added keywords

- The archive will also add keywords to allow for additional data searches
  - Program ID and PI information
  - Basic weather information (if available)
  - Added for both Level 0 and 1
Basic pipeline requirements

• Must produce instrument calibrated data using only files from the Level 0 archive
  • No configuration/instrument information should be stored elsewhere
  • Relevant calibration data identified through keywords and heuristics
• Should require minimal human intervention to process a full night of data
• Must have a scriptable mode, i.e. not GUI only
• Must provide some data quality metrics
  • Discuss as part of DMP definition
Level 1 example keywords

• Same telescope, instrument, source keywords as Level 0
• Processing keywords
  • Pipeline settings
  • Calibration files used
  • Derived wavelength calibration
  • Warnings or errors found during processing
• Data quality metrics (if available)
Archive interface

• User-specified, web-based searches on metadata, including
  • Source name and position
  • Program name
  • Observation date
  • Wavelength range
  • Instrument mode, exposure time, calibration type

• Level 0 data results
  • Science and calibration file lists
    • Relevant calibration files identified with same heuristics as pipeline
  • Files can be downloaded individually or as complete set meeting search criteria
  • Links to CCD data visualizations

• See koa.ipac.caltech.edu to try out searches on your own
KOALA advanced search interface
Example Level 0 visualization

- Example from KOA: HIRES data on 51 Peg

Available Quicklook JPEG Previews for KOAID: HI.20040819.42748.fits

- CCD: 1 (blue) (HDU 01)
- CCD: 2 (green) (HDU 02)
- CCD: 3 (red) (HDU 03)
Archive interface – Level 1 data

- Search returns file list and links to visualization of spectra
- Includes data metrics from Level 1 processing