Astronomical Data Analysis Software and Systems: XXIV
ASP Conference Series, Vol. 495
A. R. Taylor and E. Rosolowsky, eds.

Design and Implementation of Data Reduction Pipelines for the Keck Observatory Archive

Christopher R. Gelino,[1] G. B. Berriman,[1] M. Kong,[1] A. C. Laity,[1] M. A. Swain,[1] R. Campbell,[2] R. W. Goodrich,[2] J. Holt,[2] J. Lyke,[2] J. A. Mader,[2] H. D. Tran,[2] and T. Barlow[3]

[1]*NASA Exoplanet Science Institute, California Institute of Technology, Mail Code 100-22, 770 South Wilson Ave, Pasadena, CA 91125*

[2]*W. M. Keck Observatory, 65-1120 Mamalahoa Hwy., Kamuela, HI 96743*

[3]*California Institute of Technology, MC 249-17, 1200 East California Boulevard, Pasadena, CA 91125, USA*

Abstract. The Keck Observatory Archive (KOA), a collaboration between the NASA Exoplanet Science Institute and the W. M. Keck Observatory, serves science and calibration data for all active and inactive instruments from the twin Keck Telescopes located near the summit of Mauna Kea, Hawaii. In addition to the raw data, we produce and provide quick look reduced data for four instruments (HIRES, LWS, NIRC2, and OSIRIS) so that KOA users can more easily assess the scientific content and the quality of the data, which can often be difficult with raw data. The reduced products derive from both publicly available data reduction packages (when available) and KOA-created reduction scripts. The automation of publicly available data reduction packages has the benefit of providing a good quality product without the additional time and expense of creating a new reduction package, and is easily applied to bulk processing needs. The downside is that the pipeline is not always able to create an ideal product, particularly for spectra, because the processing options for one type of target (eg., point sources) may not be appropriate for other types of targets (eg., extended galaxies and nebulae). In this poster we present the design and implementation for the current pipelines used at KOA and discuss our strategies for handling data for which the nature of the targets and the observers' scientific goals and data taking procedures are unknown. We also discuss our plans for implementing automated pipelines for the remaining six instruments.

1. Introduction

The Keck Observatory Archive (KOA; `https://koa.ipac.caltech.edu`) archives data from all eight active and two inactive instruments on the twin W. M. Keck Telescopes in Hawaii. These instruments cover wavelengths from the optical to infrared and produce data sets such as images and single object, multi-object and integral field spectra. KOA serves raw science and calibration data for all instruments and, when possible, matches science data to the calibration files needed for off-line data reductions.

The scientific quality of raw astronomical data is often difficult to assess, however. For example, for ground-based data obtained at wavelengths longer than $\approx 1\mu$m, the radiation from the sky can be bright compared to the stars and galaxies that are of interest to archive users. Likewise, to the untrained eye, a raw spectrum provides almost no insight into the nature of an observation because the raw data contain artifacts from the sky and have no wavelength information. Therefore, it is highly desirable to provide users with a product from which they can

quickly determine the science content. To accomplish this, an automated processing pipeline is needed to calibrate the images to a level where the data content is easily discernible.

KOA currently employs automated pipelines for four instruments (HIRES, NIRC2, LWS, and OSIRIS) that produce browse quality data sets in addition to the raw data. In this paper we discuss the aspects of a data reduction pipeline that should be considered, KOA's current slate of instrument pipelines, and our future plans.

2. Pipeline Considerations

For an archive to run efficiently, data should proceed from the telescope, to the archive, to the user with little or no interaction from the archive maintainers. At KOA all processing steps from the data preparation and packaging to ingestion and distribution are automated. Since a data reduction pipeline is executed during this telescope-to-user processing stream, it also needs to be automated. Therefore, there are many aspects of a pipeline that need to be considered.

2.1. GUI or Command Line Driven?

KOA automates data ingestion through command line scripts. Any modules or DRPs that are added to the pipeline also need to be command line driven, rather than GUI-based, and require no interaction from archive maintainers beyond what is needed to initiate the processing. Although it is highly desirable to use an existing DRP as the basis for a pipeline, many of the publicly available DRPs to reduce data from the Keck instruments are driven by a GUI, require knowledge of the observing methods, or require user interaction. Obviously, a DRP that requires user interaction is not suitable for an automated pipeline. For the Keck instruments, the observers intentions for a particular data file are generally not known and the pipeline must work based on the information stored in the file headers or a metadata table. Of course that means that the metadata needs to sufficiently describe the data such that the pipeline knows *how* to process the data (Tran et al. 2014). For imaging data it needs to contain the filter (wavelength), exposure time, and sky position. For spectral extractions the metadata must also include information such as the slit width and length, which allows the DRP to determine how the spectra are aligned on the raw frame.

2.2. Output Content

Automated pipelines for astronomical data archives generally fall under one of two categories: those that produce fully reduced, science ready data and those that produce browse products allowing users to assess data content. Pipelines like the former are often limited to instruments for which the data collection follows a rigid and predictable set of templates such as for space-based telescopes like the *Hubble* and *Spitzer* Space Telescopes. Pipelines for these telescopes can combine the data from many exposures into a single, deeper exposure.

The Keck telescopes are classically scheduled–that is Principal Investigators (PIs) are assigned specific nights on which they personally execute their observations. The instrument settings observers use are usually constrained within acceptable boundaries, but the methods for obtaining data are not. With a near-infinite array of observing parameters, it is easier to produce a quick look browse product rather than a fully calibrated product. KOA employs this "quick look" approach to its pipelined products.

2.3. Robustness

The quality of data from a ground-based telescope is often limited by the atmospheric conditions (clouds, seeing, humidity, wind speed and direction) during the time of observation. A pipeline must be able to work with data of any quality. This is not an issue for imaging data, but can be problematic for extracting spectra from faint, fuzzy, or elongated objects. Of course, a good quality extraction may not be possible for some observations and it is important to provide a flag or quality assurance score to inform the user of possible extraction difficulties, as we do for our HIRES pipeline.

It is not possible to calibrate a raw science image without the proper calibration files (darks, flats, biases, etc.). Keck observers are not required to acquire calibration frames every night, so the pipeline must not rely on calibrations obtained the same night as the science data. For instruments that are very stable over several months or years, "master" calibration files can be used to process data from many nights. Again, since the goal is to facilitate an assessment of the science content and not produce an absolutely calibrated image, a master calibration file is often sufficient.

3. Current Pipelines at KOA

KOA has pipelines in place to create browse products–that are served along with the raw data–for four instrument modes. Each of these pipelines presented unique challenges.

- **HIRES:** The HIRES pipeline is the modification of a more general purpose DRP called MAKEE (`http://www.astro.caltech.edu/$\sim$tb/makee/`). MAKEE was upgraded to support a new 3-CCD detector and to produce diagnostic output files for assessing the quality of the automatic spectral extractions. MAKEE is in wide use by the HIRES community and has over 100 citations.
- **NIRC2:** No suitable DRP existed for NIRC2 imaging data so one had to be written for the KOA pipeline. The script determines the appropriate dark and flat calibration files from information provided in the data headers, but is unable to identify a sky frame that is needed for NIRC2 data. Nonetheless, the pipeline creates an improved product that facilitates the assessment of the data.
- **OSIRIS:** OSIRIS is an integral field spectrograph that obtains a grid of spectra over a small field of view. Although raw data are stored in a 2-axis FITS file, it is very difficult to assess the caliber of the science content of the data. A DRP exists (`http://www2.keck.hawaii.edu/inst/osiris/tools/`) that processes the raw data into a data cube that provides spatial and wavelength information. The DRP consists of a set of scripts that can be executed either via GUI or command line–what's needed is a parameter file providing the list of calibration frames and processing options. For the KOA pipeline (Holt et al. 2014), in most cases we are able to determine these calibration files for a particular file directly from information in the FITS header.
- **LWS:** Raw LWS imaging data are 6-dimensional data files in which entire observing sequences (image repeats, small dithers, sky frames) are stored. A quick look representation of the raw data is impossible, so we process the data to produce a single, 2-dimensional data file. The basic code to create this product already existed (`http://www2.keck.hawaii.edu/inst/lws/lws-idl.html`); we modified it to return a FITS file rather than a floating point array.

4. Future Plans

Automated data reduction pipelines are desired for all Keck instruments and their modes (most spectrometers can obtain images and most imagers can obtain spectra). We are exploring the various DRPs publicly available for these instruments in order to determine the ease with which they can be generalized and automated for inclusion in our processing pipeline. Below is a summary of the high-level modes that are present for the Keck instruments and our plans on DRP automation

- **Near-IR Spectra:** No DRPs exist to process NIRC2 or LWS spectra. MOSFIRE has two DRPs, but both either require knowledge of the observing sequence or use a GUI to drive the software. NIRSPEC DRPs heavily rely on user interaction during the extraction process so we have started work on a new automated NIRSPEC DRP (expected implementation in Summer 2015).

- **Near-IR Images:** There are no publicly known DRPs that are designed to process images from NIRC, the OSIRIS Imager, MOSFIRE, and NIRSPEC. Given the similarities to NIRC2 data, it is likely that we will modify the NIRC2 script to accommodate these modes.
- **Optical Spectra:** DEIMOS, ESI, and LRIS can obtain single object and multi-object spectra. Many different DRPs exist to process such data and we are currently examining the ease with which they can be automated in the hope that one will be used in our pipeline.
- **Optical Images:** No DRPs are known that process images from DEIMOS, ESI, and LRIS. Fortunately, optical image processing has much in common with near-IR image processing, so we will investigate the application of an adapted NIRC2 pipeline for this mode.

Acknowledgments. KOA is collaboration between the W. M. Keck Observatory (WMKO) and the NASA Exoplanet Science Institute (NExScI). Funding for KOA is provided by NASA. WMKO is operated as a scientific partnership among the California Institute of Technology, the University of California, and NASA.

References

Holt, J., Tran, H. D., Goodrich, R., Berriman, G. B., Gelino, C. R., & KOA Team 2014, in Astronomical Data Anaylsis Softward and Systems XXIII, edited by N. Manset, & P. Forshay, vol. 485 of Astronomical Society of the Pacific Conference Series, 127

Tran, H. D., Holt, J., Goodrich, R. W., Mader, J. A., Swain, M., Laity, A. C., Kong, M., Gelino, C. R., & Berriman, G. B. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 9152 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2. 1408.5187