

## Data and Metadata Management at the Keck Observatory Archive

G. B. Berriman,<sup>1</sup> J. M. Holt,<sup>2</sup> J. A. Mader,<sup>2</sup> H. D. Tran,<sup>2</sup> R. W. Goodrich,<sup>2</sup>  
C. R. Gelino,<sup>1</sup> A. C. Laity,<sup>1</sup> M. Kong,<sup>1</sup> and M. A. Swain<sup>1</sup>

<sup>1</sup> *NASA Exoplanet Science Institute, Infrared Processing and Analysis Center,  
California Institute of Technology, Pasadena, CA 91125*

<sup>2</sup> *W. M. Keck Observatory, 65-1120 Mamalahoa Hwy, Kamuela, HI 96743.*

**Abstract.** A collaboration between the W. M. Keck Observatory (WMKO) in Hawaii and the NASA Exoplanet Science Institute (NExScI) in California, the Keck Observatory Archive (KOA) was commissioned in 2004 to archive data from WMKO, which operates two classically scheduled 10 m ground-based telescopes. The data from Keck are not suitable for direct ingestion into the archive since the metadata contained in the original FITS headers lack the information necessary for proper archiving. The data pose a number of challenges for KOA: different instrument builders used different standards, and the nature of classical observing, where observers have complete control of the instruments and their observations, lead to heterogeneous data sets. For example, it is often difficult to determine if an observation is a science target, a sky frame, or a sky flat. It is also necessary to assign the data to the correct owners and observing programs, which can be a challenge for time-domain and target-of-opportunity observations, or on split nights, during which two or more principle investigators share a given night. In addition, having uniform and adequate calibrations is important for the proper reduction of data. Therefore, KOA needs to distinguish science files from calibration files, identify the type of calibrations available, and associate the appropriate calibration files with each science frame. We describe the methodologies and tools that we have developed to successfully address these difficulties, adding content to the FITS headers and “retrofitting” the metadata in order to support archiving Keck data, especially those obtained before the archive was designed. With the expertise gained from having successfully archived observations taken with all eight currently active instruments at WMKO, we have developed lessons learned from handling this complex array of heterogeneous metadata. These lessons help ensure a smooth ingestion of data not only for current but also future instruments, as well as a better experience for the archive user.

### 1. Introduction

The Keck Observatory Archive (KOA) curates data acquired by all facility instruments at the W. M. Keck Observatory (WMKO) since their date of commissioning, extending for some instruments as far back as 1994, when the observatory began operations. The metadata are produced by instruments built by different teams, and by observers with a large degree of flexibility in executing their observing programs. Thus the metadata are heterogeneous and are neither designed with modern archiving in mind nor intended for public release. The data files may, for example, contain duplicated UTC keywords,

and *unknown*, *null*, and *invalid* values in the keywords. A major task for KOA has been to rectify this state of affairs for all the data.

We will describe here some specific examples of how KOA has retrofitted data sets, and describe methods by which KOA has used the expertise developed in archiving data to benefit operations at the telescope and to support ease of use of the data sets. A thorough exposition of these topics is in Tran et al. (2014) and references therein. KOA produces calibrated browse-quality data sets ("level 1" products) for the HIRES, NIRC2 and NIRSPEC instruments by automating existing pipelines and leveraging existing calibration algorithms, but these will not be described here; see Tran et al. (2014) for details.

## 2. Retrofitting and Processing Level 0 Data

A data evaluation and preparation (DEP) process at WMKO makes all raw ("level 0") data for each instrument archive-ready by adding all necessary metadata as FITS keywords; the process incorporates instrument-specific knowledge as needed. The metadata added include a unique file identifier, as well as the PI and program information; these are essential in protecting the proprietary rights of PIs, who have exclusive access to their data for at least 18 months after the date of observation. Other fields added include instrument status, and the file type: whether it is a science or calibration file, and the type of calibration file.

### 2.1. Image Typing

The type of data, whether on-sky science target or calibration, is not always explicitly recorded in the FITS headers. Thus an important part of the DEP process is to perform this "image typing." The process is custom to each instrument. The KOA team has developed algorithms based on the construction and operation of each instrument.

For HIRES, NIRC2 and NIRSPEC, KOA exploits the retrofitted set of keywords to match the calibration files with the science files and thereby determine the optimum set of calibration files for each dataset. This process of "calibration association" involves matching the instrument and detector configurations within pre-defined tolerances. These calibration associations are returned to users as part of the query results, if requested.

### 2.2. World Coordinate System (WCS) keywords

In general, the imaging data obtained at Keck do not contain World Coordinate System (WCS) information. As part of data preparation, KOA added WCS keywords to the FITS headers for all NIRC2 data. The WCS information is determined from the keywords representing Right Ascension (keyword RA), Declination (DEC), image rotation angle (ROTPOSN), using an instrument PA zero-point (INSTANGL) of  $0.7\text{Å}$  with a correction of  $-0.252\text{Å}$  for the narrow camera, and the pixel scales in the FITS header. In addition, for vertical angle and stationary rotator modes, the parallactic angles (PARANTEL or PARANG) and elevation (EL) are also used. In vertical angle mode, the field is maintained at a fixed angle relative to the vertical (elevation) axis, while in stationary mode, the rotator is not tracking but remains stationary at a fixed rotator angle.

OSIRIS was delivered with a data reduction pipeline to enable astronomers to manage the reductions of the complex output from this instrument. KOA uses this pipeline to serve WCS information in the headers of the reduced ("level 1") data.

KOA plans to add WCS keywords to the imaging data from the LRIS, DEIMOS, and ESI instruments when the full set of keywords is ingested in the future.

### 3. Tools For Managing Metadata

#### 3.1. Observer's Tools: PIG and CAT

We have provided observers at WMKO tools for use during observations, which ensure that a complete set of metadata are written at the telescope and which streamline the observing process and observatory operations. For data currently obtained at the telescopes, the *Program Identification GUI* (PIG) (formerly *Split-Night GUI*) sets certain keywords (i.e., OBSERVER for the observer(s) name(s), and OUTDIR for the output data directory) in the header to facilitate the assignment of data to the correct PIs. A new release of the PIG supports programs that are not scheduled, such as "targets of opportunity" (ToO) programs, which interrupt scheduled observing for observations of transient phenomena. With a list of approved ToO programs from the time allocation committees, part of the GUI expands to allow selection of one of these programs once an interrupt is triggered. This process allows ToO data to be assigned to the proper PI and written into a new directory. The PIG logs all actions from interface input to monitor changes in the above keywords. These logs can then be parsed for data assignment.

Another tool is the *Calibration Acquisition Tool* (CAT), a simple interface that allows observers to easily select and acquire the necessary calibrations for the science data all at once. With one press of a button, observers can obtain a sequence of calibrations, rather than manually reconfiguring the instrument for each type of calibration. Observers select the calibration types, the exposure time, and the quantities for each type and click *Go*. There is an option to shutdown the instrument after the calibration acquisitions are finished, so observers can retire after their observing sessions in the morning without having to wait for the sequence to complete. This tool ensures that the necessary calibrations for the science files are available and uniformly obtained and recorded in the FITS headers.

#### 3.2. File Translator

The file names produced from WMKO instruments are not by design standardized. To overcome the ambiguity that resulted in file names, KOA standardized the names. These unique identifiers, called KOAIDs, consist of a two letter code representing the instrument, the UT date of observations (in the form YYYYMMDD), and the number of seconds after UT midnight. These KOAIDs are, however, unfamiliar to WMKO PIs and observers, and not friendly to data reduction tools that may expect the original names of the files as they are written at the telescope. Thus KOA provides a UNIX script that uses FITS keywords in the header to translate the KOA filenames back to the original filenames that PIs and observers more readily understand.

#### 4. Lessons Learned

We summarize here the lessons learned in archiving the Keck data; analysis of the previously acquired data was particularly valuable in this regard. This experience can help us anticipate what to expect with new instruments, and to take proactive steps to avoid potential problems. Some of the main issues encountered include:

- **Definition of keywords.** A number of times we found that the nature of the keyword changed for a particular instrument, after KOA had specified the structure of the metadata in an interface control document (ICD). This required changing the ICD, and often changing the code that handles that specific keyword. To address this issue, for NIRC2 (and subsequent instruments), we performed a complete inspection of all headers, rather than a subset as done for NIRSPEC and HIRES, in order to bring potential problems to light early in the ingestion process.
- **Split nights and backup.** Early on, an algorithm was developed that attempted to properly assign each science file to the proper PI, sometimes a challenge for nights that are split between multiple PIs. However, this was originally done on a file-by-file basis, with no greater context from the entire stream of files taken on that night. An improved algorithm was evolved, and with the addition of observing tools like the PIG, this smoothed out handling nights split between multiple PIs, those with unscheduled ToO programs or with instruments used as backup but not on the telescope schedule.
- **Complete keywords to support archiving.** Perhaps the best lesson is that in designing the instrument and its data product, the data should be acquired with the archive in mind. As such, it should be ensured that the full complement of KOA keywords conform to proper standards and are incorporated in the FITS headers well before the instrument enters operations on the telescope.

**Acknowledgments.** The Keck Observatory Archive (KOA) is collaboration between the W. M. Keck Observatory (WMKO) and the NASA Exoplanet Science Institute (NExScI). Funding for KOA is provided by the National Aeronautics and Space Administration (NASA). WMKO is operated as a scientific partnership among the California Institute of Technology, the University of California, and NASA. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. NExScI is sponsored by NASA's Origins Theme and Exoplanet Exploration Program, and operated by the California Institute of Technology in coordination with the Jet Propulsion Laboratory. This material is based in part upon work supported by NASA under Grant and Cooperative agreement No. NNX13AH26A.

#### References

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, SPIE Conference 9152