NOAO SDM Document PL106/CDR-19

ODI Automatic Pipeline Data Products

R. Swaters[1], F. Valdes[1]

National Optical Astronomy Observatories
Science Data Management

V1.0: Oct. 17, 2011

[1]NOAO Science Data Management, P.O. Box 26732, Tucson, AZ 85732

Copyright © 2012 by the authors.

Table of Contents

List of Tables

1 Introduction

This document describes the data products generated by the Automatic Calibration Pipeline (AuCaP, also known as the Tier 1 pipeline). The table below describes the different data products that are generated by AuCaP and the content of each of these products.

2 Data Product Standards

AuCaP will generate FITS files, PNG images, and ASCII log files.

The detrended images, stacks, and map files will be FITS files. The FITS files will follow the FITS standard v3.0 as defined in [1]. FITS compliance will be verified through the fitsverify utility [2]. Image metadata is stored in the FITS header.

Catalogs will be in the form of a binary FITS table, following the FITS v3.0 standard and verified with the fitsverify. Metadata is stored in the FITS header.

Graphics files will be in PNG format. Metadata in the graphics file will be stored following the Astronomy Visualization Metadata Standard version 1.1 [3].

Log files will be in ASCII format. Log files will be collected at the end of pipeline processing and compiled into a single bzip2-compressed file. This file will also contain a README file with relevant metadata.

Each FITS files will inherit all keywords from the raw data file. In addition:

1. If multiple exposures are combined (e.g., calibration data, stacks), the header of the first image (in time) is inherited;
2. The pipeline may update keywords with more accurate values, or add relevant calibration keywords (e.g., the astrometric solution, the photometric zeropoint; see Section 4.1);
3. If needed, the pipeline will add keywords describing the observation type and the data product classification (see Section 4.2);
4. The pipeline will add provenance metadata (see Section 4.3);
5. The pipeline will add keywords describing the associations between its data products (see Section 4.4).

All FITS files the pipeline produces will be Rice compressed following the tiled-image compressed convention provided by fpack [4].

3 Pipeline-Generated Metadata

The raw ODI will contain metadata describing the instrument, the telescope, the observations, and many other parameters. These keywords are described elsewhere, and are not repeated here.

Table 1: AuCaP Data Products

Data product	**Description of content**
Detrended images	Fully calibrated data products, i.e., bias and dark subtracted, flat fielded, pupil and fringe corrected, dark sky flat applied, astrometric calibration, photometric characterization, and corrected for pixel size variations
Stacks	If more than one exposure of the same field is available, these are projected to the same tangent point and orientation and combined into a single, deep image
Data quality maps	Each data product is accompanied by a data quality map, containing integer codes, where zeros indicate valid data and values larger than zero indicate masked data; the value in the mask represents the reason the pixel is flagged, e.g., due to saturation, or cosmic ray hits
Weight maps	Each data product, including calibration data, is accompanied by a weight map; the weight for each pixel is the inverse of the variance, except for pixels that have nonzero values in the data quality map, for which the weight is zero; the weight map is constructed by propagating the uncertainties associated with each calibration step into the Poisson noise model defined for the raw data
Coverage maps	For stacks only, a coverage map provides the cumulative integration time for each pixel; because of e.g., dithering, gaps between detectors, missing cells that were used for guiding, the coverage map can be complex in appearance
Source catalogs	All science exposures, i.e., both detrended images and stacks, are accompanied by a source catalog derived automatically by AuCaP; this catalog will contain all sources detected at or above 5 sigma; its main use in AuCaP is for astrometric calibration and photometric characterization, but it may be useful to end users
Graphics files	All data products (science and calibration data) are accompanied by graphics files; these may be used by the portal to provide a visual representation of the data, and by operators when inspecting the data products
Calibration data	Calibration data are derived from both special calibration programs required by WIYN, and from static data taken by the observer if possible; most of the calibration data are derived by AuCaP (such as the zero, dark, dome flat, pupil and fringe templates, and the dark sky flat), whereas others are derived outside the pipeline (such as the bad pixel mask and the initial astrometric solutions)
Processing log files	The processing log files are intended for internal use and are designed to make it possible to review the details of the processing after the fact

3.1 Keywords Added or Updated by the Pipeline

Before the start of processing, AuCaP will do basic validation of the data to ensure that data can successfully be processed by AuCaP. In rare cases, there may be problems with the keywords in the header. If possible, AuCaP will correct the problem and proceed with processing.

As part of the normal AuCaP processing, the astrometric solution provided at the telescope is refined, and the standard astrometric keywords in the header are updated. Also, the RA and Dec of the four image corners are provided. AuCaP will also characterize the photometric zeropoint.

3.2 Observation Type and Data Product Classification

The observation type of a data product will be given in the OBSTYPE keyword. Permissible values of OBSTYPE include 'dark', 'dome flat', 'fringe', 'object', 'pupil', 'sky flat', 'twilight flat', and 'zero'.

The data product type will be given in the PRODTYPE keyword. Permissible values of PRODTYPE include 'catalog', 'dqmask' (data quality maps), 'expmask' (coverage maps), 'image' , and 'wtmap' (weight maps).

The meaning of the PROCTYPE keyword is as follows. 'Raw' data are data that have not been processed by AuCaP (note that metadata of raw data may have been updated). 'MasterCal' are calibration files generated by the pipeline, either from raw calibration data or from science data. 'InstCal' are data that have been calibrated with all the applicable and available calibrations to remove the instrumental signature. 'InstCal' data are not the result of combining individual images. 'Resampled' refers to detrended data that have been resampled to a standard orientation. 'Stacked' refers to a combination of 'Resampled' data.

3.3 Provenance Metadata

All pipeline-reduced data will include provenance information in the keywords given in the table below.

3.4 Associations between Data Products

The associations reveal the links between different pipeline data products. For example:

1. which calibration files were applied to a particular science expousure
2. which exposures contributed to a mastercal
3. which exposures contributed to a stack
4. to which exposures was a mastercal applied

To capture these associations, AuCaP adds the following keywords to the headers:

To convey the associations to the archive, a task is run at the end of the pipeline processing. This task harvests all the association information into a data summary. It is necessary to do this because

Table 2: Calibration and Charaterization Keywords

ASTRMCAT	Catalog used for astrometric calibration
CDi_j	Transformation matrix from pixel to intermediate world coordinate
CORNiDEC	Dec position of image corner i (i runs from 1 to 4, counterclockwise from the SE)
CORNiRA	RA position of image corner i
CRPIXi	Location of the reference pixel along the axis I (in pixels)
CRVALi	Value of the world coordinate at the reference point (in degrees)
CTYPEi	Name of coordinates along axis i
DEC	Declination of the center of the field of view (updated from pipeline astrometry)
DPTHADU	Depth in the image (in counts)
EQUINOX	Equinox for the celestial coordinate system (in years)
MAGZERO	Photometric zeropoint
MAGZERR	Uncertainty in photometric solution
MAGZREF	Reference photometric zeropoint
MAGZSIG	Dispersion in photometric solution
NAXISi	Number of pixels along axis i
PHOTBW	RMS width of the filter (in nm)
PHOTCLAM	Central wavelength of the filter (in nm)
PHOTDPTH	Photometric depth (in mag)
PHOTFWHM	Full width at half maximum of the filter (in nm)
PHOTMCAT	Catalog used for photometric calibration
PIXSCALi	Pixel scale along axis I (in arcsec/pixel)
RA	Right ascension of the center of the field of view (updated from pipeline astrometry)
RADECSYS	Name of the reference system
SEEING	Seeing across the image
SKYBG	Background of the sky (in counts)
SKYMAG	Magnitude of the sky (mag/arcsec^2)
SKYNOISE	Dispersion of the background noise (in counts)
WATi_nnn	IRAF-specific description of the nonlinear portion of the transformation from detector pixels to the world coordinate system (i represents the axis, nnn a number increasing from 001
WCSCAL	Was an astrometric solution found?
WCSXRMS	RMS in x-direction of astrometric solution
WCSYRMS	RMS in y-direction of astrometric solution

Table 3: Provenance Keywords

PIPELINE	the name of the pipeline software that processed the data
PLVER	the software version of the pipeline used to process the data
PLPROCID	the processing ID under which the pipeline ran when processing the data
PLFNAME	the name of the file as given by the pipeline
PLOFNAME	the name of the file as originally given during the observations
PLQNAME	the name of the dataset that was processed
PLQUEUE	the name of the pipeline queue from which the data was processed

Table 4: Association Keywords

DARKFIL	Dark calibration applied to the data
FLATFIL	Flat calibration applied to the data
FRNGEFIL	Fringe calibration applied to the data
IMCMBnnn	List of the nnn images that contributed to a stack or a mastercal
PUPILFIL	Pupil calibration applied to the data
REJECT	Was this exposure included in the stack?
RSPGRP	All exposures in a stack, including those rejected from the stack, share this value
RSPTGRP	All exposures that are stackable without resampling share this value
SFLATFIL	Dark sky flat applied to the data
ZEROFIL	Zero calibration applied to the data

not all the associations can be captured in the keywords. For example, keywords cannot capture all the exposures to which a particular flat was applied. An advantage of creating the data summary prior to submission to the archive is that the file names are not changed by the submission process.

The task that harvests the association is run as part of the pipeline, but it provided and maintained by the archive group to ensure that the data summary is compatible with the archive.

Submission of the data summary to the archive signals that the processing of a dataset by the pipeline is complete.

3.5 Other Keywords

In addition to the keywords described in the previous subsections, AuCaP may also add additional keywords that are not otherwise required, such as keywords describing aspects of the data reduction, or characterizing the data. All keywords added by the pipeline will be documented, both in the pipeline documentation, as well as with a brief description in the header itself.

3.6 Metadata for Graphics Files

The graphics files generated by the pipeline will contain metadata describing the observations, including basic information on the observations (e.g., instrument, telescope, date and time), astrometry, photometric characterization, and associations, following the Astronomy Visualization Metadata Standard.

4 Data Product Description

4.1 Image Data

The input data for AuCaP consists of a sequence of exposures. The data for each exposure is a set of 64 files, one file for each OTA. Each of these files is multi-extension FITS (MEF) file with 64 extensions, one for each of the cells. AuCaP will merge all the cells for a given OTA into a single FITS file. Thus, the automatic calibrated data that AuCaP generates will consist of 64 images. In those cases where there multiple exposures of the same target are available, AuCaP will reproject the data and create a stack. These data will consist of a single, large FITS file. AuCaP will not produce MEF files, but the archive will be able to ingest MEF files because of the raw data.

4.1.1 Calibration Data

Master calibration data produced by the pipeline from sequences of calibration exposures will be in the format of 64 individual FITS files, one for each OTA. Each FITS header will contain all original metadata for that OTA (except where updated by AuCaP), additional metadata generated by the pipeline for that OTA, and relevant metadata derived from all OTAs. The individual exposures that contributed to the final master calibration file will be provided in the header.

4.1.2 Detrended data

Detrended data produced by the pipeline will be in the format of 64 individual FITS files, one for each OTA. Each FITS header will contain all original metadata for that OTA (except where updated by AuCaP), additional metadata generated by the pipeline for that OTA, and relevant metadata derived from all OTAs.

4.1.3 Reprojected and Stacked Data

Reprojected data and stacked data will consist of one single FITS file. The FITS header will only contain metadata relevant for the combined image; any metadata specific for individual OTAs will be removed. For stacked data, only metadata for the reference exposure (usually the exposure with the best transparency) in the stack is used, and where necessary the metadata should be updated to reflect the properties of the stack. References to the individual exposures used, as well as weights, will be included. Reprojected data are an intermediate data product that is not made available to the user except by request.

4.2 Data Quality Maps

The data quality map provides integer-value codes for pixels that are not scientifically useful or are suspect. The format is a set of 64 individual FITS files. The metadata is inherited from the companion image data, except for format indicators (e.g., the PRODTYPE keyword).

4.3 Weight Maps

A weight map is generated for each data product, including calibration data. Pixels that are flagged in the data quality map will have zero weight. The format is the same as for its companion image data product, i.e., a single FITS file for stacked and reprojected data, or a set of 64 individual FITS files otherwise. The metadata is inherited from the companion stack, except for format indicators (e.g., the PRODTYPE keyword).

4.4 Coverage Map

The coverage map provides the total exposure time for each of the pixels in stack. Only pixels with valid data contribute to the stack, i.e., masked pixels do not contribute. The format is a single FITS file. The metadata is inherited from the companion stack, except for format indicators (e.g., the PRODTYPE keyword).

4.5 Catalogs

Source catalogs for images must only contain sources detected at or above the 5-sigma level. The catalogs will contain at least the x,y pixel position, the RA,Dec equatorial position, the pipeline-calibrated magnitude, and a shape/ellipticity parameter. The uncertainties on the sky position and magnitude will also be included. The format is a binary FITS table, one for each of the 64 OTAs. The metadata is inherited from the companion image data, except for format indicators (e.g., the PRODTYPE keyword).

4.6 Graphics Files

Graphics files are generated for each science image data product. The highest resolution graphics file to be delivered to the archive will be binned 2x2. The portal also needs lower resolution images for its interface (down to 200x200 pixels), and these are also provided. All different resolutions are packed into a single tarball. The lower resolution images will also be used for operator review. The graphics files will contain basic header information about each exposure, but need not contain the complete set of header keywords. At a minimum, the headers will include basic information on the observation (time, date, instrument, observatory, filter), data product classification, photometric and astrometric calibration. The header will also contain a means to associate the graphics file with the original image, so that other keywords can be found through association with the original image.

4.7 Log Files

Each of the processing steps in the AuCaP will produce a log file. At the end of the pipeline processing, all log files will be collected. In addition, a file named README will be generated with KEYWORD=values pairs that contain basic information on what the log files pertain to, including instrument, observatory, observing time and date range, identifiers of all images that were processed. All the log files and the README file will be tarred and compressed.

5 References

[1] FITS Standard Document,
http://fits.gsfc.nasa.gov/fits_standard.html

[2] FITSVERRIFY – A FITS File Format-Verification tool,
http://heasarc.gsfc.nasa.gov/docs/software/ftools/fitsverify/

[3] The Astronomy Virtualization Metadata Standard,
http://virtualastronomy.org/avm_metadata.php

[4] fpack & funpack – FITS image compression programs,
http://heasarc.nasa.gov/fitsio/fpack/