NOAO SDM Document PL105/CDR-18

ODI Automatic Calibration Pipeline Design Elements

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

National Optical Astronomy Observatories
Science Data Management

V1.0: February 17, 2012

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

Table of Contents

Purpose of this Document

This Document describes the design elements of the workflow logic and algorithms by which the requirements in the ODI Pipeline and Archive Science Requirements Document (PASRD) for the Automatic Calibration Pipeline (AuCaP, known as Tier 1 in the PASRD) will be achieved. References are made to these requirements throughout the document.

1 Introduction

This document describes how the Automatic Calibration Pipeline (AuCaP, also known as the Tier 1 pipeline) will meet the requirements as defined in the PASRD and the Use Case Workshops.

The document consists of a narrative describing the operation of AuCaP, with use cases quoted were relevant. This narrative is a high-level description of the operation of AuCaP, but does include a description of details of how the calibration steps are performed. However, it does not describe how the various components of AuCaP interact with e.g., each other, OGCE, XMC cat, or the archive.

2 Summary of Data Flow through AuCaP

AuCaP has a hierarchical structure, optimized for parallelization across a large number of processing nodes, each having multiple cores. This is achieved by a modular design of AuCaP (R3.7). The data processing in AuCaP will be organized in small scripts (also called modules, usually IRAF scripts) that address a very specific step in the data reduction. These modules are organized into NHPPS pipelines, which orchestrate parallelization within a multi-core processing node. The NHPPS pipelines are wrapped as OGCE services, which are parallelized across multiple processing nodes. How AuCaP is parallelized is described in [1].

AuCaP will be triggered by the data engine when all raw data have been received into the archive. AuCaP will retrieve from the archive (R3.3) all raw data and metadata it needs to process the data (R3.1).

AuCaP will operate on a dataset that contains all exposures from a block of nights. This block consists of a mix of calibration data and science data. A block typically consists of three or four nights because, as work with the MOSAIC pipeline has shown, better-quality data products can be derived when more science exposures are available. However, the number of nights in an observing block can be adjusted as needed in the data engine.

The exposures are sorted by observation type. Calibration data are processed first, and the resulting calibrations are stored in the calibration library so that the results are available when the science data are processed. On-sky observations taken in static mode (possibly as part of a standard calibration program, or simply science data taken in static mode) are processed next, and any calibration data derived from them (fringe template, pupil ghost template, dark sky flat) are applied to these data and stored in the calibration database as well. Finally, data taken in coherent or locally-guided mode are processed, using all available calibration data.

AuCaP will gracefully handle missing cells or detectors (R4.5). This feature is already present in the MOSAIC pipeline, which had to deal with missing detectors or detectors with no valid data. This feature is also needed to exclude the guiding cells from the integrated science data.

Upon completion of processing, all required metadata are generated, and the data are submitted to the archive (R4.20). AuCaP will generate standard FITS files for image data, and standard text files (ascii, xml, csv) as needed (R3.6, G3.1).

The flow of the data through AuCaP is orchestrated by dedicated components of AuCaP that sort and stage data, and communicate with other components. Because this document focuses on how AuCaP meets its design requirements, the orchestration components are not described in this document.

At some points during the processing, AuCaP will require interaction with an operator. In the early stages of AuCaP use, it is anticipated that the calibration products will be reviewed by an operator before being stored in the calibration library. As the experience with the calibration products increases, the quality assessment of the calibration data will be automated, although some products, such as the dark sky flat (based on experience with the MOSAIC pipeline) may continue to require operator review. Any calibrations flagged by the pipeline as being of low quality, as well as the final science products, will always be reviewed by an operator.

Each version of AuCaP will always run on the raw data in the same way (R4.2). If the user is requesting additional data products (UP2.1, UP2.2), these will be made available only to the user and are not archived. The standard AuCaP will always run on the data even if a user is requesting special data products (R4.2).

3 Calibration Data

3.1 Calibration Library

Another component of AuCaP is the calibration library (R3.4). In order to find the most suitable calibration data, AuCaP queries this library whenever it needs calibration data. The calibration library contains two kinds of calibration data. Static calibrations are calibration data that are fairly constant over time (such the higher order components of the astrometric solution, the crosstalk coefficients, and the bad pixel mask). These static calibration data are provided to AuCaP by others, such as instrument scientists, pipeline scientists, and pipeline operators. Calibration data that may vary over time (such as dark, bias, flat, fringes, and pupil ghost) are generated by the pipeline (R3.4) and stored in the calibration library. To make sure these calibrations are available when a block of observations is processed, AuCaP creates the calibrations before processing the science data. All calibration data will be public (R3.9). When needed, separate calibrations are stored, e.g., for different filters, read-out modes, binning (R4.4).

The calibration data are stored with associated data quality information. This information is used as part of the algorithm for selecting the ‘best’ calibration for each exposure. One data quality criterion for calibrations derived from multiple exposures, for instance standard bias, dark and dome flat sequences, is the number of exposures. The algorithm will be strongly weighted to selecting longer sequences, say a minimum of 5 exposures, but if only a single or small sequence is available AuCaP will still continue.

3.2 Creating Basic Calibrations

AuCaP is required to process calibration data in a consistent way (R4.1). To make this possible depends in part on the implementation of standard calibration plans, which are beyond the scope of this document. The pipeline will ensure consistent processing by close control of the software. If changes to the calibration software are needed, the pipeline will update the version information in its data products (R4.3).

AuCaP will identify bias, dark, dome flat, and twilight flat sequences. Each sequence of these calibration files will be combined by AuCaP into a master calibration file.

Processing of the calibration data will start with crosstalk correction (R4.6). The details of how crosstalk will impact the data will have to be evaluated with engineering data. Like is the case in the MOSAIC pipeline, the crosstalk correction will likely consist of scaling and subtracting those components that affect a particular detector. If no adequate correction is possible, the data quality map is updated to reflect this (R3.10). The QUOTA experiment (which used different controller electronics) showed very little crosstalk, so it is possible that there will be little or no crosstalk for ODI.

All calibration data will be trimmed and overscan subtracted. Overscan subtraction will be done either on a line-by-line basis, or by a fit to the overscan region, depending on the properties of the overscan. The noise may not increase by more than 20% (R4.7). Because of the large number of pixels used to determine the overscan, even for line-by-line overscan subtraction (based on our experience with the MOSAIC pipeline), the noise will increase by a much smaller amount.

Bad pixels (from the static bad pixel mask) will be replaced by local interpolation, and flagged in the data quality mask.

After outlying exposures have been rejected (e.g., abnormal noise, too many counts), bias and dark exposures will be combined into master bias and dark calibrations by taking the average at each pixel position while rejecting the highest and lowest values.

After outlying exposures have been rejected (e.g., overexposed or underexposed), the dome and twilight flats will be combined into master dome and twilight flats by taking the average at each pixel position while rejecting deviating pixels using sigma clipping. If the filter in which the flat was taken suffers from a pupil ghost pattern, then it is removed by dividing it out. This ensures that the pupil pattern is not incorrectly flat fielded out of the science exposures. Light contributed by the pupil pattern should be subtracted out of the science exposures. A minimum of three flats are required in a sequences for it to be accepted as a master calibration.

To ensure a consistent photometry across all detectors, the dome and twilight flats are normalized by the average across all flats. The dome flats will be used to update the bad pixels masks (R4.17). This will be done with a separate component in AuCaP, e.g., a stand-alone tool that is run by the operator in order to ensure consistent bad pixel masks (R4.2).

The quality of each the master calibrations created by AuCaP is evaluated, based on image statistics and comparison with references in the calibration library. Master calibrations found to be bad or of poor quality are identified for later operator review.

After the master calibrations have been generated, they are submitted to the archive for storage (R3.4). All master calibrations are public (R3.9).

4 Processing of the Static Exposures

Processing of the static exposures will start with the crosstalk correction (R4.6), as described in Section 3.

The crosstalk correction is followed by overscan subtraction, either line-by-line or by subtracting a fit to the overscan region, depending on the properties of the overscan. The noise may not increase by more than 20% (R4.7; see also Section 3).

Next, the data are trimmed to remove the overscan regions and the cells from a single OTA put together into a single image based on the pixel geometry of the underlying detector architecture. Non-imaging and guiding pixels are flagged in a data quality mask and the values set to an appropriate blank value.

Pixels that are not photometric, e.g., because they were in the static bad pixel mask, affected by saturation, or bleeding, will be replaced with a local interpolation, and added to the data quality mask.

AuCaP will then query the calibration library for the best calibration data. Generally, calibration will be selected from data taken during the same observing run, but it is possible that calibration data are selected from other runs. This is possible because calibration data are public (R3.9). When the best calibration data have been retrieved, first the bias is subtracted (R4.8), then the dark is subtracted (R4.9), and finally the dome (or twilight) flat is applied (R4.10). The data products will indicate which calibration data were used (R4.3).

After the basic calibrations have been applied, a single-image transient-event detection will be done (R4.15) using a moving block average, with the central pixel and a number of additional highest pixels excluded. Identification of cores of objects is avoided by excluding detected objects as cosmic ray candidates. This transient-event detection pass will be set to be conservative to not compromise the science data, and because there may be a second, more sensitive, detection pass for data taken with multiple exposures of the same field. Detected transients will be added to the mask but not replaced in the image (R4.15).

For each exposure the astrometric solution is determined (R3.5) by matching sources detected in the image (R4.19) against those in astrometric reference catalog. To select the reference sources, the telescope pointing information in the header is used. The current plan is to use the USNO-B catalog because of its complete sky coverage. The astrometric solution will be determined using the entire field of view (and not per OTA or per cell) to avoid discontinuities and to make use of the physical constraints provided by the focal plane layout. For low-density fields, AuCaP will determine the lower order terms (such as scale, offset, and rotation), while using the higher order terms from the astrometric solution in the calibration library. If the source density is high enough, AuCaP will be able to locally refine the astrometric solution. However, it needs to be established during commissioning whether these higher order terms are reliable enough, or whether it is better to use the astrometric solution from the calibration library.

The source catalog generated by AuCaP will also be used to determine the seeing for each images, and the variations of the seeing across the image (R4.19).

Even though determining a photometric zeropoint is not a requirement for AuCaP, it will compare the instrumental magnitudes for the detected sources against the magnitudes of the reference sources while determining the astrometric solution. This comparison will provide characterization of the photometric zeropoint. However, because the USNO-B catalog is the likely reference catalog, the photometric accuracy is limited to about 0.5 mag. During development of AuCaP, other catalogs will be considered to improve the photometric accuracy (UP2.5). It is possible that different catalogs are used for the astrometry and photometry to achieve the best possible photometry for a given field.

It is a goal (G4.3) to achieve better than 10% photometric accuracy on photometric nights by assigning a standard photometric zeropoint if the guide star variance data indicates a night was photometric.

For each image a variance map is generated (R4.16). The weight map (R3.10) can easily be calculated by taking the inverse of the variance map. The variance map is determined based on a Poisson model of the raw data and then propagating the uncertainties associated with each of the calibration steps.

5 Deriving Advanced Calibration Data from Static Exposures

Fringe and pupil templates, as well as dark sky flats, need to be derived from a set of static exposures. These can either be static exposures taken as part of a fixed calibration program, or they may be observations taken by an observer for their science or for further calibrations.

First, AuCaP will determine whether it is possible to derive a pupil ghost template from the dataset. This will be done using a heuristic algorithm that uses information such as the number of exposures and the presence of extended objects. The details of this algorithm will need to be determined during commissioning, but it is likely that the heuristics developed for the MOSAIC pipeline will be a good starting point.

If a template could be derived from the static exposures in the given block, by fitting a function with radial and azimuthal components, it is then stored in the calibration library (R4.12).

To remove the pupil ghost image from the static exposures, AuCaP will query the calibration library to find the best available pupil ghost template. Normally this template will have been derived from data taken during the same block, but if needed data from another block is used. This template is then scaled to each individual exposure, masking out sources and the pixels from the data quality map. The scaling of the template to the data can be done either locally (e.g., per OTA) or globally, i.e., across all OTAs. The latter option gives better results in the MOSAIC pipeline, so it is the anticipated default behavior for AuCaP as well. Finally, the scaled pupil ghost is subtracted from the data (R4.12).

Deriving the fringe template and applying the fringe correction happens in much the same way as the pupil ghost. First, a heuristic algorithm (using the MOSAIC version as starting point) will determine the exposures that are suitable for constructing the fringe template. Next, the fringe template is constructed by median filtering and fitting for the large scale structure and subtracting it to leave the fringe pattern. This pattern is stored in the calibration library (R4.11). Then AuCaP retrieves the best fringe template from the calibration library and scales the pattern amplitude it to the data (using a globally determined scale factor as default). Finally, the scaled template is subtracted from the data (R4.11).

For the MOSAIC pipeline, the dome flat fielded data are used to create a ‘dark sky flat’, which provides a correction for the difference between how the dome flat and the sky illuminate the detector. Applying this dark sky flat results in data that have better background flatness (R4.13). The expectation is that this is also the case for ODI.

Therefore, AuCaP will analyze the static exposures to determine heuristically (following the logic from the MOSAIC pipeline) which exposures are suitable for constructing the dark sky flat. This includes logic to decide whether the dark sky flat has a high enough signal-to-noise ratio to be used at the level of individual pixels, or whether it should be smoothed to provide only the large scale illumination correction. If a dark sky flat can be derived, it is submitted to the calibration library.

Next, AuCaP will query the calibration library to find the best dark sky flat, and apply it to the data.

Even though the pupil template, fringe template, and dark sky flat are derived from science data, they are calibration data and therefore public (R3.9).

As each of the advanced calibration steps is applied, the uncertainties associated with these steps are propagated into the variance map (R4.16).

After the pupil ghost and fringe pattern have been subtracted and the dark sky flat has been applied, the data are divided by the Jacobian of the astrometric solution to correct the data for the variations of the on-sky pixel size (R4.14). This ensures that the calibrated data can be used to derive accurate photometry. However, depending on the amplitude of the pixel-size correction, this correction may result in data with a non-flat background. Because the correction is based on the astrometric solution, it can be reversed at any time (R4.14).

6 Calibration of Coherent and Locally Guided Data

Like for the static exposures, described in Section 5, the processing of the coherent and locally guided data starts with the crosstalk correction (R4.7), done in the same way as for the static data. Overscan correction and trimming is also done in the same way as for the static data. However, because of the orthogonal transfer shifts done during coherent and local guiding, the pixels near the edges of the cells may shift ‘out of a cell’. These pixels are flagged in the data quality map (R3.10).

Any pixel that was shifted through a bad pixel will be flagged as bad. This is achieved by convolving the bad pixel mask with the shift history. Saturated pixels and bleed trails will be detected without considering the shift history. Flagged pixels will be replaced with a locally interpolated value.

Next, AuCaP will retrieve all necessary calibrations: zero (R4.8), dark (R4.9), flat (R4.10), pupil ghost template (R4.12), fringe template (R4.11), and the dark sky flat (R4.13) from the calibration library. Each of these will be convolved with the particular shift history for each exposure (R3.5), and applied to the data as described in Section 5. Which calibrations were applied to the data will also be stored in the metadata (R4.3).

After that, a single pass transient detection is done (R4.15). It is anticipated that the transient detection will be done in the same was as for the static data, but during commissioning the effect of the orthogonal transfers on these high-contrast events will be investigated.

In the following step the astrometric solution (R3.5, R4.19) and the photometric characterization are done. For the data taken in coherent guiding mode, the processing will be the same as described in Section 5 for the static data. For data taken in local guiding mode an additional step is necessary. Different blocks of cells and detectors are guided independently and hence have different shift histories. By assuming that the exposures are sufficiently long that the average of the shift history reflects the true position of the sources, the offset between the average and the initial starting point can be used to tie all data to a common grid. After that, the astrometric solution can be derived in the same way as the static and coherently guided data.

As described in Section 5, the matching sources found in both the processed data and the astrometric reference catalog will be used to obtain a characterization of the photometric zeropoint; it is possible that other catalogs will be used to improve the accuracy of the photometry.

While the calibration data are applied, the uncertainties associated with each of the steps are propagated into the variance map (R4.16). Finally, as described in Section 6, the fully calibrated data is divided by the Jacobian to ensure that the data are photometrically correct (R4.14).

7 Resampling and Stacking

After all the calibrations have been applied and the astrometric solution has been determined (as described in Sections 5 and 7), the data are reprojected to a common tangent point, with standard orientation (north up and east left) and fixed pixel size. During reprojection, sinc interpolation is used in order to preserve the noise properties. The use of sinc interpolation is one of the reasons bad pixels and transients are replaced with a locally interpolated value. If this were not done, the sharp edges of these bad pixels and transients would cause ripples during the re-projection because of the sinc interpolation.

After the data have been reprojected, AuCaP will determine, for observing sequences with more than one exposure at the same position on the sky, which data are of sufficient quality for use in the stack. Data with poor photometry are excluded from the stack. Users may want to use different criteria to create their own stacks, and this can be done in the Tier 2 pipeline.

Before data are stacked, the contributing data are scaled to a common photometric zeropoint, and weighted. The weights will be lower for images with higher noise, and for images with seeing significantly worse than the median seeing. This weighting strategy gives good results over a large range of observing programs in the MOSAIC pipeline. At a user's request, AuCaP may also generate stacks with other generally useful weighting strategies (e.g., for the best seeing; UP2.4).

These products will be made available to the user but are not stored in the archive (UP2.1). The user may select different weighting strategies in the Tier 2 pipeline.

Next, the pipeline will create a first-pass, so-called harsh stack. In this harsh stack, a median filter is applied. This removes any transients from the stack that is produced. However, the median filter may also affect the photometry. This is not a problem, because the sole use of the harsh stack is to serve as a reference. For all exposures in a sequence the difference between each exposure and the harsh stack is calculated. A source detection algorithm is then run on this image, and detected sources are flagged as transients (R4.15) in data quality map, and the transients events are replaced with a local interpolation. This transient detection method can detect cosmic ray hits, light trails from planes and satellites, and (in some cases) stray light. However, it will also very effectively remove moving sources. It is a goal to include detection of moving astronomical sources in the pipeline (UP2.3). Such software has already been developed for the MOSAIC pipeline.

With the transients now masked, a second stack is made, with the same scale factors and weights, but now the average of all exposures is used. The result is a stack, free of transients, that is photometrically correct.

8 Automated Calibration Pipeline Software

Most of the AuCaP processing modules will be written as IRAF CL scripts. Changes to these modules will be tracked with version control software such as CVS. If such changes result in a change in the output data products, the AuCaP version identification will be changed as well (R4.3).

The AuCaP software will be made available to users (G3.2), but there will be no support of the software to end-users.

9 Documentation

Modules will have in-line documentation (R3.8). This in-line documentation will be written in reStructuredText format, which is human-readable markup syntax that can also be converted to webpages and standalone documents.

Further documentation will be provided in the ODI Data Handbook and an NOAO/SDM document describing the details of AuCaP (R3.8). It is a goal to write a science paper on the ODI pipeline.

10 References

[1] F. Valdes, *ODI Pipeline Flow Design*, SDM Pipeline Document PL013, NOAO/SDM, Oct 2009, http://chive.tuc.noao.edu/noaodpp/Pipeline/PL013.pdf

11 Requirements and Derived Requirements from the PASRD

Below is the complete list of requirements, and derived requirements

Requirement 3.1. A software pipeline must exist in order for ODI to be a productive scientific instrument. The pipeline must process ODI data consistently, guaranteeing the accuracy of the calibrations. The pipeline must be able to take raw images and meta-data (OT shift history files, initial astrometric solution, etc.) from the telescope. The pipeline should also be capable of capturing relevant meta-data from any existing Tier 0 pipeline at the mountain. The pipeline must produce output that can be productively used by ODI observers.

3.1.1 Need Interface Control Document (ICD) for Pipeline to Tier 0 interface

3.1.2 Need version tracking

3.1.3 Need policy for versioning and rollout

Requirement 3.2. Continued, long-term support for the pipeline software must be provided for ODI users.

3.2.1 Need End-user operational support strategy

3.2.2 Need End-user software/UI support

3.2.3 Need software maintenance

3.2.4 Need long term hardware commitment

Requirement 3.3. A long-term archive that is tightly coupled to the reduction pipeline is required to meet the science goals expressed in the ODISRD. The raw data must be archived. There is a strong science case for storage of the detrended data as well as the raw data. In any case, it is essential that the pipeline be able to (re)-process data contained in the archive.

3.3.1 Need an archive

3.3.2 Archive must deliver input data to pipeline with low latency/high bandwidth

3.3.3 Need two plans for "on-the-fly" and archive of Tier 1, include costing

Requirement 3.4. The ODI pipeline software must have access to libraries of calibration files that are applicable for a finite time and are updated at definite periods. This will require the ODI

pipeline process to have access to multiple sets of PI data to generate calibration images from science images. These calibration images should be stored in an archive (possibly separate from the entity of the science archive). The pipeline must be able to produce relevant calibration files (such as combined Zero, Dark, Flat, Fringe, Illumination and Pupil images) from raw images to keep the libraries updated.

3.4.1 Need a separate calibration pipeline (must do dark, flat, fringe, illumination, and pupil ghost corrections)

3.4.2 Must be clearly defined calibration plan describing the acquisition and use of calibration frames

3.4.3 Archive must store calibration data

3.4.4 calibrations are valid for finite and definite time periods

3.4.5 pipeline needs ACLs to control access to calibrations

Requirement 3.5. The pipeline must be able to process data taken in the primary imaging modes of ODI—static imaging, coherent guiding, and local guiding. Video streams from shutterless photometry mode need to be archived for PIs to retrieve, but need not be processed by the pipeline. Images taken in Non-sidereal tracking mode will be treated by the Tier 1 pipeline in the same way as images taken with sidereal tracking and detrended accordingly, with the understanding that obtaining the same absolute astrometric accuracy as the sidereal tracking images may not be possible.

3.5.1 video must be archived

3.5.2 Need ICD for video

3.5.3 Need usage information from WIYN

3.5.4 Need to support all imaging modes (static, coherent guiding, and local guiding)

Requirement 3.6. The output of the Tier 1 pipeline (and indeed all processed output) must be in standard format—i.e. images must be in FITS format, and any catalogs or ancillary tables must be in FITS or VOTable format or another common and easy-to-interface format.

Requirement 3.7. The pipeline tasks should be modular in nature.

Goal 3.1. The input and output to each step for Tier 2 should be in a format that allows access by external programs (FITS or VOTable, for example).

Requirement 3.8. The pipeline tasks must be fully documented; that is, the algorithms used in each step of the pipeline must be described, as well as the input and output data specifications and elements of the meta-data that are used and produced in each step.

3.8.1 Need end-user documentation describing the action of the pipeline

3.8.2 Need documentation for development of new capabilities/bug fixing

3.8.3 Need documentation for ongoing maintenance

Goal 3.2. Users should be able to run the pipeline assuming access to the data and metadata. Therefore, the pipeline software should be available.

3.2.1 Pipeline software provided but no support provided

Requirement 3.9. Master Calibration files shall be public.

3.9.1 Need definition of Master Calibration file

3.9.2 Need download service

3.9.3 Need to tag observations with corresponding Master Calibration

3.9.4 Master Calibrations and metadata must be archived

Requirement 3.10. A weight map (and if possible a data quality map) should be provided with the science images in the pipeline and be part of the final data products.

3.10.1 Pipeline must create weight (uncertainty) and quality maps for each exposure

3.10.2 Pipeline must create weight (uncertainty) and quality maps for calibration frames

3.10.3 Must be a way to visualize weight and quality maps

Requirement 3.11. Images must be released from the pipeline to PIs and from the archive to the public continuously, as soon as they pass quality control criteria, rather than in coordinated data releases. Raw data should be available to the PIs upon ingestion into the archive.

3.11.1 Pipeline Scientist needs special UI for determining quality

3.11.2 Need definition of what quality control criteria are

3.11.3 Data/Metadata are made available as soon as possible

Requirement 3.12. Data distribution of the output of the pipeline and from the archive must be available at least via electronic transfer.

3.12.1 Data must be made available via some electronic form (e.g. http, ftp, etc.)

Goal 3.3. A media distribution center should be considered, as it would provide access to processed data to PIs who do not have fast connections to the Internet.

3.3.1 Come up with proposal and costing model for physical media distribution

Requirement 4.1. Calibration data should be obtained and processed in a consistent way as part of a standard calibration plan as envisioned in the ODI SRD. WIYN should develop policies and procedures to ensure that standard calibrations are acquired even if they are not required for the science goals of the PI observing. These policies are especially important if ODI operates in visitor mode.

4.1.1 Do what WIYN says with regard to the calibration plan

Requirement 4.2. Tier 1 data reduction should be applied in a consistent way, and after a validation period (1-2 years, possibly), long-term stability of data reduction recipes should be preferred. The Tier 1 archive data reduction processes should be controlled by WIYN, and data stored in the archive should be reduced using the standard pipeline even if the

PI opts for non-standard reductions. The metadata obtained at the telescope and the raw data should always be available to PIs who wish to do their own independent reduction.

4.2.1 Release new pipeline version is infrequent (following commissioning) and controlled by WIYN

4.2.2 WIYN monitors and determines what is put into the archive

4.2.3 Tier 1 must have a standard pipeline version

4.2.4 Raw data and metadata must be made available to PIs

4.2.5PI must make a request to pipeline operator for non-standard processing

Requirement 4.3. It must be possible to improve or add pipeline algorithms if necessary, and to reprocess all the data taken up to the point when major pipeline improvements are released. Data stored in the archive for both PIs and general users must have attached metadata clarifying which version of the pipeline has been used to generate it, and which calibration files were used.

4.3.1 Reprocessing must be possible

4.3.2 Metadata must be stored

Requirement 4.4. The pipeline must be able to keep libraries of standard calibrations that are appropriate for each allowed combination of readout, binning, and imaging mode.

4.4.1 must store libraries of master calibrations and associate science images with proper library

Goal 4.1. The pipeline should provide an operator interface that allows selection and generation of calibration files. This is particularly important early in the pipeline's lifetime when there will necessarily be re-evaluations of the extent of applicability of the master calibration files.

4.1.1 Must provide an operator interface

Requirement 4.5. The pipeline must be able to proceed in the absence of data from individual cells and detectors. Missing data should be masked out so that it does not interfere with subsequent processing.

4.5.1 Need masks

Requirement 4.6. ODI data will be corrected for crosstalk contribution from other cells, to the accuracy to which the relevant crosstalk coefficients can be calculated.

4.6.1 Need crosstalk correction

4.6.2 If not correctable should be flagged

Requirement 4.7. ODI data will be overscan corrected and trimmed. Overscan corrections should not leave residual noise greater than 1/5 of the readout noise. In particular, readout noise-dominated exposures (e.g. narrow-band images) must not be degraded by the overscan correction.

4.7.1 Images should be overscan corrected and trimmed

Requirement 4.8. ODI data will be Bias/Zero subtracted, using master bias calibration images generated from sufficient individual exposures to ensure that the master bias frame is not read-noise dominated but instead reveals the fixed pattern noise. The pipeline must be able to create these master bias frames for all supported readout modes.

4.8.1 Pipeline must perform bias subtraction

4.8.2 Pipeline must create master bias frames

Requirement 4.9. ODI data shall be corrected for dark current and other instrumental signatures that scale with exposure time. The pipeline must be able to handle both convolved and unconvolved dark frames.

4.9.1 Must be able to find appropriate dark frames

4.9.2 Must be able to scale dark frame

4.9.3 Must apply appropriate convolution (if needed)

4.9.4 Must apply dark frame

Requirement 4.10. ODI data shall be flat field corrected. For most filters, this will involve an iterative correction with the fringe and ghost pupil correction. The pipeline must provide software to remove pupil ghosts from flat fields and fringes from night sky flats. For narrow band filters, it will almost certainly be impossible to accumulate sufficient sky counts, so a dome flat field is likely to be necessary. For broad-band filters, the exact mix of dome versus sky flat calibration must be determined during commissioning. A sufficient number of flat field exposures must be combined by the pipeline for the flats to have a resulting $<<$1% pixel-to-pixel noise.

4.10.1 Must be able to remove pupil ghost from flat field

4.10.2 Must be able to remove fringe from flat field

4.10.3 Must be able to apply flat field

4.10.4 Must be able to remove pupil ghost from science images

4.10.5 Must be able to remove fringe from science images

Requirement 4.11. ODI data shall be corrected for fringing, where necessary. The pipeline must be capable of producing master fringe maps from static night-sky observations for the reddest filters (probably including the i' band and redder filters). The fringe maps will have to be convolved with the OTA guide history before being applied to images taken in local and coherent guide modes.

4.11.1 Fringe map must be convolved with OTA shift history

Requirement 4.12. ODI data shall be corrected for ghost pupil and other light reflections that do not vary with telescope orientation, where necessary. The pipeline must be capable of producing libraries of ghost pupil models. The pupil ghosts will vary with each filter and will have to be calculated and removed from both dome flat images and sky images. The pupil ghost model images will have to be convolved with the OTA guide history before being applied to images taken in local and coherent guide modes.

4.12.1 Pipeline must produce libraries of pupil ghost models

4.12.2 Pupil ghost model must be convolved with OTA shift history

Requirement 4.13. The pipeline should produce flatness of the background less than or equal to 1% over a significant fraction of the field of view, as required in the ODISRD, where the calibrations data allow it.

4.13.1 1% flatness if possible

Requirement 4.14. ODI data shall be illumination corrected. This is to be the default mode for the ODI pipeline, so that individual images may be photometrically calibrated. The image stacking software (in Tier 2) must be able to 'undo' this correction.

4.14.1 pipeline must apply an illumination correction

4.14.2 pipeline should be able to apply a pixel scale correction

4.14.3 Pixel scale correction must be reversible

Requirement 4.15. The ODI data pipeline will identify and flag cosmic rays in the individual exposures. Rather than correcting the image itself, affected pixels will be flagged in the data quality image, so that they may be ignored in resampling and stacking procedures.

4.15.1 Cosmic rays must be flagged on individual images only

Requirement 4.16. The ODI data pipeline must create a variance map, and as much as possible a data quality map, which will be associated with each observation.

4.16.1 must create a variance map

4.16.2 must create a data quality map

4.16.3 Will at least double storage needs

Requirement 4.17. The ODI data pipeline must be able to create bad pixel masks from sets of flat fields taken with different exposure times. These bad pixel masks must be periodically updated to account for degradation of the detectors.

4.17.1 pipeline must create static bad pixel mask from flat fields

4.17.2 static bad pixel masks should be able to be updated

Goal 4.2. The Master data calibration products that are produced by the pipeline should be made available for use in the ODI Tier 0 data pipeline, and possibly to any independent fast transient detection pipelines.

4.2.1 Tier 0 must have "fast" access to master calibrations

4.2.2 Independent fast transient pipeline access to master calibration

Requirement 4.18. The ODI data pipeline must update the World Coordinate System of each exposure. In particular, for exposures taken with OT guiding, the pipeline must correct for shifts in the WCS of each guided cell due to the OTA shift history. It is possible that this correction will be already undertaken in the tier 0 pipeline, in which case the ODI tier 1 pipeline must be able to apply this correction. Global pointing accuracy should be enabled by matching to the best available astrometric catalog for the location and filter

choice in question (e.g. USNO, 2MASS, SDSS, etc.).

4.18.1 Final WCS for each cell must be corrected based on best available astrometric catalog

Requirement 4.19. The pipeline must generate source (as opposed to science) catalogs for use by the pipeline in calculating astrometric and photometric properties, as well as for computing data quality flags. The catalogs will also be used to derive a map of PSF variations across the image, if this is not already available as part of the tier 0 metadata from the telescope. These catalogs must be associated with the exposure in the archive, although the expectation is that these catalogs should not be construed as final catalogs for PI or archival science.

4.19.1 Pipeline will produce a source catalog containing position and magnitude

4.19.2 Source catalog will be archived

4.19.3 Catalog must be associated with exposure

Goal 4.3. For data gathered during photometric nights, Tier 1 data shall be assigned a standard (average) photometric zeropoint in the image meta-data to allow photometric accuracy at better than the 10% level. Guide star variance data shall be quantified and used to measure non-photometric conditions.

4.3.1 pipeline should generate photometric zero point

Requirement 4.20. The Tier 1 pipeline must store the detrended image, with all its attached meta-data, in a long-term archive. Reprocessing by the archive pipeline is expected and it should replace the existing archive contents for that exposure. Dates of processing and calibration files used must be associated with the archived files.

4.20.1 evaluation of storing Tier 1 vs. computing Tier 1 "on the fly" both costing and performance

4.20.2 Processed data has been processed with the latest version ofthe pipeline

4.20.3 Dates of processing and calibration files along with other metadata must be associated with Tier 1 data

Requirement 4.21. The main reduction steps for Tier 1 pipeline (see appendix 2) and an archive capable of storing data from the pipeline and serving data to the pipeline shall be functional and ready for commissioning by the start of ODI science commissioning (expected to be in June 2010). Operational prototypes of many of the steps should be available when ODI lab-testing and engineering commissioning commence in April 2010. The individual steps detailed in the Tier 1 pipeline must all exist at the start of commissioning, and they must be able to process taken with successively more complicated observing modes as these modes become ready for commissioning.

4.21.1 Must work with ODI team to prioritize parts needed for commissioning

4.21.2 Need to find out what is critical for commissioning and what the commissioning plan is.

Goal 5.1. A Tier 2 photometric calibration should be capable of delivering up to 1% absolute photometric accuracy using the calibration data provided by WIYN, provided that the calibrating

dataset (either standard star observations or photometrically calibrated catalogs such as SDSS) support this level of precision. Note that this level of precision requires that a very detailed plan for calibrations be adopted by WIYN operations.

5.1.1 Standard star frames must be identified by WIYN to the pipeline

5.1.2 Frames should be flagged as to whether the night was photometric or not

Requirement 5.1. An astrometric distortion map shall be applied to each image that takes into account both the geometric distortion of the camera/optics and (if significant) any residual airmass-dependent terms (for example if the ADCs are not engaged for the observation).

5.1.1 Find out if this is different from req. 4.18

Goal 5.2. The goal of the geometric correction is to assign global coordinate values to all objects with an absolute accuracy per object better than 0.2 arcsecond ($<$2 pixels). This will allow the coordinates to be usable in slit and fiber spectrographs (with the possible exception of STIS) without additional processing.

5.2.1 Per object accuracy better than 0.2 arcsec in static mode

5.2.2 Per object accuracy better than 0.2 arcsec in local mode

5.2.3 Need a catalog that will accommodate this

5.2.4 Pipeline should provide astrometric accuracy no worse than the input catalog

Requirement 5.2. The Tier 2 pipeline software must include software to perform stacking operations on sets of dithered exposures taken in all of the supported imaging modes (with the possible exception of non-sidereal tracking).

5.2.1 Pipeline will provide a "standard" stack

5.2.2 Stacking must support all imaging modes

5.2.3 Standard stack is for dither sequences

Goal 5.3. The stacking software should be available for users to run on their own machines if resources allow, in order to reduce pressure on a central processing facility from PI requests.

5.3.1 Need a plan if central resources are maxed out

Requirement 5.3. The stacking software must allow for user input in the stacking, and therefore be available to PIs (and eventually to archival users, perhaps in the form of a resource queue) to be run on request.

5.3.1 Need a user accessible tool to produce individual resampled images

5.3.2 Need a mechanism for end user to have input into stacking procedure

Requirement 5.4. The stacking software must allow for differencing of pairs of images, be they individual images or the output of the stacking software, or indeed the comparison of a single image to a stacked template. This software is not anticipated to be run as part of the standard software, except for pre-defined ON-OFF sequences. Furthermore, it is not expected that the output of the software necessarily must be available on fast timescales (less than a

few days). However, as with the stacking software, the capability for PIs to run this software with user determined parameters (e.g. PSF matching) is a requirement for the software.

5.4.1 predefined on-off sequences part of standard processing

5.4.2 User must be able to select inputs for non-standard differencing

Requirement 5.5. Tier 2 must include software to allow filtering of images, i.e. convolution of detrended images and stacks with simple spatial functions. The software must allow for user input, and therefore be available to PIs (and eventually to archival users, perhaps in the form of a resource queue) to be run on request.

5.5.1 Must have filtering software

5.5.2 User must be able to have input into the filtering

Requirement 5.6. Tier 2 must include software to allow splitting of stacked as well as resampled individual exposures into multiple pieces for analysis by PIs with existing software tools. The slices should inherit the calibrations, data quality and variance images and the appropriate metadata. The WCS for the image slices must be properly transferred, and the provenance of the image must be recorded.

5.6.1 Must determine what standard download format is

5.6.2 Need to have slicing software and defined slices

Goal 6.1. The output of Tier 1 detrending should be available to PIs on the timescale of (a small number of) days, although it is possible and indeed expected that final tier 1 reprocessing will be necessary on the timescale of weeks to accommodate the availability of improved calibration files. For this reason, it is important that the output of Tier 1 calibration preserve information about the date of processing and the calibration files used.

6.1.1 Sufficient data for Tier 1 must get to Data Capacitor before processing can begin

6.1.2 Metadata (date and calibrations) must be associated with images

6.1.3 Must be able to reprocess Tier 1 with improved calibration data

Goal 6.2. WIYN should consider the possibility of a mechanism for requesting priority pipeline processing for projects that do not require immediate (timescale $<$1 day) processing, but for which speedy processing (1-2 day timescale) could prove scientifically essential.

6.2.1 Consider speedy processing

12 Feedback from Use Case Workshops

UP 2.1 Some of the pipeline algorithms have parameters with several reasonable variables (e.g., coordinate system, astrometric projection, stacking technique). Users would like the ability to specify some of these, and have the pipeline produce user-specific data products in addition to the defaults.

UP 2.2 Intermediate data products generated by the pipeline should be kept for some (TBD) time before deletion, to enable (1) users to retrieve them, if desired, or (2) effective diagnosis when users discover problems with their data.

UP 2.3 Techniques should be developed to identify scientifically interesting transients in ODI data.

UP 2.4 Several alternative methods of stacking dithered exposures should be adopted (and documented) and stacks produced using all of them.

UP 2.5 Attempts should be made to provide as accurate a photometric calibration in the pipeline as possible. Observing conditions should be documented/archived to support future archival research.