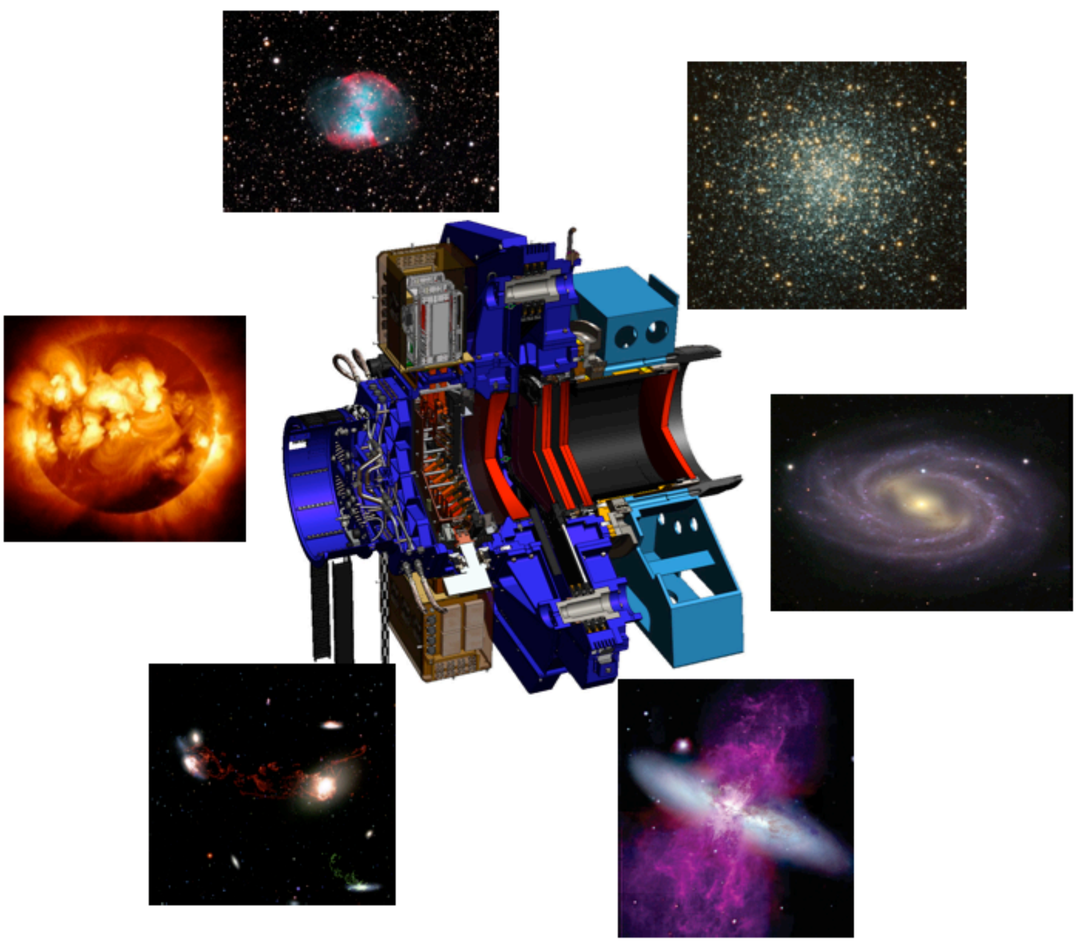
One Degree Imager Pipeline Software and Archive Science Requirements Document

Version 1.5: 7/23/09

Ian Dell'Antonio, Daniel Durand, Daniel Harbeck, Knut Olsen, John Salzer

Final Draft Edition: Pierre Martin



1. Introduction

This document defines the requirements and technical specifications for the WIYN One Degree Imager (ODI) data pipeline and the requirements for a data archive for the ODI data (the ODI-PASRD). The pipeline is intended for use both by the WIY partners and by the general community through the NOAO access to the data. This document is mainly the result of a meeting held in Tucson on June 15-16, 2009.

In the document, we have attempted to separate out the requirements for the pipeline and archive from goals for the pipeline and archive. Roughly speaking, for the ODI pipeline, **requirements** are processes that will be required by all or nearly all of ODI users, as well as processes that are essential for the successful completion of the science drivers detailed in the ODI Science Requirements Document (ODISRD). **Goals** for the pipeline are processes and methodologies that would increase the scientific usefulness of the pipeline. Within the goals, we have made some attempt to prioritize the importance to ODI. For the archive, requirements are considered to be functionalities required for PI science with ODI and functionalities that are necessary for mandated access by the community to the data. Goals are functionalities that will increase the science output and usability of the ODI data by both the University partners and the community.

We note here in the introduction that achievement of the specifications described in the ODISRD requires precisely the type of pipeline procedures that make for a usable archive. The pipeline will produce uniformly calibrated dataset; the archive will need to provide data back to the pipeline as calibrations evolve. We therefore envision a system in which the pipeline and archive are heavily intertwined.

We have not specified the exact algorithms to be used in a pipeline in most cases, but it should be understood that there is a significant consensus in the astronomical community as to the “best practice” algorithms for most if not all of these steps, and that these should be adopted for the pipeline. For some novel modes and specific reduction tasks, however, such best practices will have to be established during ODI commissioning. We therefore envision that the pipeline must be able to evolve during the progressive steps of commissioning described in the ODISRD. We also point out that there are many cases were issues of WIYN operations cannot be separated from the pipeline. In particular, achieving the science requirements of ODI will necessitate a standardized program of calibration files. The pipeline requirements can detail how these calibrations are to be used, but it is a matter for WIYN operations to determine how to acquire these necessary calibrations.

The document is organized as follows. We begin with a summary of the charge to the committee, then general remarks about the design of the pipeline and archive, followed by a discussion of the requirements for the four areas of emphasis identified (removal of instrumental signatures, advanced processing, time domain requirements, archiving and data mining), and finally a response to the specific questions listed in the charge. Appendix 1 includes the definitions of the different reduction steps as envisioned for the ODI pipeline, and appendix 2 presents sme guidelines on the priorities for the implementation of the pipeline and archiving.

2. The ODI-PASRD charge

2.1. Scope

ODI will be used for a large number of diverse scientific programs. It is probably unrealistic to attempt to cover all of the possible scientific applications in developing this PASRD. The PASRD should be concise (< 20 pages) yet as complete as possible. It is, however, important to emphasize again that the goal is not to design the pipeline in this exercise, but rather to offer a set of scientific requirements to be used later on as the pipeline project moves forward.

We request the following approach:

1. Identify and define the most common scientific requirements for ODI regarding the data products, such as removing the main instrumental signatures from the raw data. Despite ODI having a non-conventional focal plane, these analysis steps are not expected to differ significantly from normal reduction recipes. If relevant, we strongly encourage the Committee to prioritize these steps.

2. Identify and define more elaborate scientific requirements, which might lead to the development of more sophisticated analysis steps for the pipeline, that might be requested by a large number of users, and prioritize them. As examples, the requirements for data stacking, astrometry accuracy, and photometry should be described.

3. Identify any specific scientific requirements for time domain programs, in particular related to the time frame in accessing detrended or fully reduced data, if relevant.

4. Evaluate the scientific importance of archiving and data mining of ODI data products for the success of the existing ODISRD, and if relevant, define the time scales, eventual needs for data re-processing, data product contents, and metadata needed to meet these scientific requirements.

Important Remark: The Committee should conduct the development of the PASRD using the ODISRD document provided. The ODISRD offers a good overview of scientific applications planned for ODI, which in return provide a strong guidance in developing the scientific requirements for the pipeline and archiving.

3. Committee Charge

3.1 General Charge

We request the Committee to draft the science requirements, illuminated with selective user-case examples, for an ODI science pipeline and archiving system that is necessary for the WIYN Observatory to see the successful realization of the ODI SRD. We further request that the

Committee prioritizes, when appropriate, these requirements. The identification of goals for the pipeline and archiving system is encouraged.

3.2 Specific Questions

In addition to the general charge above, we request the Committee to specifically respond to the following questions:

1. What are the scientific requirements regarding the removal of instrumental signatures of ODI that would be the most commonly required by the users in order to achieve their scientific goals? What are their priorities?

2. What are the scientific requirements related to more advanced data reduction and analysis steps (e.g., stacking) that are likely to be required by ODI users in order to achieve their scientific goals? What are their priorities?

3. What are the scientific requirements of time domain programs (e.g., dataflow time scale vs. quality of data reduction) that should addressed by the reduction pipeline?

4. What are the scientific needs for archiving the ODI data? If archiving is needed: What data products should be archived, for how long, and what sophistication might be needed for data mining?

5. Are there any other scientific requirements to ensure success of the data pipeline? For instance, how important is it for users to have some control over the reduction process? How important is it for users to be able to provide their own reduction routines? How important is the ability to merge external data products into the archiving system (e.g., for VO applications?)

Since this list of questions is non-exhaustive, the Committee is invited to add any relevant requirements or information that they judge would be useful later on in defining a data analysis pipeline and archiving system for ODI.

3. General Requirements for the Pipeline and Archive

The central conclusion of the committee, as elaborated below, is:

A data reduction pipeline is essential for the successful deployment of ODI—even if the camera works flawlessly, if the data it produces cannot be adequately calibrated and analyzed, the entire project will be a failure.

Because many of the science goals (astrometry, weak gravitational lensing, wide-field or deep imaging surveys, monitoring and variability studies) require the combination of data taken over the interval of years, **it is also essential for a long-term archive of ODI data to exist**. Not archiving the data for periods of >5 years would result in a failure to achieve the science goals detailed in the ODISRD. Because the uniformly analyzed data required for the science goals is potentially immensely valuable, we consider it essential that ODI store at least the raw images for reprocessing by the pipeline for the lifetime of the instrument, particularly if the pipeline has primary responsibility for the calibration so that uniformly processed datasets can exist. In fact, it will almost certainly be necessary to reprocess large amounts of data, as calibration files and practices are likely to be under development for the first few years of ODI operations.

We recognize that although the pipeline and archive are technically separate entities, they must be tightly coupled to be effective. For example, the pipeline must be able to (re)-process raw data from the archive, especially as improvements in calibration files become available. This will be very difficult if the bandwidth between the pipeline site and the archive site is insufficient. Likewise, the archive's usefulness depends on the existence of a central pipeline that allows uniform reduction of data.

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.

Evidence for this requirement is amply supplied by a range of scientifically successful modern instruments used for wide field instruments, including SDSS, CFHT cameras, and NEWFIRM.

Any data reduction pipeline should be considered as a long-term operational endeavor. The complexity of ODI data will certainly lead to several iterations on processing, calibrations and algorithm fine-tuning. To ensure the success of the pipeline, it is essential that user support be provided by WIYN on a long-term basis, not only during the development phases of the pipeline.

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

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.

The calibration requirements described in the ODISRD have implications both for the pipeline design and the data collection. In order to achieve the photometric, sky flatness and astrometric precision required by the science goals, the ODI data pipeline must have access to a uniform calibration dataset. In particular, a library of calibration files that is kept up-to-date should be available to the pipeline. (The calibration files will be listed in section 4). Whether the calibration files are requested of PI observers, generated by the ODI project, or part of a queue program is an operational question, but a centralized pipeline (and incidentally a usable archive) requires a uniform calibration dataset. A further benefit to the centralized pipeline is that calibration images that need to be generated by using scientific data (most notably night sky flats and accurate fringe images) can be available to the pipeline without compromising data rights of individual PIs and surveys. It is an operational question whether the ODI pipeline calibration data are available this way or through periodic ODI campaigns to generate the required data (keeping in mind that night sky flats, ghost pupil images and fringe images have to be calculated from images taken without OT guiding).

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.

Production of the calibration images should be as standard as possible—and if reduced data are stored in an archive this standardization will be essential. The existence of a WIYN ODI Queue for observing would greatly facilitate the acquisition of the necessary images. If not, WIYN operations will need to adjust to make sure the image libraries are up to date.

ODI data can be acquired in a variety of different imaging modes, as detailed in the ODISRD. We anticipate that the majority of observations will be taken in three modes: static imaging, coherent guiding, and local guiding. The pipeline must support all three. A fourth mode, non-sidereal tracking, is not expected to see much use, but processing of data taken in this mode should also be supported because for all the processing steps up to image combining it will behave like a static imaging or a coherently guided one, depending on the rate of tracking. One other mode, shutterless photometry, is supported by the camera but produces very different data, and it is not anticipated that the data reduction pipeline described here will be able to handle data from that mode, particularly if only the photometry stream is output. The archive must, however, be able to ingest and serve to PIs the video mode output of the stars (which is not the default mode but is allowed), so that PIs may process the video streams, as they desire.

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.

For many applications, scientific measurements will be made using existing software (for cataloging, shape measurement, etc.) on the fully detrended scientific exposures. As a result, it is critical that the output of these exposures be in a format that is compatible with existing and impending tools.

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.

Although it is anticipated that most of the users of the pipeline will want only the fully detrended images (or even the stacks), we anticipate that some users might want specialized reduction steps. It is beyond the scope of the pipeline to expand to accommodate all user interactions. It should instead be regarded as a goal for the pipeline to be modular, so that users could request the output of each task. This would relieve the pressure on the pipeline organization to over-customize what should be a very straightforward detrending pipeline.

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).

In any case, the documentation describing the individual steps of the ODI pipeline must be public so that astronomers know what analysis is done to the ODI data.

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.

Similarly, although the existence of a central pipeline is a requirement for success of ODI as a facility instrument, there will be users and institutions that have resources to run the pipeline on data locally. The existence of a central pipeline facility should not preclude running the pipeline or its individual functional modules on other machines. Although we do not wish to place any strict requirement for open-sourcing the pipeline, the software should be available.

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.

This has implications about access to the calibration and metadata files. This is a question of WIYN policy, but it is our strong recommendation that the calibration library images be publicly available immediately when they are made, even the calibration files constructed from PI science observations that are not public.

Requirement 3.9. Master Calibration files shall be public.

To ensure that the precision required for the science cases envisioned in the ODISRD can be achieved, and to enable proper image stacking, it is important that errors be tracked and propagated through the pipeline, as much as possible. This means that variance and possibly data quality images must be associated with science images and be considered part of the science output of the detrending and stacking part of the pipeline.

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.

Although ODI will be used for some large surveys, it is not a unitary survey. Thus, image release to the public from an archive should be continuous, in accordance with WIYN and TSIP release policies, rather than in seasonally timed data releases as in the SDSS.

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.

Access to the data from the pipeline and archive is envisioned to be primarily electronic.

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

We recognize that this is a potential roadblock to access to the pipeline data for users from institutions without adequate access to the Internet. However, there are very significant management costs to maintaining a data distribution system. If there are sufficient resources, a goal of the pipeline and archive systems should be to maintain a data distribution system to provide access to data that cannot be transferred over a slow Internet connection.

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.

4. Pipeline processing of Images—Detrending of Instrumental Signatures (Tier 1):

Identify and define the most common scientific requirements for ODI regarding the data products, such as removing the main instrumental signatures from the raw data. Despite ODI having a non-conventional focal plane, these analysis steps are not expected to differ significantly from normal reduction recipes. If relevant, we strongly encourage the Committee to prioritize these steps

The processing of ODI data, particularly data taken in the OT modes, is sufficiently complex and computationally expensive that a standardized reduction scheme is strongly recommended even for individual PIs. Furthermore, ODI archive data have the chance of becoming useful for users other than the original PI only if data are calibrated consistently and with a dependable quality control mechanism. Science data and calibration have to be taken in a well-defined way, even if WIYN/ODI should be operated in visitor mode. It is therefore recommended that:

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.

A further benefit of a pipeline processing is the potential of multi-year consistency of processing. This is essential for long-term projects (monitoring, surveys, proper motions), but also allows the archive collection to be generally useful.

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.

Because the ODI camera represents a completely new implementation of CCD imaging, it is expected that changes in the processing procedures and especially in the calibration data required will crop up during the first year(s) of operation. As a result, we expect that there will be a need for re-reduction of data archived, especially during the initial deployment of ODI.

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.

It is anticipated that ODI will have the option to acquire data either in unbinned or in binned mode. The pipeline must be able to handle binned data (at least with 2x2 binning). ODI will also potentially have a “slow” readout mode optimized for narrow-band imaging.

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.

Of course, not all permutations will be allowed or will make sense—in general, because OT guiding occurs at the native pixel scale, if the data are binned at readout it may be difficult to reconstruct the OT-shifted calibration files required. WIYN should create policies on which permutations of readout modes should be supported.

Because this places a significant calibration burden on observations, it is recommended that the Tier 1 pipeline provide an operator interface that guides through the generation and selection of the best end-of-run master calibration products.

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.

The pipeline must also be robust to the absence of data from particular cells or even detectors. At the cell level this is a natural consequence of the OT guiding.

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.

We envision the pipeline to encompass the following standard data processing steps, and that all data taken with ODI in the supported modes should be corrected for these steps.

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.

The level of the crosstalk subtraction required is not known, and likely will not be known until the populated focal plane array is delivered. However, the intention of this requirement is that the software not be the limiting factor in cross-talk correction. If the cross-talk coefficients can be accurately determined, they should be subtracted as completely as possible. This is potentially a computationally expensive problem, because it means that the pipeline must be capable of handling cross-talk between all rows of cells—i.e. potentially up to 512 separate cross-talk terms. Coefficients used in the cross-talk removal should be logged for each image. In addition, the pipeline should provide tools for measuring cross-talk coefficients from specially selected observations (e.g. observations of star clusters).

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.

Note that it is quite possible that there will be different-sized overscan regions based on the readout mode chosen (i.e. "slow" readout mode allowing a larger overscan region to further depress the read-noise contribution. The pipeline must therefore be able to work with at least two separate overscan configurations, ideally with a completely general algorithm. With ODI's synchronization scheme, there is the potential for the overscan to vary line-by-line, so the pipeline should be capable of either line-by-line overscan subtraction or lower-order fitting over multiple lines as required. The overscan region must be trimmed off the images, and the overscan correction that is applied should be logged for each image. The pipeline shall recognize and flag as potentially bad situations where the overscan of an individual cell jumps or shows significant deviation from "normal" values.

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.

.

Although dark current itself is expected to be small, there will likely be other effects (e.g. amplifier glow) that will need calibration exposures of varying time taken with the shutter closed.

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.

Observers will be able to take exposures with a very wide range of exposure times, but the calibration set will be necessarily limited to a set of standardized times. The pipeline therefore has to be able to interpolate dark current calibration frames to match the actual dark time (not necessarily the same as exposure time) of an exposure. The pipeline must handle enough calibration frames to ensure that shot noise will not significantly increase readout noise.
Dark current calibration frames potentially have to be convolved with the guide history of OTA corrections. Therefore, the pipeline must be able to handle both unconvolved (for static imaging) and convolved dark frames and potentially produce an individual dark frame for each exposure.

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 $\ll$1% pixel-to-pixel noise.

The pipeline must be able to generate and keep updated a library of master dome flat fields and master night-sky flat fields for each filter and readout mode. The pipeline also must provide mechanisms to remove pupil ghosts and fringes from flat fields. Note that this will almost certainly involve an iterative process involving both dome flats and sky flats. The details must be established during ODI commissioning, as they depend on the relative intensities of fringe and ghost pupil signatures relative to the sky background in each filter, neither of which is known precisely at this point. The requirement on the pixel-to-pixel variation ensures that application of the flat field does not contribute significantly to the statistical photometric errors. Flat field combination should include outlier rejection (e.g., removal of cosmic rays) and cleanup for bad columns/pixels. The flat field step should also provide gain normalization across the entire image (producing a flat image modulo real sky gradients). Flat field images may have to be convolved using the OTA guide history before they are applied. For sky flats, this may mean that a set of images taken in static imaging modes for each filter may need to be acquired. Given the potential for scattered light sensitivity of ODI, specific locations in the sky may need to be identified for sky flat construction. The procedure for acquiring these calibrations is a matter for WIYN operations.

Flat fielding (or the illumination correction described in requirement 4.14) may also have to correct for illumination patterns that depend on the orientation of the ADC prisms, and thus might vary from exposure to exposure in a predictable way. The pipeline should be prepared to correct for such second order flat field effects.

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.

Fringe maps will have to be rescaled before being applied to individual images and the pipeline must have a mechanism for determining this scaling. Fringe maps are expected to be stable long-term at least during photometric conditions (at least this has been the experience with Mosaic, but ODI operations should periodically (on the timescale of ~1 year) monitor the fringe images.

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.

The pupil ghost is to be distinguished by more complicated forms of scattered light (i.e. off axis light scattering of elements in the telescope or dome and entering the detector). Unlike the pupil ghost, these forms of scattered light will be very difficult to remove via an automatic pipeline. Projects that require the most exquisite background subtraction (studies of intracluster light, very low surface brightness galaxies or unresolved outer halos of nearby galaxies) may need to be designed so that scattered light signals can be modeled out separately. In any case, we believe that correcting for these features lies outside the requirements of the Tier 1 and Tier 2 software.

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.

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.

Flat fielding corrects for a non-uniform pixel-scale and creates a flat background, but it results in a photometric zero-point gradient for point sources. The pipeline should generate an illumination correction based on the WCS and any known vignetting. The choice of applying the illumination correction to all ODI images as a default simplifies the pipeline operation, and allows science output from individual, unstacked images.

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.

Cosmic rays are to be identified with single-image algorithms such as LA-cosmic, which should be easier to implement given the small pixel scale of ODI. Cosmic ray flagging is particularly essential for long, binned-mode narrow band imaging exposures, for which a non-negligible fraction of the pixels may be affected. Rather than modifying the pixel data, it is preferable that the potential cosmic ray locations be flagged in the data quality image.

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.

The data quality mask should flag pixels as bad or questionable for a variety of reasons, each with its own (binary) code. Examples of flagging include:
* pixels compromised because of saturation
* pixels compromised because of cosmic ray
* pixels compromised for other reasons (bad pixel, bad columns)
* all pixels lost at the boundaries of cells because of the OT shifts as well as the OT guiding cells themselves.

The data quality mask is essential for the analysis of single images, and also will enable higher quality data stacks to be generated. The variance map is required for error propagation—keeping track of the uncertainties introduced at each calibration step will be necessary to ensure that the photometric precision goals stated in the ODISRD are met.

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.

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.

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.).

Even though there is a requirement for the data calibration products to be public, it would be even more valuable if they were available both to the on-the-mountain validation reductions and to projects (such as fast transient detection) that will necessarily not be well served by the tier 1 pipeline as envisioned in this document.

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.

At a minimum, it is expected that these catalogs contain the following entries:
X&Y pixel coordinates, as well as WCS-corrected RA and DEC; a global and an aperture magnitude and flux measurement (e.g. the SExtractor MAG_AUTO and MAG_APER and associated fluxes); a bounding box for each detected object (X_MIN, X_MAX, Y_MIN, Y_MAX); shape measurements for the objects (major and minor axis sizes, orientation angle, elongation and ellipticity); the sky level and the FWHM for the object; data quality flags and possibly some star-galaxy separation index. It is understood that where appropriate errors in these quantities should also be cataloged.

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.

Once the final throughput of the telescope in the various filters and observing modes has been calibrated, it is expected that rough photometric offsets should be associated to each exposure at this step. This is especially important for the archive, as this associates an estimate of the depth of the exposure to the metadata allowing for efficient searches. Of course, there will be PI science programs for which this level of accuracy will be sufficient, independent of better photometric estimates envisioned for the tier 2 software.

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.

We gave consideration to the idea of long-term storage of only the raw data, with the archive, with reprocessing of data from the archive by the pipeline to a staging area. However, such a system has complexity and management costs that likely exceed the storage cost of archiving the detrended data. In particular, we note that the only archive that has successfully attempted the "process on demand" model is the HST archive. Most of the access to ODI data by both PIs and community users is expected to be through the detrended images. It is likely, however, that resampled data for stacking should not be stored, as the resampling is likely to alter the noise properties and would have to be different for different stacking prescriptions.

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.

This "requirement" is different from the other requirements in this document. It may in fact not be achievable, but the committee felt that it was important to emphasize how pressing the time constraints are and to focus the attention of WIYN on the urgency to produce working software. In order for ODI commissioning to transition smoothly into science operations, users must feel confidence that their data will be calibrated and that they will be able to do science with ODI. If there is no prospect for a pipeline by then it could prove disastrous.

5. Advanced Pipeline features (Tier 2 software)—Photometry, Astrometry, Image Stacking and Image Differencing

Identify and define more elaborate scientific requirements, which might lead to the development of more sophisticated analysis steps for the pipeline, that might be requested by a large number of users, and prioritize them. As examples, the requirements for data stacking, astrometry accuracy, and photometry should be described.

Beyond the basic pipeline reduction steps for de-trending the images, the panel attempted to identify the most fundamental analysis tools that would be essential to a wide range of science programs and astronomers. These tasks are considered to be so fundamental that they must exist, at some level, in order for ODI to be considered a successful instrument.

The basic tools we see as most important are:

A) Automated Object Identification and Classification
B) Photometry
C) Astrometry
D) Image Stacking
E) Imaging Differencing (e.g., ON minus OFF in narrow-band imaging, identification of time variable sources to the Poisson limit)
F) Image filtering tasks (Gaussian convolution, median filtering, PSF normalization)
G) Stack slicing—dividing the stacked images into manageable pieces.

In what follows, we present our thoughts on each of these tools. In some cases, the discussion involves requirements on observing operations, since the proper handling of some of the tasks requires a combination of software and a specific implementation plan in order to achieve the requirements detailed in the ODISRD. In addition, it is worth pointing out that the details for how to carry out some of these tasks are naturally less detailed than the specifications given for the Tier 1 processing, because there is less standardization in the astronomical community as to how these should proceed. Despite this, we hope that the requirements laid out here will still lead to a well-defined path forward for an ODI pipeline. We discuss each task in turn. Note that the ordering is not meant to be a priority ranking for the various functions.

Automated Object Identification and Classification

The committee realized that it was important to draw the distinction between source catalogs (that are useful as input to calibration tasks and for rough quality checks) and science catalogs (that contain the level of precision and completeness for scientific analyses). Although science catalogs will definitely need to be produced for scientific output, the committee felt that there was too much variety needed in the catalog construction for science catalogs to be part of the pipeline. In this sense, because ODI is not a purely survey instrument, there is no uniform catalog that will serve all the ODI PI users, let alone archival users.

We note, however, that source catalogs must still be produced by the pipeline to successfully perform other calibration steps, as described in and after requirement 4.18, and we require that these catalogs be available to the PIs and be linked to the images in the archive.

Photometry

The panel envisioned this as having two fundamental parts: calibration, which involves primarily operational issues, and all-frame photometry, which is more completely a software problem. Since the two cannot be totally separated, we include a discussion of both.

If ODI is going to be able to reach its goal of providing 1% photometry (or better), non-standard calibration methods must be developed. That is, a typical user will not be able to come to the telescope and fully calibrate their data in the traditional way. Rather, the WIYN staff will need to carry out a program of photometric calibrations to properly specify the photometry obtained with ODI.

A key task that clearly needs to be done by the staff is the measurement of the magnitudes of standard stars with a range of colors on all of the individual detectors (cells?) via a step-and-stare grid observation pattern in which the individual stars are observed on all the detectors. The purpose of these observations would be to calibrate the relative zero-points for each detector, and to determine the color terms for each filter as a function of location on the array. This should be done periodically, to check for long term variations in the calibration values. Doing this at a minimum of once per year is essential, and perhaps more frequently during the early stages of ODI's use.

Measuring nightly zero-point constants would best be achieved using a standard set of fields and exposure times. The WIYN staff is strongly urged to develop a complete set of fields with cached coordinates and exposure times for all broad-band filters. A similar set of spectrophotometric standard fields should be defined for narrow-band calibrations

The panel recognizes the broad range of techniques used for actually carrying out the process of flux measurement on a CCD image, and does not think it is appropriate to elaborate specifications at this time. The software must be capable of delivering on the 1% photometric accuracy figure. Whether this can be achieved by the basic photometry code associated with the cataloging software (section A), or a more sophisticated approach is required, needs to be evaluated.

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.

Astrometry

Again, this task is seen by the committee as requiring a combination of operational and software components.

We expect that the raw data will enter the pipeline with a coarse WCS already in place, with a field center specified to significantly better than 1 arcsec. Given this, a key next step is to apply an accurate geometric correction based on detailed observations of a standard astrometric field (e.g., NGC 188). Generating this correction map is a task that should be carried out by WIYN staff during commissioning, and repeated on a regular basis to check for stability. We assign a goal of 0.2 arcsec precision on the absolute coordinates of any object on the camera after the application of this correction. It is reasonable to expect that this correction could be applied during the Tier 1 pipeline processing, rather than in Tier 2.

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).

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.

Determining independent cell-by-cell astrometric solutions will not be possible in general due to the density of astrometric standards in the general field. For many (most?) applications, the 0.2 arcsec global astrometric precision specified above will be adequate. Developing software for calculating a general astrometric solution that results in substantially higher precision (e.g., a full solution over say a 4 x 4 grid of cells) should be considered, but only if there is a clear science case for it. The impact of the 0.2 arcsec global astrometric precision limit also needs to be evaluated in the context of the imaging stacking software. In any case, for the stacking, it is expected that the relative astrometric precision will be much greater, because many more objects will be available in most ODI observations that can be used for relative astrometric determinations.

Image Stacking

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).

This is considered a fundamental task for many programs and essential to the success of ODI and is therefore a requirement of the pipeline. The applications of image stacking are quite broad, ranging from combining a set of images taken with a basic dither pattern (a fundamental task that must be available at first light) to combining survey images taken over a range of months or years for selected deep survey fields. There should not be an a-priori limit on the number of images to stack, but it is possible to envision a limit on the area of sky covered by the stack (i.e., to establish a contrast between "stacking" and "mosaicing").

The fundamental parts of the stacking process, and their attendant software requirements, are:

i) Determination of the relative astrometric transformations between the images. For most applications, this will be accomplished by matching coordinates of objects detected in the standard catalogs of each detector cell, and should allow the (non-linear) transformation to be computed at the detector level with <<1 pixel accuracy. However, there are applications (short exposures, very blue or narrow band filters) for which the source density is insufficient to allow offsets to be calculated the detector level. In this case, the pipeline must be able to allow either integer pixel shifts or shifts to be determined in larger-than-detector cell blocks.

ii) Flux scaling between the images. This needs to be carried out with an accuracy of better than 1%, in order to avoid degrading the native photometric precision. It is expected that, provided that the photometric corrections in the tier 1 pipeline are correctly applied, there will be sufficient objects in all but the shortest narrow-band exposures for this level of precision to be achieved.

iii) Weighting of individual images. Specific weighting schemes would need to be user specified, and could be as simple as uniform weighting, or as complicated as taking into account the relative scaling and PSF values of each image and the variance maps associated with the image by the Tier 1 pipeline. Given the need for some user input in the stacking, it is anticipated that the stacking of anything other than pre-designated dither sequences should be carried out upon request, either at a central facility (via some sort of resource queue) or via the availability of the stacking code to individuals and institutions with significant computing resources.

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.

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.

Image Differencing

This process also has a wide range of applications, but is crucial for most narrow-band imaging programs (e.g., creating ON - OFF images) and for variable source observations, and therefore it is a requirement that the Tier 2 ODI software allow for image differences to be produced.

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.

In principle, image differencing is a slight variation on the image stacking task, with the key difference being that the images being processed are not added, but subtracted from each other. The process needs to include the following functions:

i) Image alignment. This is similar to the stacking process, and has the same requirements as for the image stacking.
ii) PSF matching (e.g., convolving with a Gaussian kernel to yield the same PSF in all images). Requirements: all PSFs have FWHMs matched within 10% both image-to-image and across a given image, except where the density of stellar objects does not allow accurate determination of the PSFs on small scales.
iii) Flux scaling of the individual images. The requirements for this step are the same as for image stacking - better than 1% accuracy where possible.
iv) Subtraction of the images.

This process can be applied to pairs of individual images, to two stacks of images, or to a stack of images relative to a single image. We note that this process could potentially be used for the detection of transients (i.e. supernovae). However, we do not consider it a requirement for the pipeline to produce image differences on a timescale sufficient for detection of fast transients. Nevertheless, we encourage the sharing of codes developed here with groups interested in setting up independent fast transient identification pipelines.

Image Filtering

Another Tier 2 functionality that will be requested by many if not all science programs is software to perform simple filtering operations on the images. Examples of such operations are Gaussian convolution, median filtering, and PSF regularization (convolution with a finite user-defined function). Again, the expectation is that this software need not be part of the processing that leads to archived data, but the ability to perform these operations on both individual detrended images and stacks is a requirement for the success of multiple science cases. As with the case for image stacks and differences, if this software runs at a central processing facility, the output needs to be available in a staging area for PIs to access/download.

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.

Image Slicing

The output of even a single stack will be an image roughly 4Gb in size. This will overburden many current analysis and data visualization programs. Especially as we do not envision a science cataloging tool to be part of the Tier 2 software, we recognize the necessity of providing a tool to divide stacked (and optionally unstacked, but resampled) images into sub-images that can be processed and displayed with existing software.
One possibility is that the image slicing software be identical to the image cutout service envisioned for the archive (see section 7).

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.

6. Requirements for the pipeline for time-domain Science

Identify any specific scientific requirements for time domain programs, in particular related to the time frame in accessing detrended or fully reduced data, if relevant.

ODI's large field of view and good image quality will make it an excellent tool for the discovery and observation of time variable objects. In particular, the ODI Science Requirements document features several science cases that require time series data, including halo RR Lyrae, stellar variables in clusters, transiting planets, extragalactic supernovae. As a precursor to LSST, ODI will also almost certainly be used to refine LSST observing strategies and to test LSST analysis tools, particularly those designed to characterize and classify the plethora of objects that will be found in the time domain. The ODI pipeline and archive should thus be responsive to the needs of time domain science to the extent possible. For the evaluation of the scientific requirements of time domain programs, we need to consider two different types of observations and objects:

A. Observations that are entirely supported by ODI imaging, such as monitoring observations, or the discovery and derivation of light curves for variable targets in a survey of a particular region of sky. These observations do not require prompt, triggered followup on ODI or any other facility. These could include:

1. Periodically variable objects.
2. Transients that appear and disappear on timescales of days or longer, e.g. supernovae and microlensing events for which prompt spectroscopy is unnecessary.

3. Transients that appear and disappear on timescales of hours or minutes (for applications that do not require followup observations of these transients).
4. Slowly moving objects, such as KBOs.

For these observations, we assume that fast knowledge of the variability of the objects is not required, only that the pipeline make images with instrument signatures removed available as soon as they are produced. We thus expect that these types of time domain observations will be well served by the Tier 1 data products.

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.

A trickier case is:

B. Observations that use ODI imaging for discovery and other instruments for follow-up, or that require rapid follow-up with ODI in different parts of the sky, including:
1. Identification of periodically variable objects, e.g. eclipsing binaries or transiting planets, for which radial velocity measurements are desired from spectroscopy.
2. Initial discovery of transients that appear and disappear on timescales of days or longer, e.g. extragalactic supernovae to be used in cosmological studies, for which spectroscopic classification is needed within days of maximum light.
3. Initial discovery of transients that appear and disappear on timescales of hours or minutes, e.g. gamma ray bursts for which a redshift is required from spectroscopy.
4. Observations of fast-moving NEOs.

For these types of observations, if the Tier 1 or more advanced data products are needed on timescales faster than they can nominally be expected to be provided by the ODI pipeline, then the investigators will be provided with the raw data and calibrations for their observations, but must themselves take responsibility for the reductions. In particular, observations of fast transients (categories B3 and B4 above) will need to be reduced by the investigators, not the pipeline. The reasons for this recommendation are that the needs of the science is likely to be very specific to individual projects, and thus not well served by a general pipeline, and that large groups have historically formed to solve the problems of the specific projects, and thus will have greater expertise than the pipeline group. However, it is important to note that if people must bring their own equipment to WIYN, it is likely that this will require additional resources from WIYN, including possibly more cooling, bandwidth and access for computers in the WIYN computer room.

For less rapid transients (category B2), the investigators may also need to take responsibility for their own reductions, depending on their needs. We do recommend, however, that if it enables the science to be done through the pipeline, that such programs be considered for priority pipeline processing. For the regularly variable objects (category B1), we expect that the nominal timescales for Tier 1 and more advanced products will serve the community's needs.

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.

7. Requirements for the archive

Evaluate the scientific importance of archiving and data mining of ODI data products for the success of the existing ODISRD, and if relevant, define the time scales, eventual needs for data re-processing, data product contents, and metadata needed to meet these scientific requirements.

Requirement 7.1. A basic archive must be in place for ODI commissioning. The archive must be capable of ingesting data from the pipeline, serving data back out to the pipeline, and serving raw or processed data to PIs (and eventually to community users). The capabilities to ingest data from the raw data stream and to serve it to the pipeline must be available from the beginning of commissioning. Other functionalities can be developed on a longer timescale.

A basic working archive must be in place right away, and should be used in conjunction with the pipeline during science commissioning. If well conceived, the archive will be a central point of contact where data and metadata are stored and fetched from by all users, be they PIs or archival users. A Science Archive so constructed will be able to maximize the full scientific potential of the ODI.

For this purpose it is essential that: The Science Archive must contain and serve all the pipeline processed data and metadata. This includes the raw and the detrended pixels, and all the housekeeping data which is essential in assessing the science quality of a given exposure as well as any housekeeping data essential to obtain the full provenance of a given products. For this purpose it is essential that all execution of important pipeline modules are traceable for a given science product. This will insure repeatability.

Requirement 7.2. The science archive should serve all Tier 1 pipeline processed data and metadata, as well as the raw data and master calibration files. Ideally, the archive should also contain the Tier 1 data although this has to be financially viable in the global context of the pipeline project.

It is clear the ODI pipeline will evolve with time because of bug fixes, added features, and the possible introduction of more challenging science calibrations. It is then essential that the Science Archive is designed to be able to cope with these changes with time.

Requirement 7.3. The science archive should be able to accommodate reprocessing, and, for at least an initial period, a replacement of calibration techniques and metadata stored in response to advances in methodology for handling ODI data.

Once established, the Science Archive should allow execution of simple and complex queries using a well-designed user interface. The basic search should be able to cope with simple space, time, and energy (wavelength) based queries, as well as queries by PI or program number. This will also greatly facilitate the usage and the inclusion of the ODI archive within the Virtual Observatory (VO). Complex queries should be able to handle exact footprint based queries. The spatially based queries should accept input co-ordinate or an object name.

Requirement 7.4. The science archive should allow execution of simple (and eventually complex) queries using a well-designed user interface.

The Science Archive should provide authenticated access to PI and survey team, and open access to the public data. The authenticated access should not disturb the regular ingestion of new data within the Science Archive. It is also important that the Science Archive must allow access to survey/PI data before all observations are complete for a given program.
In addition to direct data access, the Science Archive must allow requests for arithmetic operations and options for a few basic operations on pixel data. The basic operations, which have to be put in place soon after first light, are image stacking and differences with the possibility of defining the input images (as described in Section 5).

Requirement 7.5. The science archive must allow authenticated access to PIs (and survey teams) during the proprietary period for the data, and open access to the public data.

The Science Archive must also provide access to an ODI cutout service. Users will be able to request any portions of a given image using a central coordinate or a list of coordinates. The size of the cutout box could also be specified (within practical limits). This is essential because the full-stacked images are too large for standard visualization tools. Finally, for expert users, it must be possible to get images, or portion of images using a very simple web based programmatic interface (using tools like wget, curl, etc.). This is a fundamental service allowing direct access to the entire collection from a user's software located anywhere on the Internet.

Requirement 7.6. The science archive must provide access to a cutout service for regions of images to be accessible. A downsampled image (e.g. binned 8 x 8 .jpeg) should be available for quick recognition of individual, smaller fields extracted from an entire image.

Requirement 7.7. The science archive must allow programmatic access to the archive via a web-based programmatic interface (with some appropriate authentication); as a goal this functionality should allow the use of images or image sections in the user's software where authenticated properly.

Requirement 7.8. Archiving models proposed to meet the above requirements should contain options on costing (e.g. "on-the-fly" Tier 1 vs. Tier 1 storage).

8. Response to the specific questions detailed in the charge

1. *What are the scientific requirements regarding the removal of instrumental signatures of ODI that would be the most commonly required by the users in order to achieve their scientific goals? What are their priorities?*

A Tier 1 pipeline must exist, and it must remove the instrumental signatures as described in section 4 of this document. All of the steps envisioned in this section are of high priority for success of the scientific goals of ODI as detailed in the ODISRD.

2. *What are the scientific requirements related to more advanced data reduction and analysis steps (e.g., stacking) that are likely to be required by ODI users in order to achieve their scientific goals? What are their priorities?*

Tier 2 software as described in section 5 of this document represent the advanced reduction tasks that an ODI pipeline center must support. The highest priority among those should go to stacking, differencing, and photometric and astrometric calibration. It is anticipated that source rather than science catalogs will be produced by the pipeline, leaving scientific cataloging to a user-specified Tier 3 of processing.

3. *What are the scientific requirements of time domain programs (e.g., dataflow time scale vs. quality of data reduction) that should addressed by the reduction pipeline?*

Programs that require photometry and/or image differencing on timescales shorter than days are not necessarily served by the envisaged pipeline. These programs must however be allowed access to the raw data stream and metadata available from the mountain as foreseen in the ODISRD. As a goal, we suggest the implementation of priority pipeline processing employing library calibration images. We also suggest that validated calibration data should be available for inclusion in private fast transient pipelines.

4. *What are the scientific needs for archiving the ODI data? If archiving is needed: What data products should be archived, for how long, and what sophistication might be needed for data mining?*

The scientific requirements of projects described in the ODISRD necessitate multi-year (and hence long-term) archiving of ODI data, metadata and calibration files. Raw data and mountain-generated metadata must be archived. Tier 1 detrended and generated auxiliary data (e.g. data quality images) should also be archived, as it will not in general be cost-effective to "generate them on the fly". We envision the archiving requirements to be long-term, approximately the 10-year lifetime of the instrument. Simple and eventually more complex queries (positional, wavelength, time queries) should be supported, as well as programmatic access to the data. An image cutout server should be deployed to maximize the usefulness of the stored data to PIs and to the community.

5. *Are there any other scientific requirements to ensure success of the data pipeline? For instance, how important is it for users to have some control over the reduction process? How important is it for users to be able to provide their own reduction routines? How important is the ability to merge external data products into the archiving system (e.g., for VO applications?)*

It is very important (a requirement) to provide for some user control of the Tier 2 software, and it should be a goal for some user control of Tier 1 where it is scientifically motivated, with the understanding that any eventual user control of Tier 1 calibrations should not feed or replace data in the science archive. We envision most user reduction routines to operate on the output data products of the pipeline rather than being incorporated in the pipeline. However, a mechanism for community to propose input of advanced software tools should be available. Merging of external data into the archive seems outside the scope of the project, best handled by a VO interface. To ensure that this is possible, archive data products must be VO enabling.

Appendix 1
Data Reduction Tier Definitions

The overall science pipeline should be divided into several layers, which are referred as Tiers. Each of the tiers provides specific functionality and the divisions of tiers are implemented along natural boundaries in the scientific workflow.

- **Tier 0:** Real-time (quick look) analysis for observers and (possibly) time-domain programs: basic analysis done on site on local machines. [Included within the ODI project].

Remark: The development and operation of the first basic reduction step are carried out at WIYN. Tier 0 is part of the One Degree Imager (ODI) project and entirely developed by the WIYN ODI project team. Tier 0 was the initial limited science pipeline done with pre-defined calibration products. Reduction is done locally, at the WIYN summit facility. Tier 0 produces individual images with basic flat-fielding and correction for the over-scan so that the observer can assess the quality of the data. Tier 0 data products could also be required for time domain programs. No re-processing is feasible. User support ends with the run.

- **Tier 1:** End-of-run, removal of instrument signatures (from master calibrations) and more advanced spectrometric and photometric calibrations on individual images (1% flatness).

- **Tier 2:** Production of optimum science products. For example: Image stacking, high accuracy astrometric and photometric solutions, PSF re-sampling, cosmic ray removal, fringing correction. Fine-tuning on stacking (e.g. specific selections of images) and production of catalogs (1% photometry accuracy).

- **Tier 3:** Image manipulation (for example, image filtering, image arithmetic, marking and examining sources, artificial star tests), display, photometry, etc.

Appendix 2
Implementation Priorities

Although the PASRD Committee explicitly recommends that **all** of the requirements of the pipeline and archiving system for ODI be eventually met, it is probably not realistic to envision that everything can be made available all at once. More likely a significant fraction of the pipeline and archiving developments, in particular those related to the more advanced observing modes of ODI, will proceed with the *incrementa*l commissioning steps planned for the instrument. The figure below suggests some guidelines for prioritizing *the implementation* of the ODI pipeline and archiving. It is likely that several components will be developed in parallel (for instance, basic stacking is likely to be needed during commissioning so its development could proceed at the same time as some Tier 1 components) but the general idea in proceeding with an incremental implementation is illustrated in this figure. Note that the figure is non-exhaustive; some requirements (e.g. documentation) are not illustrated.