AUSTRALIAN NATIONAL UNIVERSITY

Created: 20 April 2005

Last modified: 26 July 2005

SKYMAPPER CRITICAL DESIGN REVIEW

SKYMAPPER SCIENCE DATA PIPELINE AND SCHEDULER SYSTEM

Stefan Keller, Annino Vaccarella, Brian Schmidt, Tony Martin-Jones and Darius Powell

Research School of Astronomy and Astrophysics

Institute of Advanced Studies

Australian National University

Revision History

Revision No.	Author & Date	Approval & Date	Description
1	Stefan Keller, Annino Vaccarella 27 April 2005	Stefan Keller 20 April 2005	Initial release for review by Peter Young and Mark Jarnyk
2	Stefan Keller, Annino Vaccarella 27 April 2005	Stefan Keller 27 April 2005	Post review
3	Stefan Keller, Darius Powell 3 May 2005	Stefan Keller 3 May 2005	Finalised DFDs
4	Stefan Keller 10 May 2005	Stefan Keller 10 May 2005	ICD incorporated
5	Stefan Keller 17 May 2005	Stefan Keller 17 May 2005	ANUSF specs
6	Stefan Keller 20 May 2005	Stefan Keller 20 May 2005	Release for second review by Peter Young and Mark Jarnyk
7	Stefan Keller 10 June 2005	Stefan Keller 10 June 2005	Release for internal review
8	Stefan Keller 26 July 2005	Stefan Keller 26 July 2005	Release for external review

SkyMapper Science System

Table of Contents

1 SkyMapper Science System

In this document we present the SkyMapper Science System. The Science System is required to optimise the acquisition of data (via the Scheduler System), conduct the subsequent data reduction and make the data products available via the Internet (via the Science Data Pipeline System). We present our requirements in detail in Section 3 and respond to this with our design detailed in Sections 4-9. In Section 10 we detail our costing and in Section 11 our risk management strategy.

2 Applicable Documents

Document ID	Source	Title
SDN08.01	SkyMapper	SkyMapper Science Requirements
SDN05.01	SkyMapper	SkyMapper Software Control System
DC-TAROS-03-CDR001	Computer Section, RSAA	Telescope Automation and Remote Observing System (TAROS) Critical Design Review Document
SDN08.03	SkyMapper	SkyMapper Scheduler to TAROS ICD
SDN09.01	SkyMapper	SkyMapper Project Plan

3 Requirements

3.1 SkyMapper Scheduler System

The SkyMapper Scheduler System must maximise our scientific output by providing automated, optimised field selection and data quality control during the nightly operation of the telescope. The Scheduler must:

- Select from the predetermined SkyMapper field centres the most opportune 5-Sec and Main Survey field in real time at the telescope. This selection should be as rapid as possible, no more than 1 second to avoid latency.
- Complete fields in all filters under photometric conditions for the 5-Sec Survey.
- Implement a dither pattern for consecutive images of a field to eliminate chip gaps and defects.
- Have the capability to respond to high wind gust conditions as reported by the weather station. The specification of maximum operationally tolerable wind gust speed awaits EOST specification and on-site validation.
- Reduce a focussing sequence image to determine the optimal focus position.
- Schedule twilight flatfields.
- The scheduler must complete fields in all filters three times for the Main

Survey first data release and six times for second data release. In addition, we require the following time intervals between observations: 4 hours, 1 day, 1 week, 1 month, 1 year. Some flexibility in these intervals is allowed: 1 hour, 2 days, 3 days, 2 weeks, 2 months either side respectively.

- The scheduler must respond to photometric/seeing conditions during a night. If the sky becomes non-photometric it must switch from 5-Sec to the Main Survey. If the seeing is greater than allowable limit for Main Survey, but sky is photometric, switch to the 5-Sec Survey.
- It will take into account the various factors that contribute to the ranking of the importance of each field.
    - In the Main Survey these are: ability to obtain a time interval, airmass, sky brightness due to the Moon, current telescope pointing (to reduce slew time latency), ability to complete set of 3 or 6 images for data release and the available time left to observe the field.
    - In the 5-Sec Survey these are: airmass, sky brightness due to the Moon, current telescope pointing, ability to complete set of 6 images for data release and the available time left to observe the field.
- In the case of the 5-Sec Survey, the scheduler must schedule standard star observations at appropriate intervals (hourly).
- The scheduler must interface with TAROS to obtain the observations.
- The system must examine the acquired data for quality assurance. Science images are to be checked for appropriate count levels, photometric zeropoint, sky background and seeing. Calibration data is checked for appropriate count and noise levels. Data which fails quality tests must then be flagged for reobservation.
- The system must monitor photometric conditions during the night and adjust the survey type accordingly.

3.1.1 SkyMapper Dither Pattern and Effective Field of View

We have explored the optimal dither pattern for consecutive images of the same field. Our aim was to optimise the number of observations per pixel over all pixels in the observed field. The dither pattern that resulted from this examination is as follows:

$$x_c = x + 0.5(i-1)G + \varepsilon \qquad y_c = y + 0.5(i-1)G + \varepsilon$$

Where *G* is the large gap between the rows of CCDs with amplifiers at their edges and ε is a small shift to account for possible pointing errors between subsequent images (7 arc seconds).

We use the outermost pixels which are completely unobscured to define the field of view of the SkyMapper detector. This results in a field of view of 2.344x2.2916 degrees (8440x8250 arc seconds). With this dither configuration:

100% of pixels have 3 observations,

99.6% of pixels have 4 observations

99.0% of pixels have 5 observations and,

90.5% of pixels have 6 observations

3.1.2 SkyMapper Field Centres

The SkyMapper 5-Sec and Main Surveys will utilise a set of field centres. Fields are arranged in bands of fixed declination such that overlap in right ascension is always equal to or greater than 2 minutes of arc. The fields in each declination band are equally spaced. Because of variations in the gaps between fields in adjacent declination bands, the declination bands are not of equal spacing but rather draw slightly closer together as declination increases to the pole. Each field is centrally aligned along a meridian of 0hrs right ascension for simplicity. Figure 1 shows the resulting tessellation of the southern sky. There are 4066 fields.

3.2 SkyMapper Science Data Pipeline

The Science Data Pipeline incorporates those processes required to produce science data products that meet our science requirements from the available image data.

The Science Data Pipeline must:

- Be invoked and shut down without operator intervention
- Be able to be distributed over multiple processors (but not necessarily multi-threaded)
- Produce and apply calibration data in order to transform the real detector system to a linearised detector system. This includes:
    - Trimming to remove the overscan region of each chip
    - Masking out bad pixels (those that do not have consistent linear response to incident light).
    - Bias subtraction. Subtract a median bias frame from data. With e2v CCDs and SDSU-III controllers we do not expect to have structure in bias frames. However, we have included this requirement for the system in case this should not be the case.
    - Linearity correction. CCDs are typically very linear devices but still suffer non-linear response at the 1-2% level at high charge levels. Linearity corrections can be determined during detector testing using a LED and are chip specific. We will apply linearity correction after bias subtraction.
    - Flatfielding, scattered light and fringe removal. Flatfielding is the process of dividing through by a response image. The difficulty is arriving at an accurate response image in the presence of fringing and scattered light.

      Our prescription is as follows:

        - Create flatfields from the median of a stack of twilight flatfields. Divide through by this flat field.

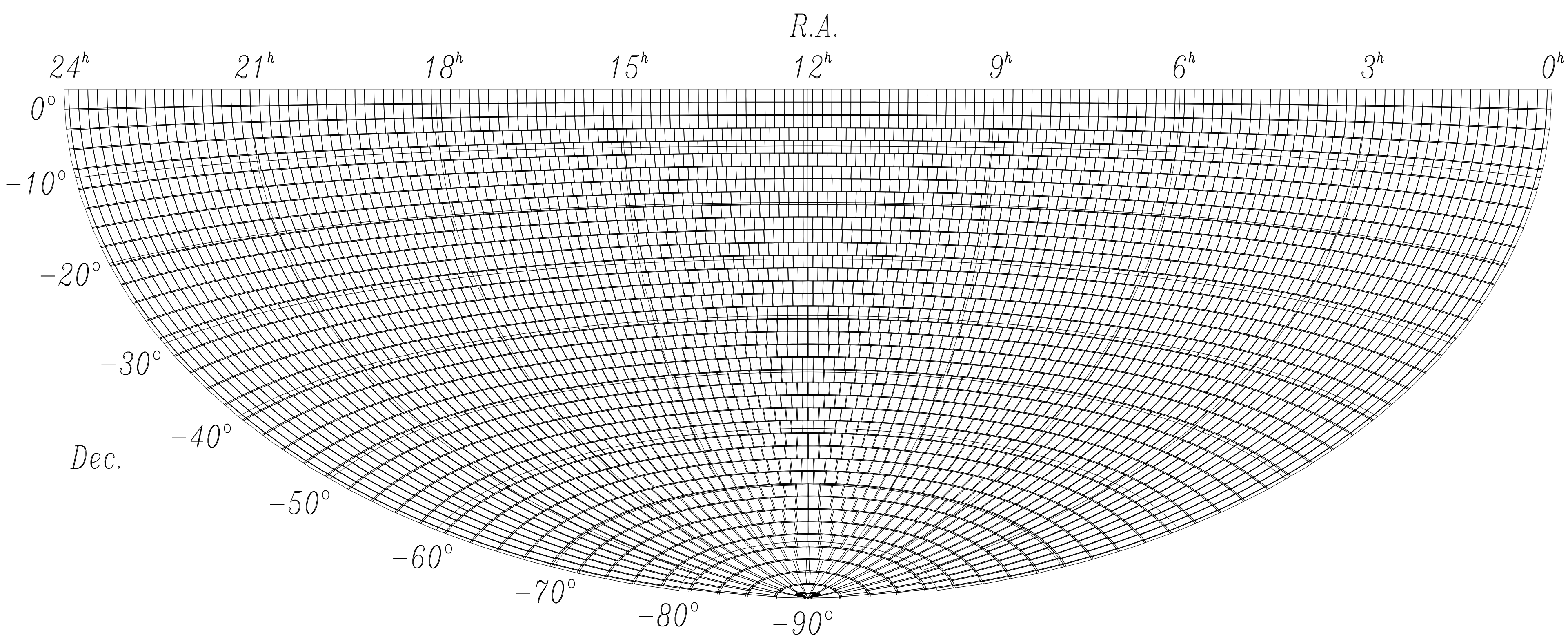


Figure 1 The SkyMapper Survey fields

    - Create fringe frames for *i* and *z* filters by comparing twilight flatfields (which are dominated by the solar continuum spectrum and hence do not exhibit fringing) and superflats (created from the median of pixels containing only the night sky, the majority of which is due to near-monochromatic sky lines that interfere within the device at red wavelengths). By subtracting the twilight flat from the superflat (after scaling to a common value), a map of the fringing pattern is produced. Scaled and subtract the fringe frame from observations.
    - Create an illumination correction to remove the effects of scattered light. Scattered light typically is the most important limiting factor to the accuracy of photometry obtained from wide-field photometry. We expect some broad scattered light from light reflecting off the CCDs and filters onto the telescope optics, as well as light reflected within the telescope structure and enclosure. We will assume that the scattered light will be different for each filter, but will not depend (at least in a complicated way) on rotator angle, and is relatively stable in time. These latter points will need to be checked on the telescope.

      The most straightforward method for obtaining a true flatfield is to create it from observations of stars. In this process, the standard flatfield derived from the superflat/twilight flat process is applied to a series of observations where the same field is observed on each CCD chip. The relative magnitude of each star in each chip is compared to the average of the star observed over all chips, and the smoothed map of average deviation of stellar photometry made for each CCD. This correction to the nominal flat creates a true flat, which can be applied to all images. These images can then be stacked and the average residual in the sky background measured. This residual image then forms the scattered light image, which can be scaled and removed from each image in the flatfielding process.

- Monitor the optical system for changes in the instrument that affect the science product. Specifically, has the flat field changed due to dust deposition and/or is there a change in bias structure due to hardware changes.
- In the case of the 5-Sec Survey:
    - determine the photometric status of each night and the nightly photometric transformation coefficients. Reject data as required
    - perform aperture photometry of each point source in the image with a 2*fwhm aperture. We choose this largish aperture to limit any problems caused by PSF-shape drift caused by scintillation and instrumental effects photometry. Apply the aperture correction required to take magnitude to infinite aperture radius. Write results to a database.
- In the case of the Main Survey:

- place each image on a common standard instrumental system utilising the overlap with the 5-Sec Survey
    - Combine the 3/6 images per filter of each field centre
    - perform photometry on science images with attention paid to galaxy photometry
    - detect transient objects,

    and write results to a database.
- Make scientific results available to the community via an image server and a data server accessible via the Internet. It is our intent to provide a web service (i.e. essentially a web API) utilising SOAP (http://www.w3.org/TR/SOAP/), XML Remote Procedure Call (http://www.xmlrpc.com). This will ensure maximum integration into the Virtual Observatory effort.
    - **Image Server:** The interface will allow RA, Dec, filter, field size searches. Images returned with Virtual Observatory Metadata. Compressed preview images and FITS images served. Scope will depend on availability of additional Virtual Observatory funds.
    - **Data Server:** The database server will support RA, Dec, field size, filter, survey queries. It will not enable fully rational database queries as we do not have the required computing power (i.e. Will not allow "find all objects i-z>2" type queries). Scope may be broadened on availability of additional Virtual Observatory funds.

4 Design

We now present our design for a system to produce science data products meeting the above requirements. Figure 2 shows the system context diagram for the Science System. The Science System consists of two components: the Scheduler System residing at SSO, and the Science Data Pipeline residing at ANUSF.

The Scheduler System interfaces with TAROS to provide telescope configurations to TAROS during nightly data taking and performs quality assurance on the observed fields returned by TAROS. The Science Data Pipeline manages the subsequent reduction of the acquired data (with operator input where required) and provides external user access to the data through a Web interface. The procedures of both systems are managed by a pipeline framework called OPUS described in the following section.

4.1 The OPUS Processing Environment

The individual tasks of the Science System are controlled within the OPUS processing environment, a third party supplied system from STSci (http://ess.stsci.edu/products/opus/faq/opusfaq_intro.html). OPUS provides the capability to construct an automated pipeline which can be distributed over multiple processors.

The SkyMapper Science System will operate OPUS in an external polling mode. In this mode, each process in the system will update a status flag in a file maintained by OPUS. OPUS will poll the status file to determine the status of each process and upon completion, initiate the next dependent process. Benchmark tests indicate that OPUS operated in external polling mode introduces a ~16% latency.

The OPUS package is controlled from the command line via a series of scripts (of our own construction). These scripts initiate the pipeline, monitor its status and shut it down.

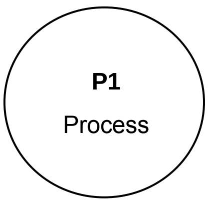
4.2 Design Methodology

In the discussion of our system design numerous requirement/design methodology systems are applicable. To avoid confusion we now describe our chosen methodology. The motivation is to capture as much information regarding the temporal and procedural operation of the system.

To capture temporal information we use where necessary process flow diagrams. These show the temporal flow of the pipeline as well as a description of the main actions of each process. These are not to be confused with state transition diagrams or state models (the processes detailed here are too simple to warrant such description).

The process modelling of the system has been represented by data flow diagrams. Data flow diagrams provide a graphic representation of the units of processing within a system and the intercommunication between them. We have chosen to adopt the methodology of Shlaer/Mellor.

The following symbols are used in the subsequent data flow diagrams:

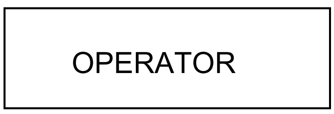


Each process is represented by a bubble. Each process has a process ID and is named by a short action clause that describes the function of the process.

Data flow is represented by a solid arrow with a description of the data. Data passed can be considered transient.

.

Data stores indicate persistent data.



Terminators represent external entities to the system.

Control flow is indicated by a dashed arrow.

System Context

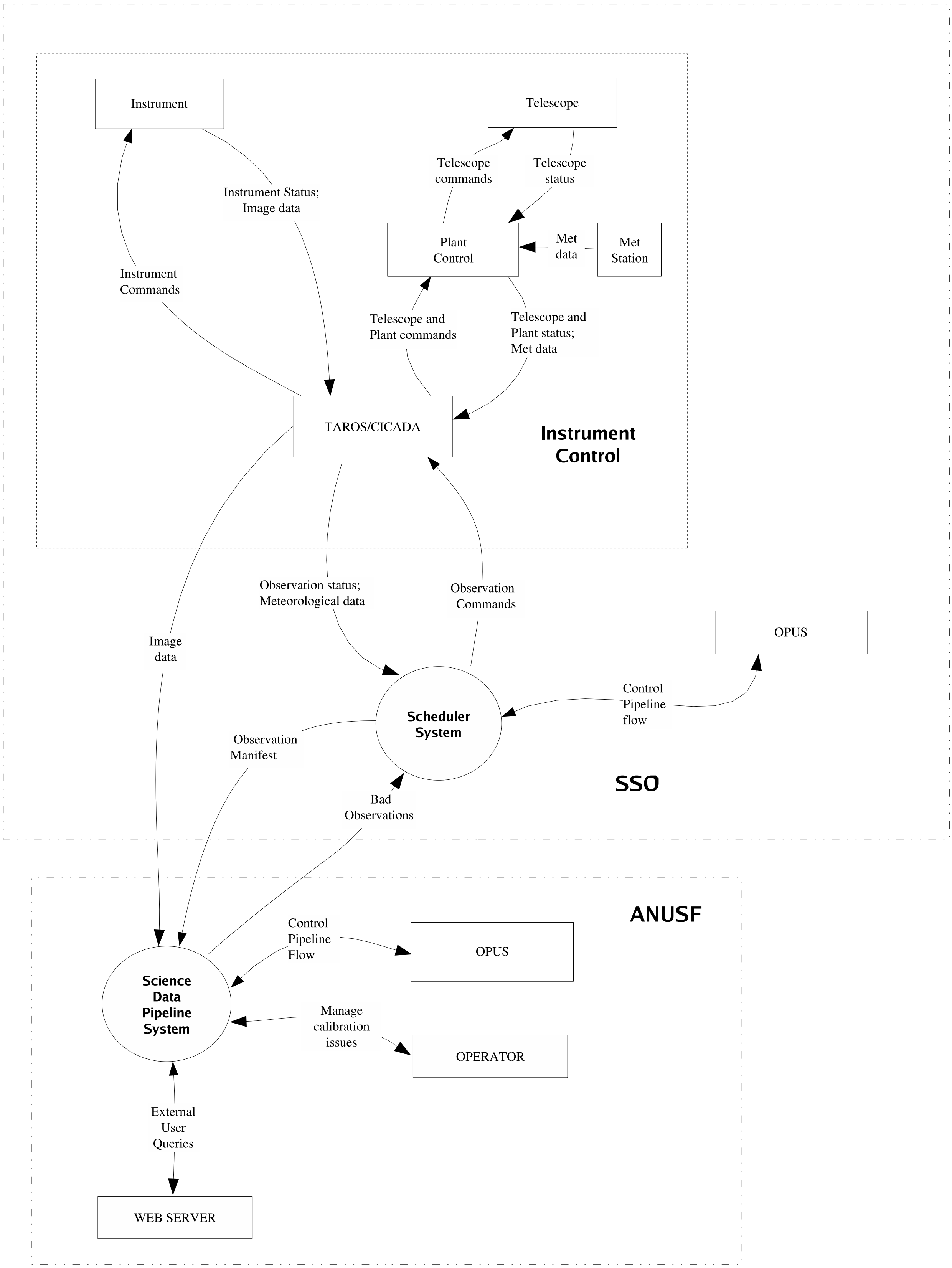


Figure 2 System context diagram for the SkyMapper Scheduler and Science Data Pipeline Systems

5 Overview of SkyMapper Science System Processes

The following is a walk through the operation of the SkyMapper Science System in order to capture the temporal flow of the various procedures. In Section 6 we describe the functional details of each procedure.

5.1 Scheduler System Operation

The Scheduler System is located at SSO. The process flow of the Scheduler System is described below. The corresponding data flow diagrams for the system are found in Figure 3and Figure 4.

1. Outside peak load times, sufficiently in advance of the night's observing, the P1.1 and P1.2 Preselectors are initiated (cron job) to generate two field shortlists, one for each survey mode. This preselection minimises compute time during the night by maintaining shortlists of high priority fields visible during the night and their pertinent information. Two shortlists are maintained such that the observing mode can be changed during the night based on prevailing conditions.
2. Ten minutes after sunset the Scheduler will request a set of Bias frames and then a set of dome flatfields for each filter.
3. At (civil twilight-10 mins) the Scheduler will request twilight flatfields be taken
4. At (astronomical twilight – 10 mins) the Scheduler will request a focussing sequence be obtained. This will be reduced locally and the focus modified.
5. At the start of astronomical twilight the Scheduler commences science image acquisition.
6. During observing the P1.3 examines the status of the telescope/instrument and meteorological conditions (in particular the wind speed, see Section 3.1) as provided by TAROS. When the instrument is in readout mode then P1.3 initiates the Obsselector (P1.4).
7. Obsselector (P1.4) selects either the 5-Sec or Main Survey on the basis of the photometric status supplied by P2. P1.4 selects from the prioritised list for the current survey mode produced by P1.1 or P1.2. On the basis on the current wind status certain azimuthal angles may be excluded. The instrument configuration for the highest ranked field at the time of observation is returned to P1.3. P1.3 then passes this to TAROS.
8. 32 ccds are read out to a raw data repository by TAROS. TAROS is responsible for conveying the raw data to ANUSF Raw Data (D7).
9. P2 polls for new data and if found performs some basic quality assurance checks. It also examines the available photometric zeropoints to test the photometric status of the night. This photometric status is conveyed to the Scheduler to switch from 5-Sec (only under photometric conditions) to Main Survey mode (non-photometric conditions tolerated) and visa-versa. P2 also establishes the seeing of the image and this is supplied to the Scheduler to switch from Main Survey to 5-Sec if seeing limits are exceeded. Upon completion of P2 and receipt from TAROS that the image has been transferred to the ANUSF data repository the image is deleted from the SSO Raw Data repository.

10. P3 polls for the presence of a Bad Observations file containing images that have failed processing through the Science Data Pipeline. Such images have their status updated in SSO Image DB accordingly.
11. At the start of morning astronomical twilight science image taking stops and the Scheduler obtains twilight flatfields for each filter (until 5 per filter per night are obtained)
12. At civil twilight + 10 mins a set of bias frames are requesed by the Scheduler.
13. Once the biases are obtained, P1.3 generates a nightly manifest of observations which is conveyed to ANUSF via TAROS.

5.2 Science Data Pipeline Operation

The remainder of processing takes place at ANUSF data repository. The following describes the process flow. The corresponding data flow diagrams can be found in Figure 5-Figure 14

1. P4.1 (see Figure 6) polls for the presence of a nightly manifest. If found P4.1 examines if all the images contained in the manifest are present in the ANUSF Raw Data repository. If they are not it waits, exits, retries. If the images have not appeared after two days the operator is flagged. If all images are present the new image entries are inserted into ANUSF Image DB together with the quality assurance information.
2. P4.2.1 (Figure 8) is triggered by P4.1. It performs a nightly sanity check on the calibration data obtained that night. It initiates P4.2.1.1-4.2.1.4 to use raw bias and flatfield data as described in ANUSF Image DB to generate processed calibration data (PCD) using this night's calibration data and then compares the results against existing PCD as described in Cal DB.

   If sanity check indicates good cal data then the time period range for the current proc cal data is increased in Cal DB (i.e. It is authorised PCD – applies for the current night – the system has not changed). An appropriate period (~1 month for flatfields for instance) will be defined as maximum length of time for a current PCD file. After this time the current proc cal data can not be reauthorised. The raw calibration data then becomes part of the raw calibration data buffer which will be used in next update calibration.

   If sanity check fails then the operator is flagged to resolve the problem. The operator can indicate that some or all of the raw calibration data is not valid and should be not used as calibration data (flagged in Cal DB). This loops, sanity check is redone.

   The operator can also indicate that the instrumental system has changed and the current PCD is no longer valid. The current PCD is marked in Cal DB as expired and new current proc cal data must be generated. In particular, if the flatfield indicates system change or flatfield has expired then the fringe frame and illumination correction will also be regenerated as the three are related.
3. P4.2.2 Generate Calibration (Figure 9) is triggered after the completion of P4.2.1. This process updates current PCD if there does not exist authorised PCD. P4.2.2.1 looks for non-existent authorised PCD and initiates P4.2.2.2-4.2.2.5 (as required) to attempt to generate new current PCD using available raw calibration data buffer in the raw calibration data buffer. Newly created current PCD are returned to the operator for inspection if requested. The new

current PCD is then authorised.

4. Upon completion of P4.2, P4.3.1 is initiated (Figure 10). P4.3.1 complies a list of science images with the requisite authorised PCD. P4.3.2 applies this calibration data to the science images. If insufficient raw calibration data files have accumulated to enable the creation of authorised PCD required then the processing of science data that require these calibration files will stall here until new current PCD is formed.
5. Exit from P4.3.2 triggers P4.3.3 which applies a WCS to each amp image.
6. P4.4 polls for images which have been reduced and deletes them from ANUSF Raw Data.

The science data is now ready to have photometry conducted upon it. Dependent on the survey type of the science data the pipeline now forks into either the 5-Sec Survey Photometry pipeline or the Main Survey Photometry pipeline.

5-Sec Survey Photometry pipeline:

1. The initiation of the 5-Sec Survey Photometry pipeline is managed by P5.1.1 (Figure 12), which polls for a night that does not yet have a confirmed photometric status. If such a night exists then the quality assurance zeropoints are reviewed and outliers rejected as non-photometric or if the entire night is of excessive dispersion then all data is rejected as non-photometric.
2. Exit from P5.1.1 triggers P5.1.2 which performs photometry on the available photometric science data. The raw photometry is ingested into the 5-Sec DB.
3. Derivation of the nightly photometric transformation coefficients is managed by P5.2.1 (Figure 13), which looks for a night of data with verified photometric conditions which has not yet been standardised in the 5-Sec Survey. If found, P5.2.1 derives the transformation coefficients for the night and stores them in Nightly Photometric Coefficient DB (D11).

   This requires operator interaction at the time of regression to remove obvious outliers. If the operator decides a good fit can not be made to the data for the night then the data is rejected from the 5-sec survey (flag data in ANUSF Image DB). If good fit is obtained the data is certified photometric and will be part of the 5-Sec Survey.
4. P5.2.2 polls for an image that has authorised photometric status and has not had standardised magnitudes created. It uses the transformation defined in Nightly Photometric Coefficient DB and transforms raw magnitudes to standard instrumental magnitudes. We note that the above procedures facilitate regeneration of pipeline photometry as the user can externally make mods to the coefficients in Nightly Photometric Coefficient DB and photometry status in ANUSF Image DB in order to rerun/modify.

Main Survey Photometry pipeline:

1. The Main Survey Photometry pipeline is initiated by P6.1 (Figure 14), which examines the ANUSF Image DB. For first data release P6.1 examines if there is a field in the Main Survey that has 3 images in a filter and has not yet been incorporated into the Main Survey DB. 5-Sec Survey data must also exist for the field. For second data release P6.1 will look for a set of six images.
2. P6.2 is triggered by P6.1. P6.2 then places the instrumental photometry on a

standard instrumental system by utilising those objects in common with the 5-Sec Survey.

3. P6.3 is triggered by the exit of P6.2. The three (or six) images are combined. The combined image is then placed in our ALTAS combined image repository.
4. P6.4 is triggered by the exit of P6.3. P6.4 generates object positions from the combined template image. This will reach to 5σ in the combined template.
5. Exit from P6.4 triggers P6.5. P6.5 performs photometry on the images and updates the Main Survey DB with calibrated photometry.
6. Exit from P6.5 triggers P6.6. P6.6 examines the individual images for transient objects and outputs those found to a Transient Object DB.

P7 (Figure 6) monitors the ANUSF Image DB for images that have failed in the Science Data Pipeline. This information is conveyed to the Scheduler System as required.

External Access:

Finally, external access will be provided for through P8 the Survey Web Service (Figure 6) accessible over the internet. This will offer external users access to the photometric databases and the combined survey images.

6 Textural Description of Data Flow Diagrams

The following is a description of each process in the Science System. The state machine of each process is sufficiently trivial that implementation will follow from the process description without the need to descend to a lower level design specification.

6.1 SkyMapper Scheduler System

6.1.1 P1 Scheduler

6.1.1.1 P1.1 Main Survey Preselector

1. The date of observations is obtained – user specified or the current date – and the corresponding start and end Modified Julian Dates (MJDs) of astronomical twilight are retrieved from the Ephemeris DB.
2. A list of uncompleted fields (fields not yet observed and those whose observations are not completed) which are observable during the night is sought. Fields in this list must be above an airmass of 2 during the night if at the equator. The maximum allowable airmass grows for southerly declinations to include the south celestial pole.
3. A score is calculated for each field. The first score modifier is on the basis of whether it is possible to obtain a 4hr interval on a certain field during the night. This requires that the field be above a critical airmass for more than four hours.

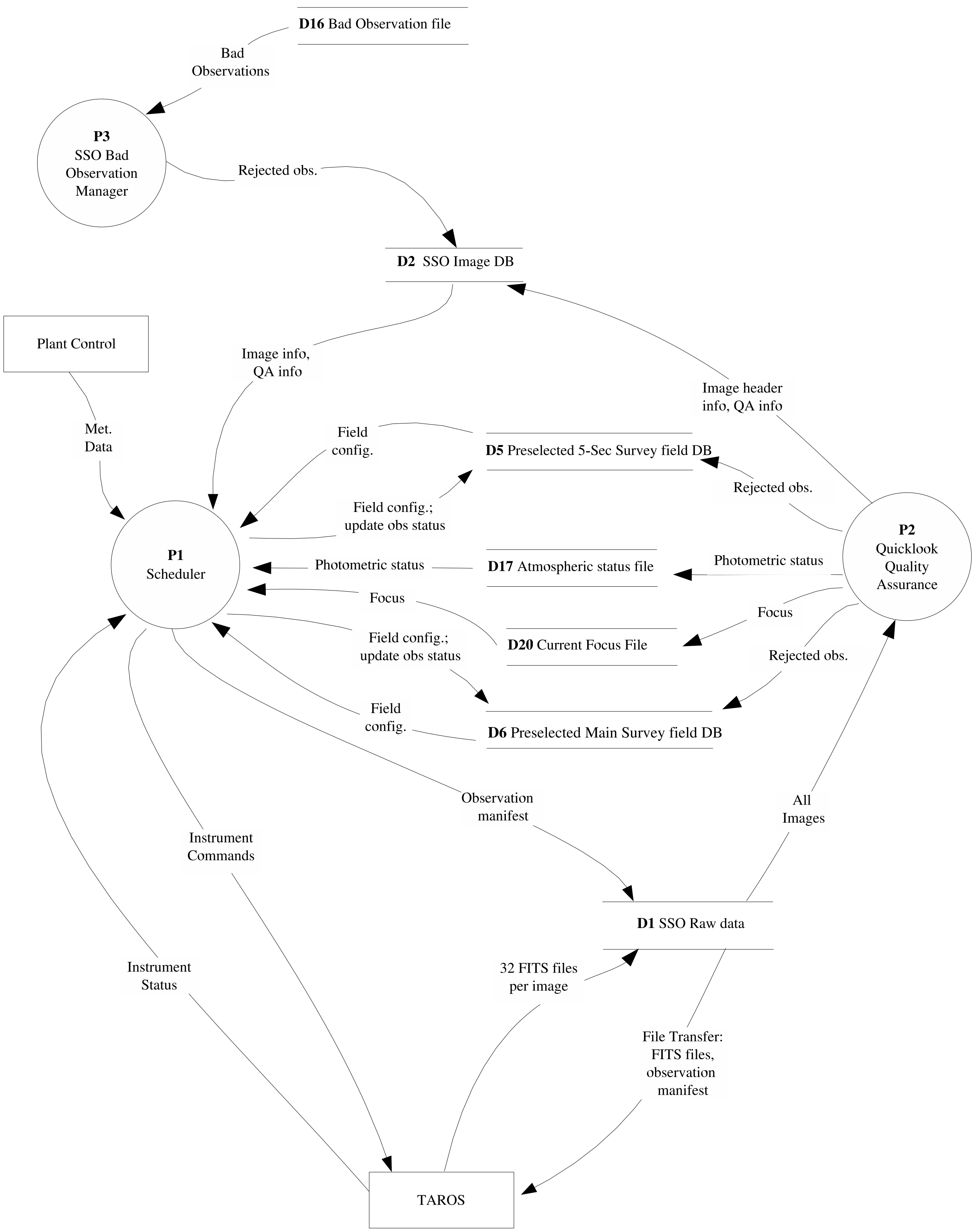


Figure 3 System level data flow diagram for the Scheduler System

4. The score is modified based on the time left in the survey when the field is observable. We loop from the current date to the ending date of the survey and calculate the number of hours observable and express this as a fraction of the total number of hours left in the survey. Expected lost time due to bad weather will be factored in.
5. The score is further modified on the basis of the position of the moon on the night. The effect of moonlight on the sky background at the location of the field is calculated; this provides an appropriate penalty for observing fields close to the moon.
6. The score is modified on the basis of whether it is possible to complete another time interval (1day, 3days, 1week, etc). Essentially the score modifier here will be an appropriate function that strongly weights fields around those time intervals. The weight function will allow a degree of flexibility to the intervals as specified in the science requirement document.
7. Finally the score is modified on the basis of ability to complete a set of 3/6 images in a filter for data release.

The fields are ranked on the basis of their score and the top 500 are placed in a Preselected Main Survey field DB.

6.1.1.2 P1.2 5-Sec Survey Preselector

P1.2 operates in a very similar manner to P1.1:

1. The date is obtained and the star and end of astronomical twilight are retrieved from the Ephemeris DB.
2. A list of uncompleted fields that are observable during the night is created.
3. The score is modified based on the time left in the survey when the field is observable. We use the values determined in P1.1.
4. The score is modified on the basis of the position of the moon on the night. We use the values determined in P1.1.
5. The score is modified on the basis of whether it is possible to complete all six filters for a given field for data release.

The fields are ranked on the basis of their score and the top 500 are placed in a Preselected 5-Sec Survey field DB.

6.1.1.3 P1.3 Observation Manager

This process is responsible for establishing the status of the instrument and supplying the best-ranked field via TAROS. This interface is detailed in Section 6.4.2.

P1.3 schedules five different types of observations – the selection is time dependent:

Start time	*End Time*	*Data requested*
Sunset + 10 mins	Civil twilight - 10 mins	Biases then dome flatfields
Civil twilight	Astronomical twilight – 10 mins	Twilight Flatfields
Astronomical twilight – 10 mins	Astronomical twilight	Focus sequence
Astronomical twilight	Astronomical twilight	Science images as per P1.4
Civil twilight - 30 mins	Civil twilight + 10 mins	Twilight Flatfields then biases

In addition, it is envisaged that focus sequences (using Shack Hartman or a modify focus, expose, shift and repeat sequence) will be performed during the night. In this way a sequence of required calibrations in defined time windows is established. This will be established in a nightly calibration database.

An example of the operation of P1.3 is illustrative; in a night 3 focus checks are required this establishes the need for the following sequence of calibration data: biases1, dome flatfields, twilight flatfields1, focus1, focus2, focus3, twilight flatfields2, and biases2. At sunset +10 minutes P1.3 is initiated for the night. It initialises the nightly calibration database. The time is in the range for obtaining biases and dome flatfields. The nightly calibration database indicates no biases have been obtained so P1.3 triggers P1.5 to supply a set of biases. The system remains idle until P2 reports the status of the biases. If all pass quality control P1.3 looks at the time and the nightly calibration database and determines the next calibration set to obtain are the dome flatfields. If some biases fail QA a successive call of P1.5 retries. All dome flats are obtained successfully but because the time is less than civil twilight P1.3 does not request twilight flatfields and the system waits. If the dome flats were not obtained before civil-10mins then P1.3 would have skipped reattempts. And so forth.

The remainder of the time P1.3 will trigger P1.4 for science image acquisition.

P1.3 also obtains:

1. meteorological data from TAROS. This information is supplied to P1.4 when TAROS requires a new field.
2. the current optimal focus. This is read from Current Focus file D20. If the value in the file has changed from the last call to P1.4/P1.5 then the focus is adjusted accordingly.

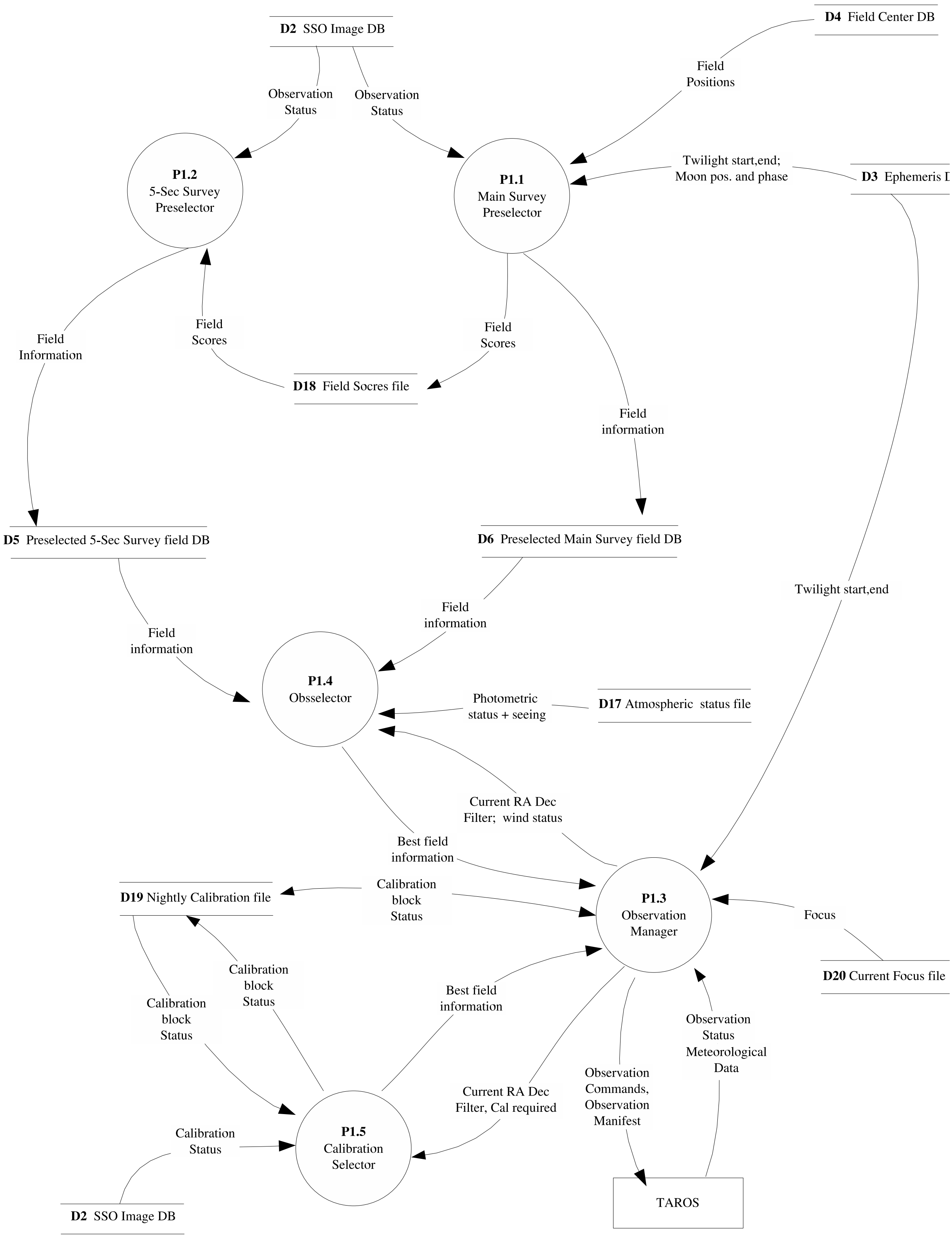


Figure 4 Data flow diagram for P1 Scheduler

6.1.1.4 P1.4 Obsselector

The Obsselector returns the highest ranked field at the time of observation. Its step-by-step functioning is as follows:

1. Obsselector obtains the current date/time and the current telescope configuration (RA,Dec,filter), and generates current MJD.
2. Examine the wind conditions (wind gusts in excess of limits?) and obtain the current photometric condition status and seeing (determined by P2). On the basis of this information decide whether to change the survey type:

Photometric?	Seeing > limit?	Observation Mode
Yes	No	5-Sec Survey
Yes	Yes	5-Sec Survey
No	No	Main Survey
No	Yes	Monitor

   If the conditions are not photometric and the seeing limit is exceeded then a monitoring state is entered. In this state an image is obtained but the data is rejected as poor quality due to excessive seeing. This field is available for rescheduling. No data is retained until seeing returns within bounds or the conditions are determined to be photometric.

3. Get fields currently observable from preselected field DB. If wind is gusting > specified limit then block angles <90° from direction of wind.
4. Score the fields residing in either Preselected 5-Sec Survey field DB or Preselected Main Survey field DB according to the current observing mode:
   a. Main Survey:
      i. 4hr interval possible
      ii. 4hr interval complete
      iii. moon position (as calculated by Preselector)
      iv. time left in survey (as calculated by Preselector)
      v. current telescope pointing
      vi. complete another interval
      vii. complete 3/6 images for data release.
   b. 5-Sec Survey:
      i. moon position (as calculated by Preselector)
      ii. time left in survey (as calculated by Preselector)
      iii. complete an image in 6 filters for data release
      iv. current telescope pointing
      v. requirement for standard star sequence every hour. We join to the preselected fields a list of standard star fields that are

currently observable and score these relative to the time that standards were last acquired. We aim to obtain a standard sequence every hour.

5. Return the telescope/instrument configuration of the highest scoring field.
6. Modify the field centre to achieve a dither pattern (an assignment of offsets in RA and Dec according to the number of images currently obtained for the field in question).

6.1.1.5 P1.5 Calibration Selector

P1.5 is triggered by the Observation Manager P1.3. It has four use cases:

1. obtain biases,
2. obtain dome flat,
3. obtain twilight flatfields and,
4. conduct focus sequence

P1.3 will supply the use case. P1.5 will interrogate the Nightly Calibration DB:

- if the current calibration has been attempted but no status has yet been returned by P2, P1.5 will check SSO Image DB for the required calibration data in the current night:
    - if all has good QA status from P2 then update the Nightly Calibration DB that calibration obtained successfully
    - if some failed QA then schedule the required number to complete the calibration
    - if some have no QA status then return a null observation block and the system will be idle.
- If the current calibration has not yet been attempted then P1.5 returns an observation block which defines the required observation (i.e. RA, Dec, filter and exposure times):
    1. Bias: set of 5 biases at park position;
    2. Dome Flat: TCS has a domeflat command in its API, obtain a set of 3 per filter;
    3. Twilight Flat: track on a relative arbitrary patch of sky and offset (30 arcsecs) between a set of at least three exposures. The exposure time will be determined as a function of time from the start of the twilight flatfielding observation period (see P1.3 table). This function will be determined by experiment. If the so calculated exposure time is $t>10$ sec or $t<0.5$ sec then the exposure will not be attempted for a given filter. If no filters are within $0.5<t<10$ sec then a null observation block is returned an the system is idle.
    4. Focus: This is a set of 5 pointings each separated in Dec by ~20 arcsec with a series of focus increments centred on the nominal focus position determined from the focus lookup table. The field will be centred such that the sequence for a bright ($8<V<10$) star is entirely located on a given central chip. (It is probable that TAROS will implement a focusing sequence in its API.)

P1.5 then updates the status of this calibration as attempted in Nightly Calibration DB.

6.1.2 P2 Quicklook Quality Assurance

P2 transfers pertinent header information to SSO image DB (D2) and computes basic quality assurance checks. This occurs as images appear in D1.

1. **Science data quality checks**: the data is examined for seeing, photometric zeropoint and sky background. This is done as follows: we apply an existing perl module to extract the zeropoint and seeing for a central chip. This module first applies a WCS (as described in Section 6.2.1.2.3.3) to the image. Then aperture photometry is performed on the objects in the image, and an aperture correction is determined and applied. The seeing is determined at this stage and the current seeing is written to D17 - Atmospheric Status File. The stars in common with 2MASS and/or Tycho (as appropriate for the filter in question) are then used to define a zeropoint for the instrumental system.

   P2 monitors photometric conditions and will enable the scheduler to switch from the 5-Sec Survey to the Main Survey during the night. Checks will commence once a sufficient number of quality assurance zeropoints (on order of 5) have been obtained for a filter. This photometric status is written to D17 - Atmospheric Status File.

   The sky background of all chips is determined (a clipped median) together with a median of each chip's overscan (or bias level). The relative sky levels of each chip in each filter will be a known constant. By comparing the observed values with this known set of values we have a useful check of the sky quality – in particular, it will be apparent if the observation was not of spatially uniform depth due to cloud unevenly obscuring the detector. The comparison of the overscan to nominal values gives a useful check on the physical condition of each chip.

2. **Calibration data quality checks**: here we check that:
   a. bias frames possess nominal levels
   b. Flatfield (dome and twilight) counts are within useful range (not too low; not saturated)
   c. Focus sequence – we actually calculate the optimal focus position from the focus image. This utilises legacy software from the automated 50”. The current focus is maintained in Current Focus File D20

P2 decides whether the data passes the basic quality assurance and returns this status along with quality assurance parameters to the SSO Image DB. P2 also updates the status of the field in the Preselected Field DB which is used by the Scheduler. In this way, a field which fails quality assurance is available for reobservation without imposing excessive latency that would be imposed if the QuickLook result were required prior to the next Obsselector operation.

Images which have completed P2 and have TAROS confirmation of transfer to ANUSF are deleted from the raw data repository.

6.1.3 P3 SSO Bad Observation Manager

P3 updates the status of images listed in the Bad Observation file (those images which have failed to be processed by the Science Data Pipeline). The Bad Observation file originates in the Science Data Pipeline and is pushed to the Scheduler System by P7. These images will be ready for rescheduling with the next call of P1.1.

6.2 SkyMapper Science Data Pipeline

6.2.1 P4 Data Reduction

6.2.1.1 P4.1 Manifest Manager

If the nightly Manifest file is present then check that all images contained in the manifest are present in ANUSF Raw Data. If any file is not present then the process sleeps for 5 minutes then exits (and polls again). If Manifest file is old (say 2 days) then flag operator that data is not present: operator to intervene to clarify. If all files are present then insert the new observations into ANUSF Image DB, together with quality assurance information.

Science Data Pipeline: System Level

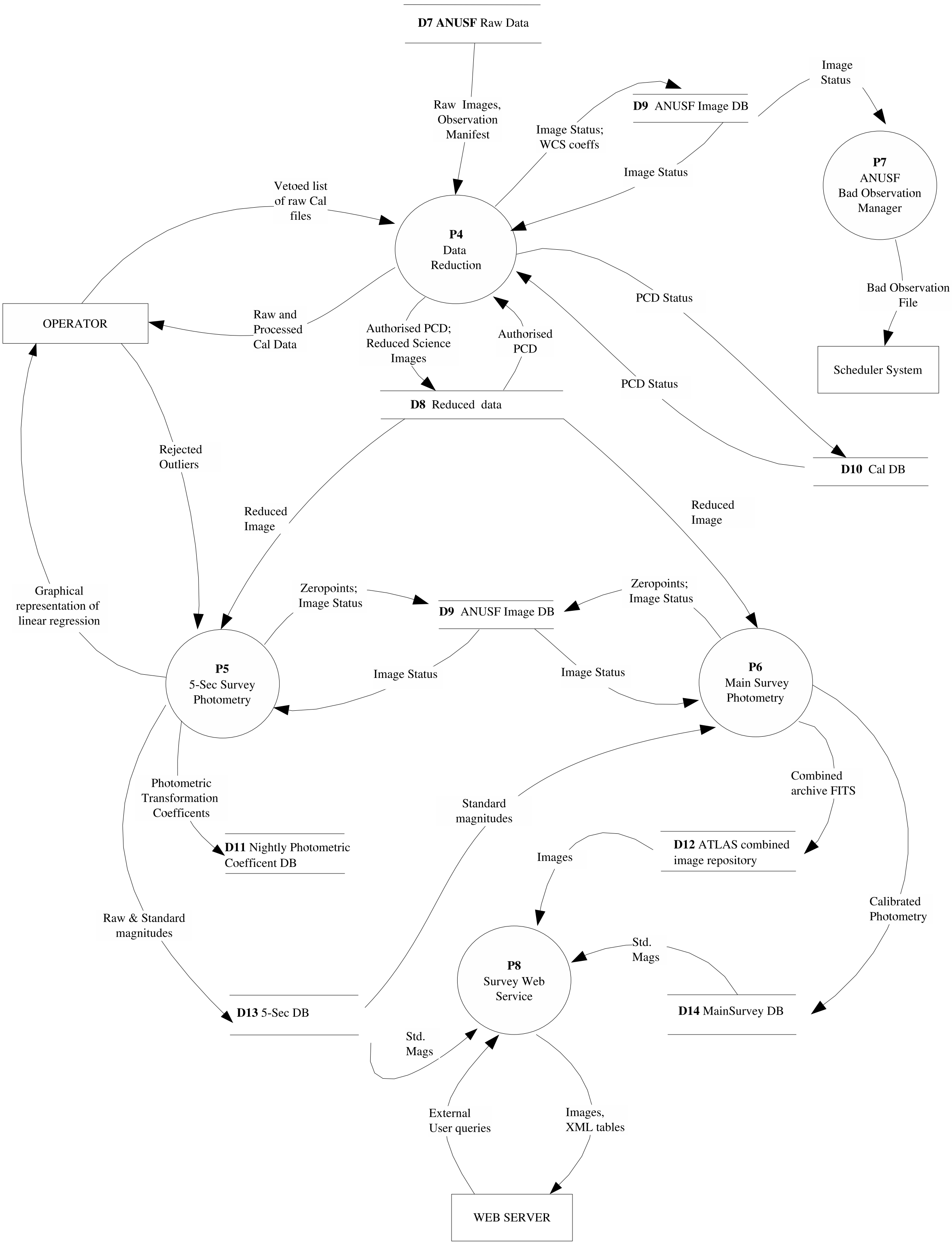


Figure 5 System level data flow diagram for the Science Data Pipeline System

P4 Data Reduction

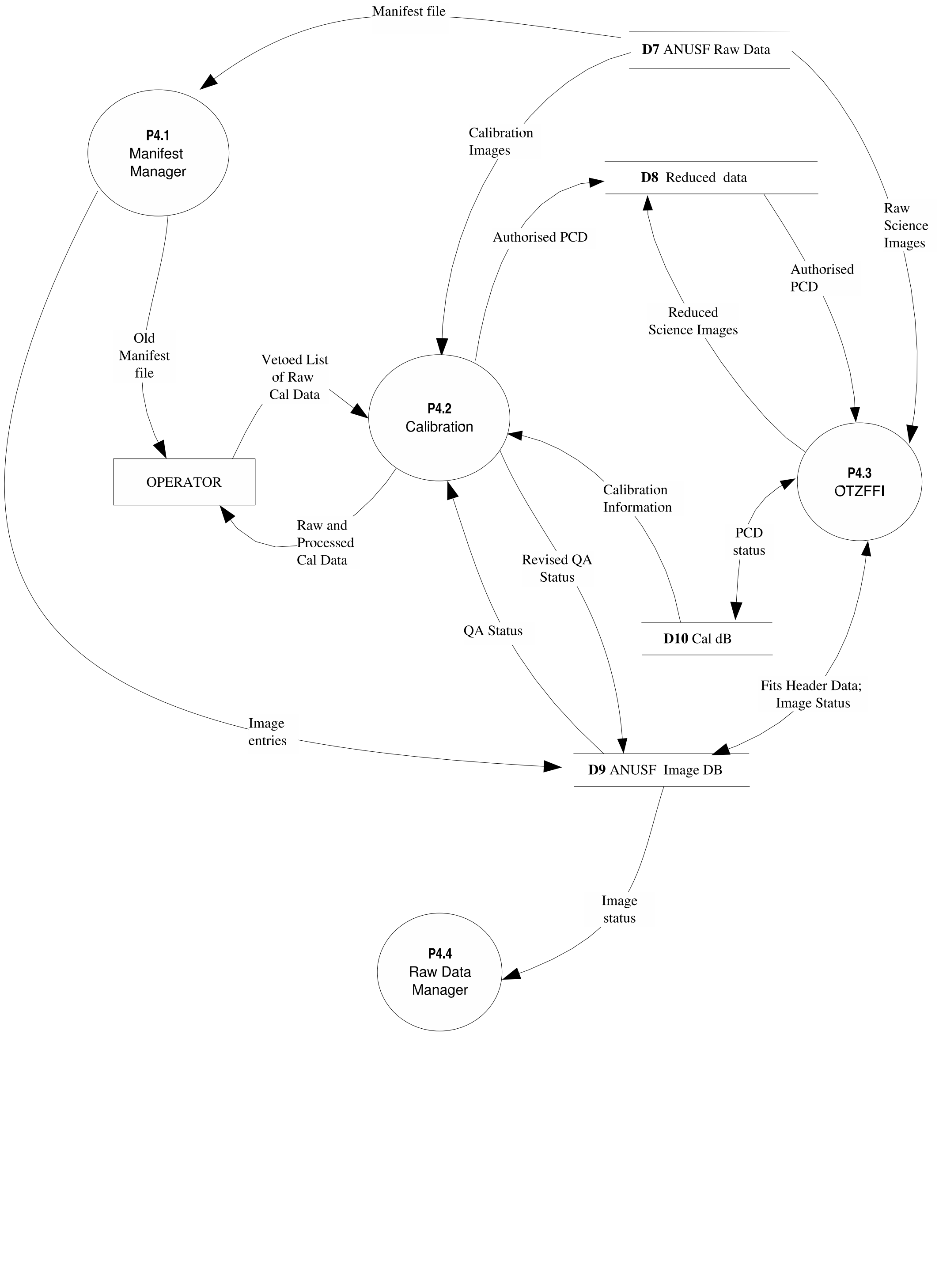


Figure 6 Data flow diagram for P4 Data Reduction

P4.2 Calibration

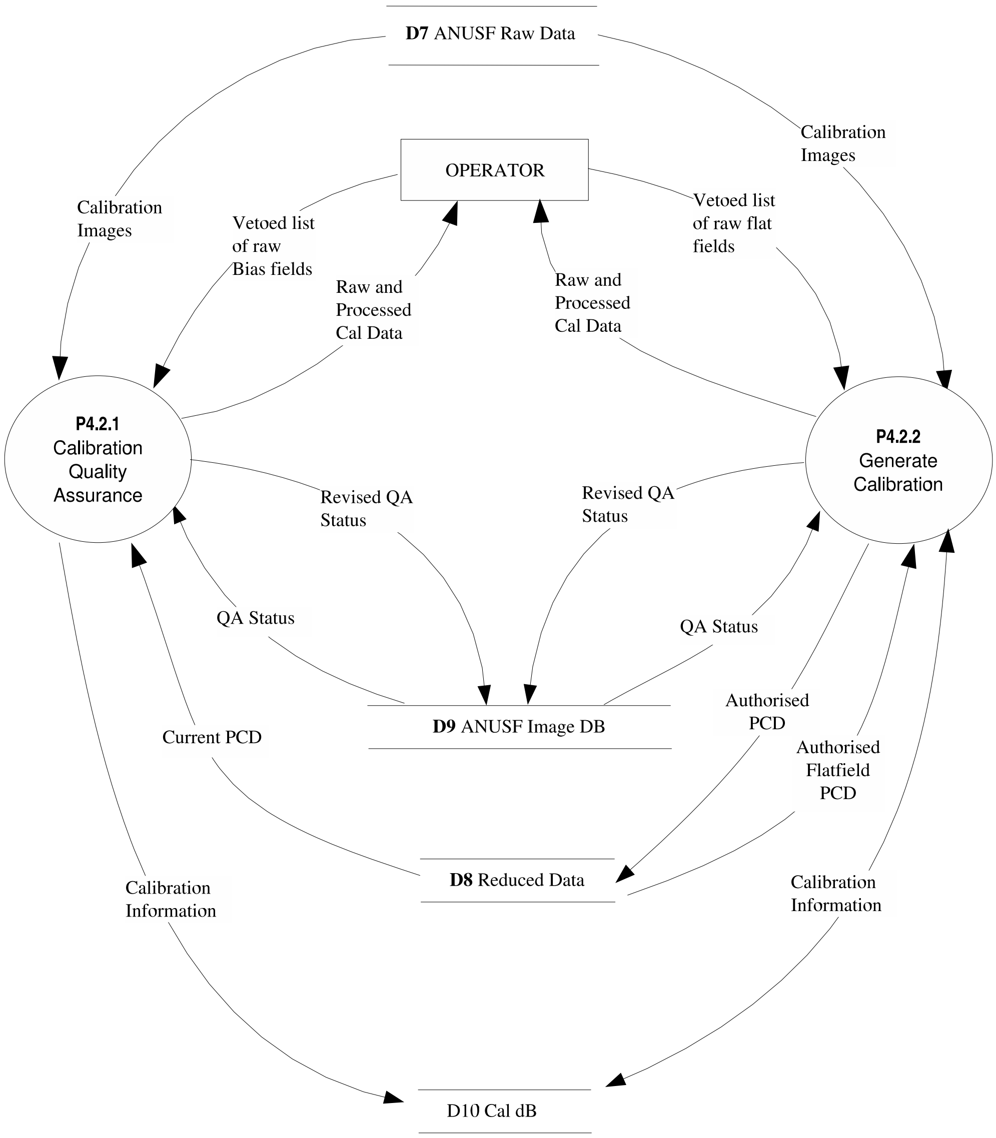


Figure 7 Data flow diagram for P4.2 Calibration

6.2.1.2 P4.2 Calibration

6.2.1.2.1 P4.2.1 Calibration Quality Assurance

6.2.1.2.1.1 P4.2.1.1 Generate Zero Calibration Data

This process (Figure 8) generates a median bias frame from those available from the preceding night as described in ANUSF Image DB. Bias frames with good Quicklook quality assurance status are used.

6.2.1.2.1.2 P4.2.1.2 Compare Bias Frames

P4.2.1.2 compares the processed calibration data (PCD) from P4.2.1.1 with the current PCD bias frame as described in Cal DB.

If the comparison is bad then the operator is sent the raw bias frames, the PCD bias and the current PCD bias and asked to flag bad raw bias frames to attempt to rectify the bad comparison. This vetoed list is then marked as bad quality in ANUSF Image DB and P4.2.1.1 is called. The procedure loops until all raw bias frames are rejected or the PCD bias passes the comparison.

If the comparison is good then the night's bias frames become part of the raw cal data buffer which will be used in the next P4.2.2 Update Calibration. This raw cal data buffer will be implemented as a temporary and dynamic database.

Once passed quality assurance, P4.2.1.2 then attempts to authorise the current PCD. First a check of the expiration date of the current PCD is made. If the current PCD has not expired it is authorised for the night (the validity date of the current PCD is updated to be valid for the current night).

If the operator rejects all raw bias frames or the current PCD has expired then the system has changed – no authorised PCD is flagged in Cal DB.

6.2.1.2.1.3 P4.2.1.3 Generate Twilight Flatfield Calibration Data

Generates a median flatfield image from those available from the preceding night. Flatfield files with good Quicklook quality assurance status are used.

6.2.1.2.1.4 P4.2.1.4 Compare Flatfield Frames

Acts identically on flatfields as P4.2.1.2 does on bias frames.

P4.2.1 Calibration Quality Assurance

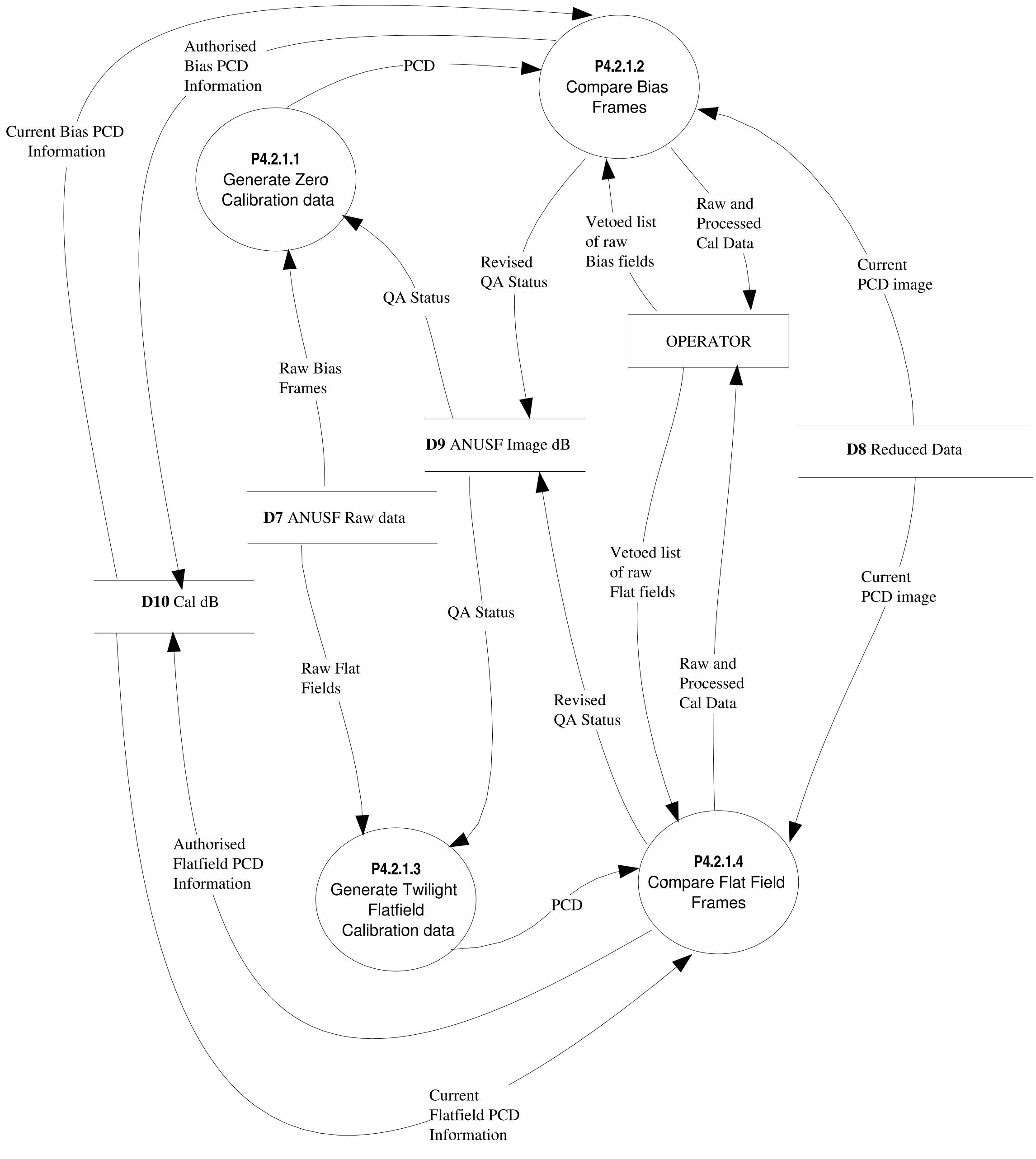


Figure 8 Data flow diagram for P4.2.1 Calibration Quality Assurance

6.2.1.2.2 P4.2.2 Generate Calibration

6.2.1.2.2.1 P4.2.2.1 Maintain Calibration Data

P4.2.2.1 (Figure 9) performs a check on the availability of current PCD for all instrument configurations. P4.2.2.1 triggers P4.2.2.2-4.2.2.5 as required to attempt to complete the set of current PCD.

6.2.1.2.2.2 P4.2.2.2 Generate Zero Calibration

Using the data in the raw cal data buffer generates a bias frame.

6.2.1.2.2.3 P4.2.2.3 Generate Twilight Flatfield

Generates a median of the available twilight sky flatfields for a filter.

6.2.1.2.2.4 P4.2.2.4 Generate Fringe Frame

Generates a fringe frame (filters I,Z only). To do this it selects a random sample of raw science images in a given filter (around 100-200 images) from the previous valid time frame, then applies the authorised PCD bias and flatfield to this image list. To construct the fringe frame we firstly construct the superflat which is a clever median/masking of stack of randomly selected images. By subtracting the appropriately scaled sky flat from the superflat we create a fringe-only frame.

6.2.1.2.2.5 P4.2.2.5 Generate Illumination Correction

To do this we must have a series of science frames of a dithered sequence. These are then processed using the current PCD (bias, flats, fringe). Delta magnitudes are formed for objects as they are placed over the focal plane array. The relative magnitude of each star in each chip is compared to the average of the star observed over all chips, and the smoothed map of average deviation of stellar photometry made for each CCD to define the Illumination Correction.

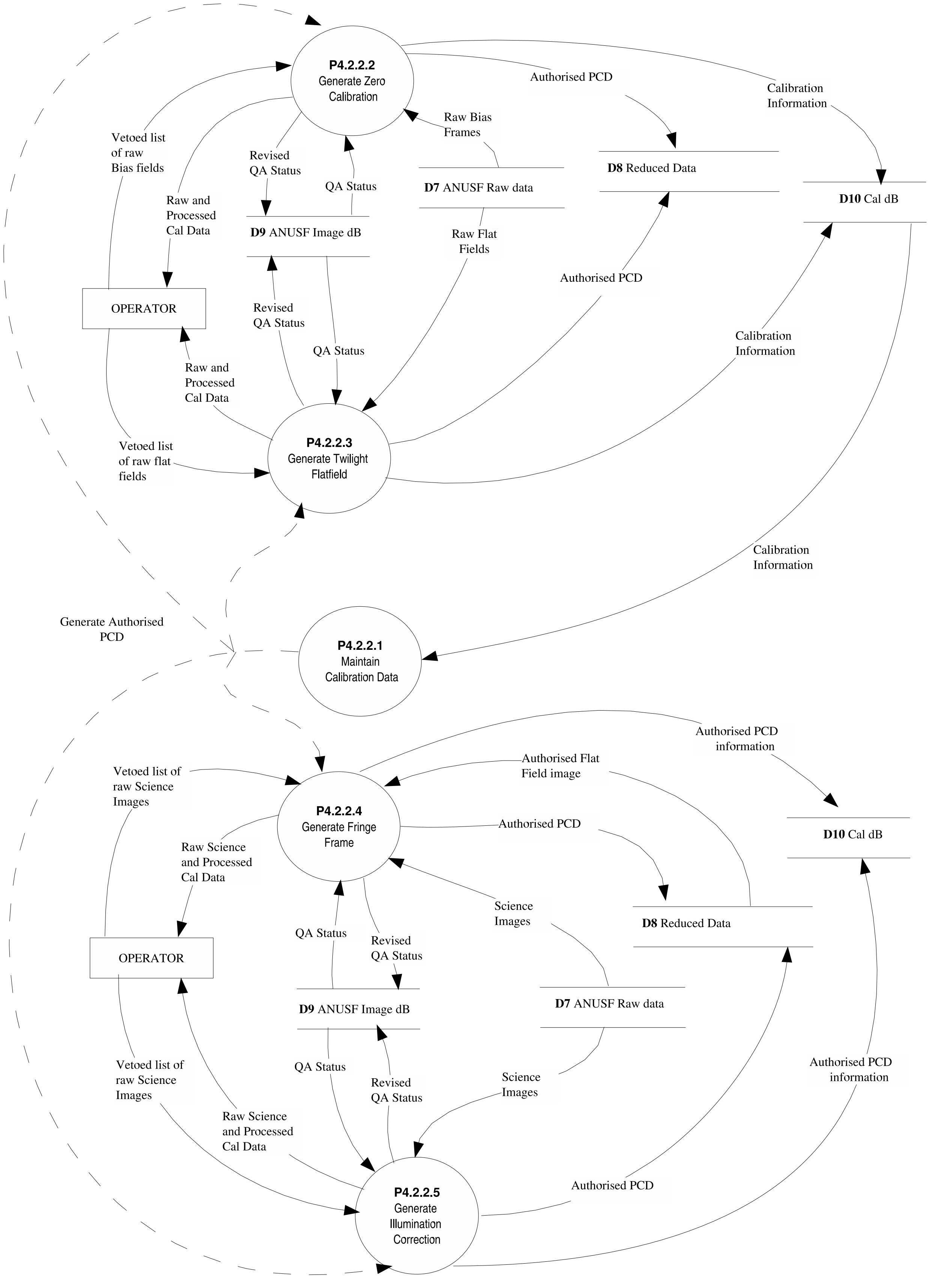


Figure 9 Data flow diagram for P4.2.2 Generate Calibration

6.2.1.2.3 P4.3 OTZFFI:

6.2.1.2.3.1 P4.3.1 OTZFFI Manager

See Figure 10. For all unreduced science images in ANUSF Image DB check if authorised PCD files exist for image configuration (amp,filter) from Cal DB. If they do then pass list of raw data with associated authorised PCD files to OTZFFI P4.3.2.

6.2.1.2.3.2 P4.3.2 OTZFFIProc

P4.3.2 implements IRAF CCDPROC to apply authorised PCD and the linearity correction to science images and outputs reduced data to Reduced Data and updates reduction status in ANUSF Image DB.

6.2.1.2.3.3 P4.3.3 DoWCS

P4.3.3 applies a World Coordinate System (WCS) to each amp image (modifies header). Updates ANUSF Image DB with WCS coefficients for each amp. These coefficients facilitate searches for objects within a given field.

To place a WCS on an image, a reference catalogue (UCAC2) is compared to a catalogue list of stars. We use Source Extractor (TeraPix Consortium) to make the catalogue, a starmatcher (Schmidt) to match the catalogue, and IRAF to fit a ZPN (Tangent plane projection with radial polynomial distortion terms). We are in discussion with the Cambridge Survey Unit to utilise their proven code for this element.

6.2.1.2.4 P4.4 Raw Data Manager

P4.4 (see Figure 6) examines ANUSF Image DB for images with reduced status. These images are deleted from ANUSF Raw Data.

P4.3: OTZFFI

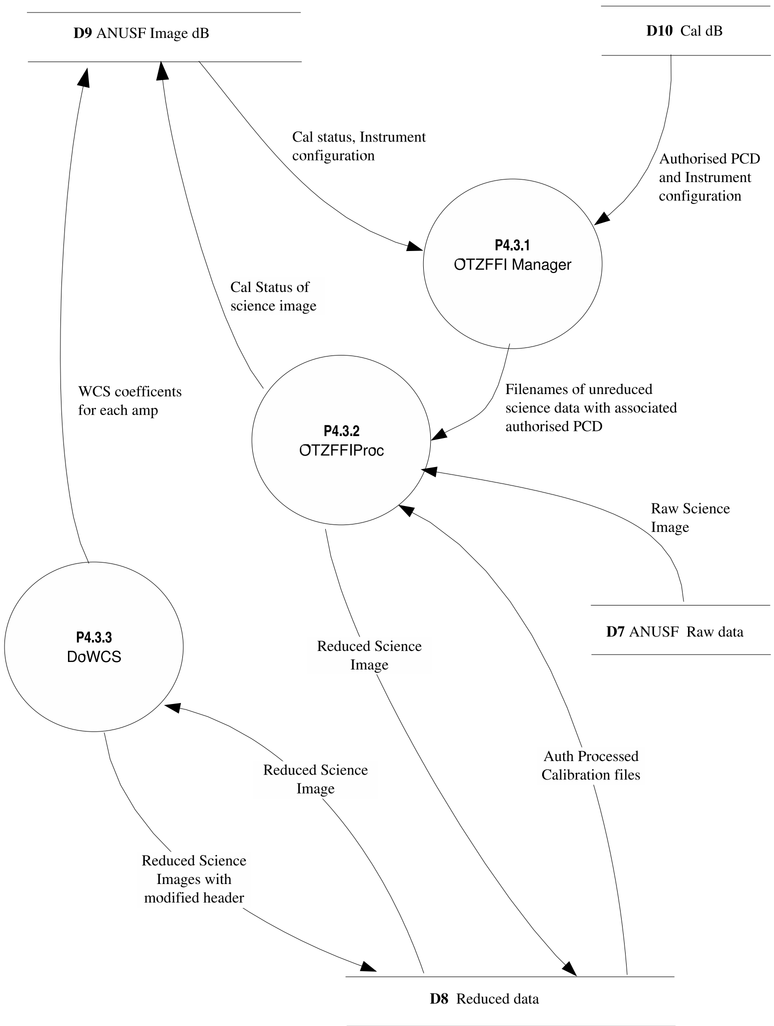


Figure 10 Data flow diagram for P4.3 OTZFFI

6.2.2 P5 5-Sec Survey Photometry

6.2.2.1 P5.1 DoPhotom

6.2.2.1.1 P5.1.1 DoPhotomCheck

Polls for a night of data without verified photometric status. The module examines the zeropoints generated by P2 for the night in question in each filter. Was the night photometric? If not, reject observations (a sigma clip) which stand out as bad. If there is a general high dispersion of zeropoints then reject the night – flag all science images for the night in Image DB as non-photometric. Else flag the science images as photometric.

6.2.2.1.2 P5.1.2 DoDOPHOT

Executes our own modified version of the dophot code on the images passed as photometric. Using the brightest, unsaturated sources in the image, an ∞ aperture correction to an aperture twice the fwhm will be calculated. This correction will then be applied to the aperture photometry of each point source in the image. The resultant tables of raw magnitudes are ingested into 5-Sec DB.

P5: 5-Sec Survey Photometry

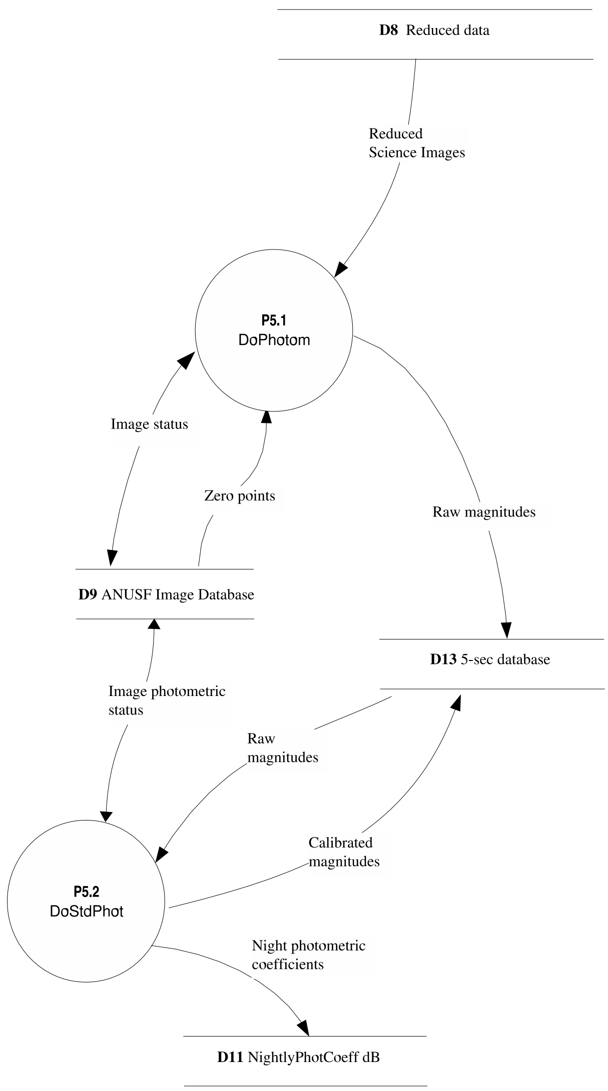


Figure 11 Data flow diagram for P5 5-Sec Survey Photometry

P5.1: DoPhotom

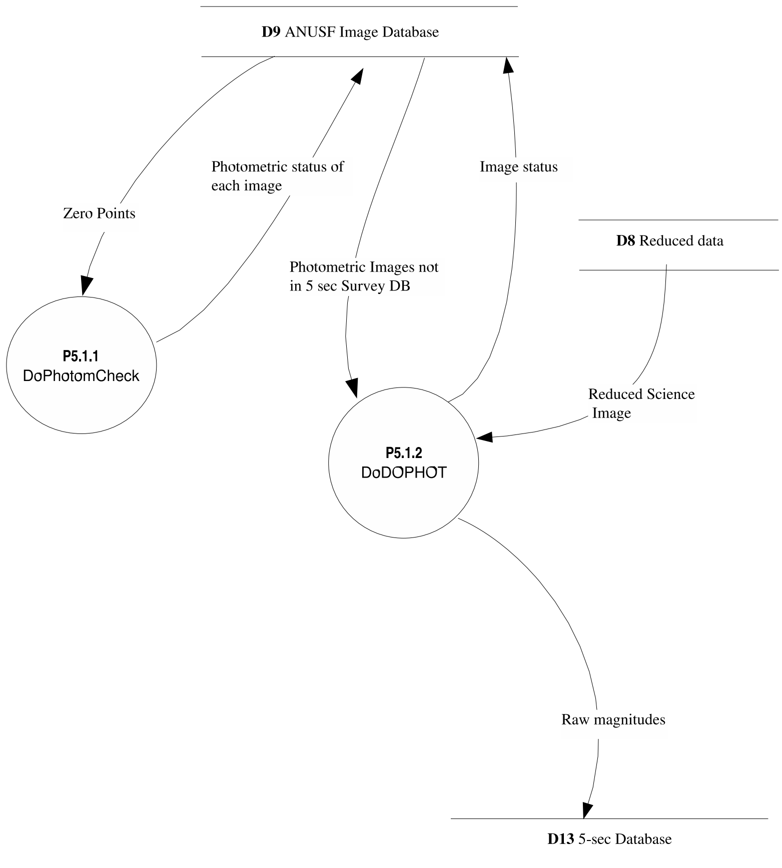


Figure 12 Data flow diagram for P5.1 DoPhotom

6.2.2.2 P5.2 DoStdPhot

6.2.2.2.1 P5.2.1 DoStandardiseNight

This process derives the transformation coefficients for the night and stores them in the Nightly Photometric Coefficient DB.

For each filter, *i*, solve $m_i = m_i^{raw} + \alpha_i X + \beta_i X(m_i - m_j) + \gamma_i$ using the raw magnitudes from 5-Sec DB, derive α,β,γ for night, or optionally fix β as quantity is not expected to vary rapidly. The operator interacts with the procedure via a simple GUI: reject outliers from the regression fit or flag the night's images as not of suitable quality for the 5-Sec Survey. For nights with a good fit the 5-Sec Survey photometry status flag of each image is modified to indicate that the data can proceed into the 5-Sec DB. The derived α,β,γ for the night is written to Nightly Photometric Coefficient DB.

6.2.2.2.2 P5.2.2 Do5secMagCalc

This procedure accesses the Nightly Photometric Coefficient DB for transformation coefficients appropriate for an image and transforms raw magnitudes to standard instrumental magnitudes accordingly. The derived standard instrumental magnitudes are placed in the 5-sec DB. A 5-Sec Survey standardised photometry status flag is set upon completion.

P5.2: DoStdPhot

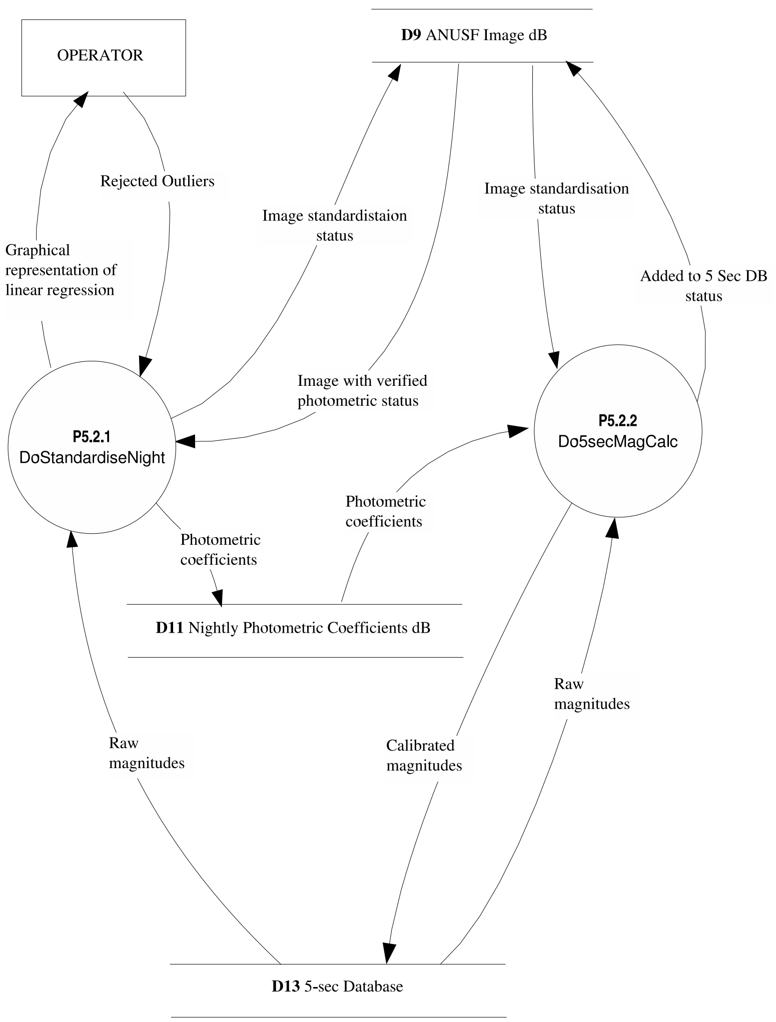


Figure 13 Data flow diagram for P5.2 DoStdPhot

6.2.3 P6 Main Survey Photometry

6.2.3.1 P6.1 Main Survey Photometry Manager

P6.1 examines the ANUSF Image DB for an image of appropriate depth (for first data release 3 images, second data release 6) in a filter, not yet incorporated into the Main Survey DB, and for which 5-Sec Survey data also exists. Its action is to trigger P6.2 via OPUS.

6.2.3.2 P6.2 DoStdFrame

uses those stars in common with the 5-Sec Survey to place the instrumental photometry of the three frames on a standard instrumental system. This is a trivial zeropoint offset. The zeropoint offsets are written to the ANUSF Image DB.

6.2.3.3 P6.3 DoCombine

Combines images using Swarp (http://terapix.iap.fr/rubrique.php?id_rubrique=49). For the subsequent second data release six images will be available allowing a deeper template to be created. The combined image is then placed in our ATLAS combined image repository D12.

6.2.3.4 P6.4 DoCreateObjCat

P6.4 masks saturated stars and grows any bleeding associated. Source Extractor (http://terapix.iap.fr/rubrique.php?id_rubrique =91/) is then implemented for object detection. This reaches to 5σ in each combined filter image.

At high galactic latitudes this is done for all filters. The union of all six filter detections is formed. This list is culled according to the following criterion: all objects within 1 arc second are to be considered the same object. This leaves a list of unique objects.

At low galactic latitudes the object list is determined from the *i*-band combined image.

The unique object list resides in D15 Unique Object DB.

6.2.3.5 P6.5 Transient Object Detector

Transient objects will not appear in the unique object list as this is formed from the median image(s). To detect transient objects we examine the individual images for new objects. First, the objects described in P6.4 are masked according to their positions and shape parameters. We do not attempt to subtract the objects as this is computationally too expensive. We then search for objects 10σ above the background with appropriate shape parameters relative to the PSF. These objects are candidate transient objects. These detections are appended to the Unique Object DB.

P6: Main Survey Photometry

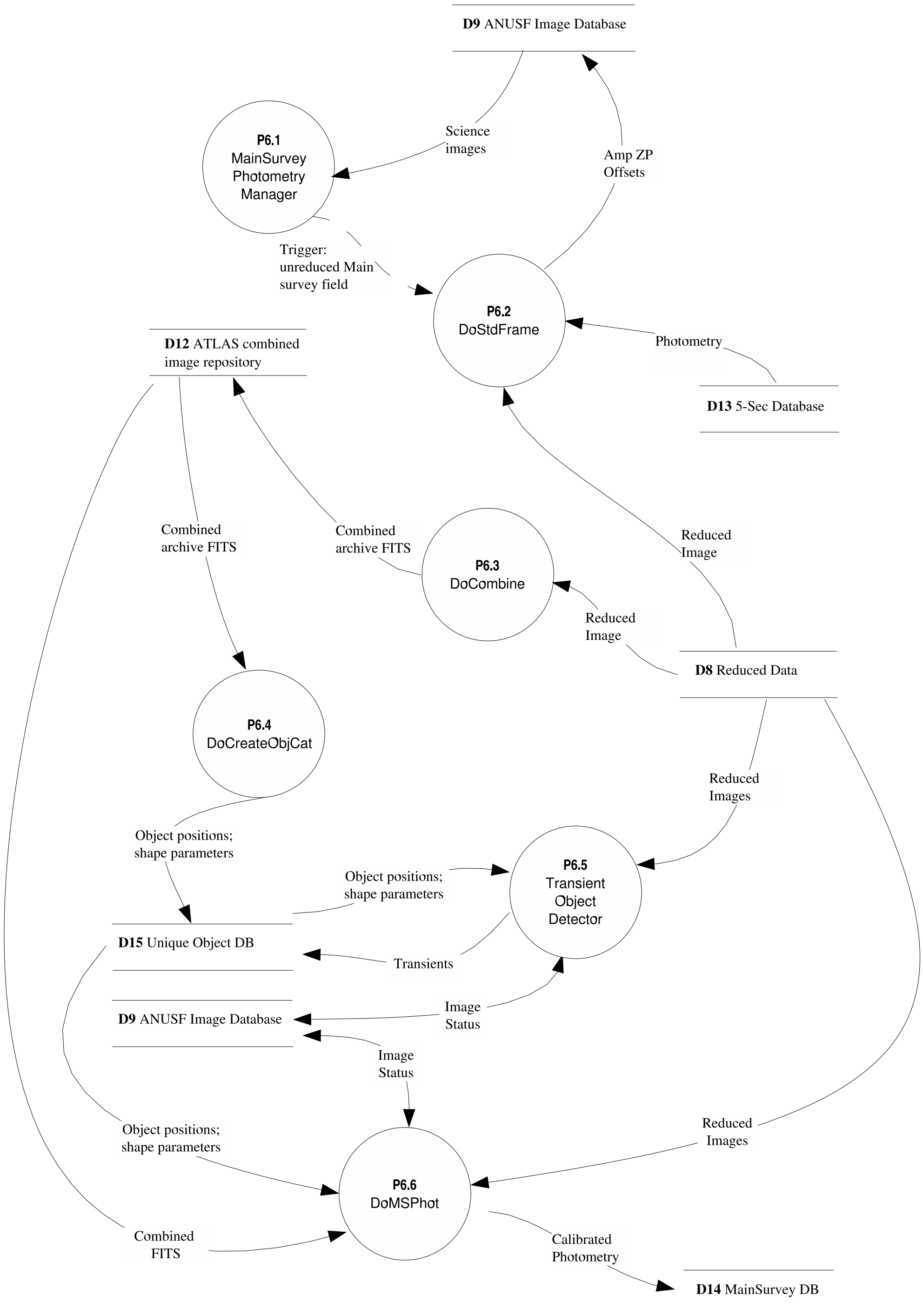


Figure 14 Data flow diagram for P6 Main Survey Photometry

6.2.3.6 P6.6 DoMSPhot

For each filter, image and amp we determine the PSF shape parameters and perform PSF-fitting photometry (using our own modified version of the dophot PSF-fitting routine) in two modes:

1. using the fixed positions from the unique object list
2. allowing the position of each object to drift somewhat (centroid)

For each object we then inspect if the object presents a reasonable fit to the PSF shape across all images in a filter. This provides the basis for object classification between star/galaxy/other.

We then treat each object as a galaxy. We apply Source Extractor on the combined image in all filters. This is applied in two modes:

1. determine the galaxy shape parameters (i.e. position angle and isophot of ellipse) from the band with the most signal. With the fixed shape parameters then determine magnitudes.
2. allow shape parameters to be determined independently in all filters and determine magnitudes.

The object type, photometry and shape parameters are written to the Main Survey DB.

6.2.4 P7 ANUSF Bad Observation Manager

This process interrogates ANUSF Image DB to find images which have failed processing in the Science Data Pipeline. Images so found are written to a Bad Observations file. This file is then conveyed to the Scheduler System via scp (See Section 6.4).

6.2.5 P8 Survey Web Service

The Survey Web Service provides a catalogue search and image server facility. Here we detail the design of the server-side processes.

ImageServer: User provides (RA, Dec, image size and filter) via Image Server Web Interface. Maximum image size allowable is set to 30' square due to file size. Using the RA,Dec the ALTAS images containing the object is located (a simple look up table of field centres). Then the boundaries of the requested image size box are considered. If the requested image is wholly contained in the ALTAS image then the image is cut out and WCS updated accordingly and returned to the user. If not wholly contained then portions of adjacent images are directly copied to the output image and the WCS updated.

Catalogue Search: Allows RA, Dec, search radius type SQL queries of the 5-Sec and Main Survey database tables.

6.3 Description of Data Locations

The following section provides a summary of the data stores discussed in the design. For the related tables in the Science Data Pipeline we provide an entity relation diagram shown in Figure 15.

D1 SSO Raw Data: files from instrument. This is a transient location for the raw data. Data remains here until quality assurance and TAROS has transferred to ANUSF. It is then deleted.

D2 SSO Image DB: table containing information about each image obtained, FITS header info, and quality assurance data.

D3 Ephemeris DB: precomputed ephemeris data for SSO such as astronomical twilight start and finish and moon position and phase for the lifetime of the SkyMapper project.

D4 Field Centre DB: a table containing the field centres of all SkyMapper fields.

D5 Preselected 5-Sec Survey DB: a table containing field information for preselected 5-Sec Survey fields

D6 Preselected Main Survey DB: a table containing field information for preselected Main Survey fields

D7 ANUSF Raw Data: transient raw data store at ANUSF Petabyte Array. Data remains here until reduced or flagged rejected whence it is deleted.

D8 Reduced Data: OTZFFI processed science data and processed calibration files. This exists on the ANUSF Petabyte Array.

D9 ANUSF Image DB: DB containing info about each image (see Figure 15).

D10 Cal DB: DB containing information about each generated processed calibration data file.

D11 Nightly Photometric Coefficients DB: contains a record of the derived nightly photometric coefficients that were applied to the data.

D12 ATLAS Combined Image Repository: This is the image repository for the ATLAS web image server. It contains all the combined images for each (field,filter).

D13 5-Sec DB: contains the photometry associated with 5-Sec Survey. Raw and standard mags exist.

D14 Main Survey DB: there will be two released Main Survey databases one for each data release. They will contain a unique object id, RA, Dec, mag, shape parameters, for each object.

D15 Unique Object file: this temporary file contains the positions and shape parameters for all unique objects within a given Main Survey field.

D16 Bad Observation file: contains the image id of images that have failed in the Science Data Pipeline system.

D17 Atmospheric Status file: contains the photometric status of the night as determined by P2.

D18 Field Scores file: a file containing the scores from P1.1 for each field.

D19 Nightly Calibration DB: generated nightly this file contains a list of required calibration files and their observation status and quality control status

D20 Current Focus file: contains the optimal focus position

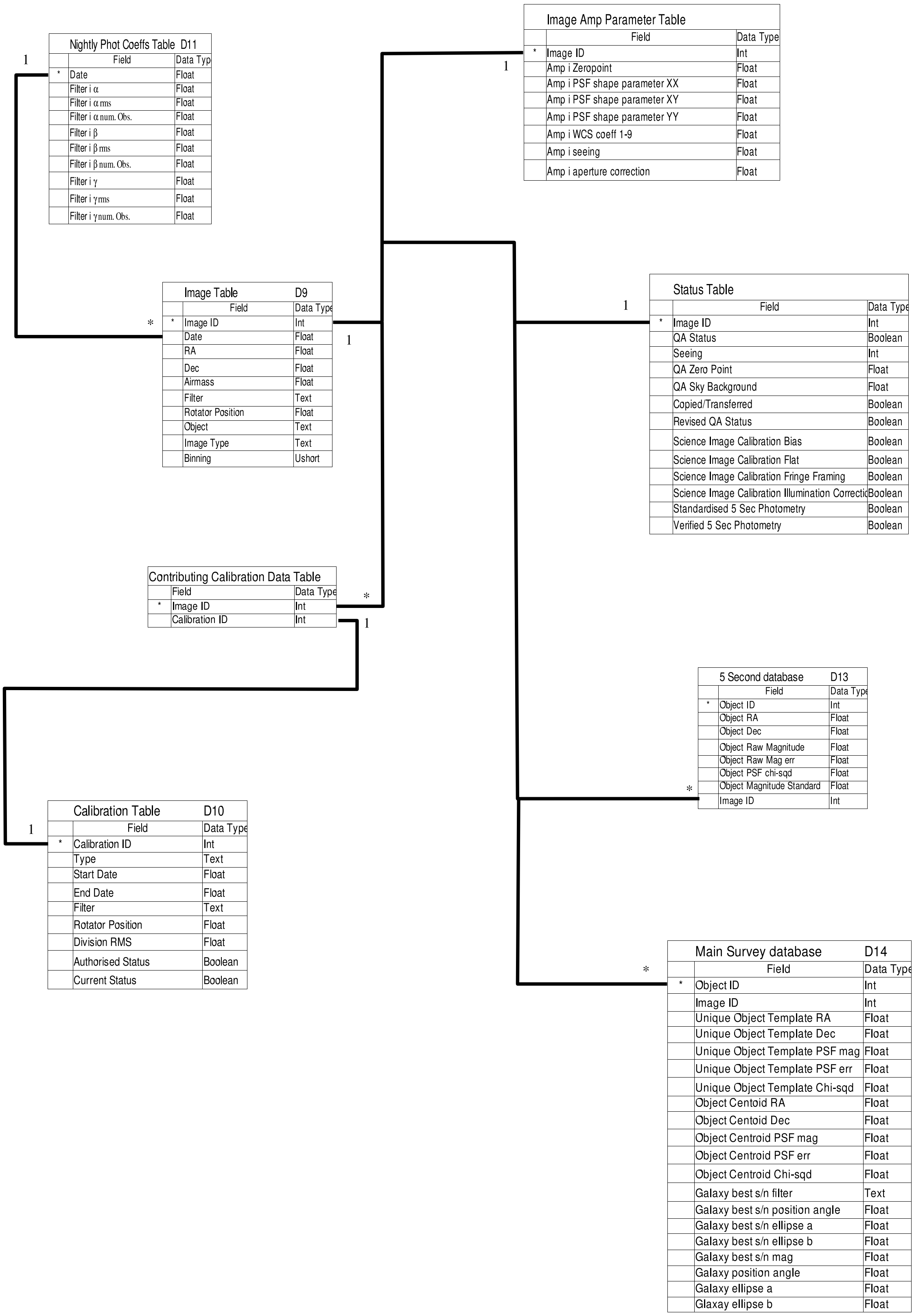


Figure 15 Entity relation diagram for related tables in the Science Data Pipeline

6.4 Interface Design

6.4.1 Scheduler System – Science Data Pipeline

This is the interface between the physically separated systems at SSO and ANUSF.

Science data is trickled to ANUSF via the gigabit link by TAROS. At the end of a night's observing an observation manifest is produced by P1.2 listing all successfully obtained images during the night. This file is transferred by TAROS to ANUSF. This file is the trigger for P4 in the Science Data Pipeline. Should the subsequent processing in the Science Data Pipeline fail for some reason the filename is placed in a Bad Observation file and pushed up to SSO by P7. At the SSO end, when P3 finds a Bad Observation file it modifies the image status in the SSO Image DB as unsuccessful.

The interface ensures robust operation in the event of link outage. If no link exists then the Scheduler System is able to continue obtaining data working with a possibly outdated Image DB. Meanwhile data reduction of previously transferred data can occur at ANUSF with the ANUSF Image DB. When the link is re-established the data flows down to ANUSF, ANUSF Image DB is added to, and the Scheduler System can ingest the Bad Observation file.

6.4.2 Scheduler System – TAROS

The Scheduler System's Observation Manager (P1.3) will communicate with TAROS via XML exchange. TAROS will accept XML RTML protocol which has been designed specifically for remote telescope operation. RTML captures all the pertinent information for the exposure in a series of keywords. TAROS will return status and meteorological data as keywords in XML.

This interface is further detailed in SDN08.03.

6.4.3 Science Data Pipeline – Web

The Survey Web page has an imbedded CGI script to send information to ATLAS combined image repository/database interrogator P8. The CGI script returns the results and displays a reply page to the user.

The Web interface accommodates two use cases:

1. Single object retrieval. The user enters the required RA,Dec directly into the Web form.
2. Multiple object retrieval. The user is able to select batch mode and supply a list of object RA, Dec.

In the case of image retrieval the user is able to select the filter(s) of required object locations and image dimension. In the case of database queries the user can choose the verbosity of database entries returned or can manually select from the full list of database fields.

6.5 Subsystem Commissioning

As construction of the pipeline proceeds each module will be commissioned. The nature of the Science System facilitates this as each module essentially stands alone. Milestones are defined as each module and subsystem (eg. P4.2) is released to LC1. Our test procedure will be one in which we pass artificial data which can be created with a range of likely data flaws and characteristics. The data will be design to stress test the specific module and we will rectify aberrant behaviour.

Included within our work package are blocks of time for the integration/testing of the Scheduler and the Science Data Pipeline Systems. This time will be used to draw out inconsistent behaviour in the systems as a whole.

7 Project Plan

The Science System has been fully incorporated as a work package into the overall SkyMapper project (SDN 09.01). Our work schedule is detailed as part of the total work package. The total time required to implement these tasks is 6020hrs or 3.58 years of effort. With current staffing of Keller (0.8FTE), Schmidt (0.2FTE) and Martin-Jones (0.3FTE) over the time from CDR to first light we have 1.62 years of effort available, a short fall of 1.96 years of effort. We request additional staffing in the form of one computing staff position and one astronomy-based postdoctoral fellow to match the current shortfall.

Skills requirements for additional resources:

Computational person:

- Proficiency in Perl, C, C++, CGI scripting, databases, Linux/Unix Solaris OS

Astronomer person:

- PhD in astronomy
- Proficiency in IRAF, Perl, C
- Extensive experience with stellar photometry
- Observational experience

8 Hardware Requirements and Specification

8.1 Data Processing

Hardware Requirements

In Table 1 we detail the computational expense of the Science System procedures. We estimate the total pass of the 5-Sec Survey through an AMD Operton 2.4Ghz processor is 376 seconds per image, and 13728 seconds per colour per field for the Main Survey. OPUS adds another 16% overhead. For the 4066 fields this amounts to 1.7 and 12.4 CPU years for the two surveys. We require the ability to process data and reprocess data without accreting a backlog.

In the event of a change in the optical system data processing will temporarily halt while new calibration files are developed. A data backlog of no more than 2 weeks is expected. This amounts to at maximum 9.8Tb (12 hour nights of 5-Sec Survey data).

System Specification

We will distribute our pipeline processes on the ANUSF LC1 cluster which consists of 152 (2.66Ghz ~ 63% of the speed of our benchmark machine) processors (see http://nf.apac.edu.au/facilities/lc/hardware.php). The data analysis pipeline for both surveys is suited to distribution over multiple processors due to its “atomic” nature.

The OPUS pipeline manager facilitates this distribution. The total load on the LC1 system from both surveys over the course of their operations amounts to a modest 6% of that available. This allows us to keep up with the data flow as it comes in, and process/re-process data that is backlogged during the day and cloudy nights. The LC1 system provides 15Tb of scratch space which accommodates our maximum expected data backlog.

Table 1 Computational demands of data processing procedures.

Process	Number of instances per colour			Processor seconds per image
	5Sec	MS-FDR	MS-SDR	
P4.3.2 OTZFFIProc	3	3	3	80
P4.3.3 DoWCS	3	3	3	152
P5.1.2 DoDOPHOT	3			144
P6.3 DoCombine		1	2	1352
P6.4 DoCreateObjCat		1	1	150
P6.5 Transient Object Detector		3	6	150
P6.6 DoMSPhot		3	6	500
General Miscellaneous		3	6	120

8.2 Database Operations

Hardware Requirements

We can expect the Southern Sky Survey Main Survey to generate a catalogue of approximately 1 billion stars and galaxies, each observed 36 times. This will much dwarf the Southern Sky Survey 5-Sec Survey. Both surveys will be made public through a searchable database which must be rapidly accessed internally.

System Specification

We expect our database, from experience with other database implementations, to be io-bound rather than processor bound. A single server with a fast ~10 terabyte RAID array will be used to house the main system database at ANUSF.

8.3 Data Storage SSO

Hardware Requirements

The maximum data acquisition will occur in a 12hour night of 5-Sec Survey observing. This will amount to 0.7 terabytes. The system must be able to store a weeks worth of raw data in case of long term link outage.

System Specification

Our system implements a 5Tb RAID array system for data storage at SSO. This will accommodate the maximum load 4.9Tb that will accumulate during a week.

8.4 Data Storage ANUSF

Hardware Requirements

In total the Southern Sky Survey will consist of approximately 150000 images, representing 72 terabytes of raw 16-bit data (compressible by 60%, and an additional 80 terabytes of reduced images data. 12 terabytes of combined reduced images (the ALTAS combined image repository) is to be made accessible to external users.

System Specification

ANUSF will provide the resources for data storage via the Mass Data Storage System (see http://nf.apac.edu.au/facilities/mdss/), a robot-driven tape library capable of storing 1.2 petabytes.

8.5 Data Transfer from SSO to ANUSF

Hardware Requirements

Data (up to 0.7Tb per night) must be provided to ANUSF without producing a backlog of data.

System Specification

An AREN (1Gigabyte per second) link from SSO to ANUSF is under construction. The AREN link could transfer the peak load of data within 2 hours.

9 Commissioning program

In this section we detail our Science System commissioning program. This program is required to characterise the photometric and astrometric properties of the instrumental system. It will also establish a series of photometric and astrometric calibrators around the sky. The commissioning program does not explicitly include commissioning of the software; this is included explicitly in the software work package.

9.1 Establishment of 5-Sec Survey Photometry Reference Fields

In order to provide accurate standardised photometry over our entire field of view we will require a series of reference fields around the sky. These fields will be chosen to be rich stellar fields of low galactic latitude. These fields will be observed once per night during 5-Sec Survey operation. The fields will provide calibrated photometry over an entire SkyMapper field-of-view. From analysis of this data we will describe the relative offsets of each chip. The chip offsets will then be considered constant for the night and standard star observations can henceforth be obtained on an individual chip. The nightly monitoring of a series of reference fields also provides another useful check on the instrumental system.

The establishment of a series of photometric reference fields will take place in the commissioning program. The fields will be centred on existing standard star fields (Landolt). The standard star sequence is dithered over the focal plane so as to place the standard star sequence in each CCD. This is done for all filters. To establish two photometric reference fields will take a night of photometric conditions to complete. As the survey progresses and the photometric reference fields become inaccessible subsequent half-nights will be required to establish additional photometric reference fields (6 and 9 months after the start of data taking).

The photometric reference fields will provide the basis for our determination of illumination correction. When required we can achieve this calibration by the acquisition of a single field.

9.2 Astrometric Corrections

We will acquire images of a photometric reference field with four instrument rotator angles each separated by 90 degrees. The astrometry derived from the four configurations will provide a shake-down of our WCS determination system. If required, additional higher order radial terms may be implemented.

9.3 Definition of Photometric Transformations to Standard Systems

The establishment of photometric reference fields will also entail the observations of a large number of existing standard star fields. We will use this photometry to derive transformations between our instrumental system and those of Cousins, Landolt, Hipparchos and DDO38.

10 Costing

In this section we detail non-labour related costs incurred in the construction and operation of the SkyMapper Science System.

AREN link: assume no cost

ANUSF: Tapes $144k, data server + disk $20k. A LIEF grant has been submitted to cover this cost.

RAID array: 2 x 5Tb systems at $15k each

Linux box: for Scheduler System operation $4k

System commissioning: 2 trips of 3 people to SSO, duration 1 week $5.5k

System maintenance: 3 trips per year, 1 person, 1 week $2.6k

11 Risk Analysis and Mitigation

The following are identified risks and how we plan to manage them.

11.1 OPUS

Availability of source code: OPUS is distributed as built executables for various operating systems. OPUS is under active development and support by STScI. Cessation of this support presents a risk to the SkyMapper Science System. Should it become necessary to update the Linux install on the ANUSF LC1 cluster and no matching make of OPUS is available the pipeline will break.

Risk: Moderate

Action: Seek source code from STScI. Brian Schmidt has initiated request

11.2 AREN Link

Availability of link: link may not be available at commencement of survey or service may be disrupted during survey for a period in excess of a week.

Risk: Low

Action: In the advent of AREN failure (or delay in establishment of this service) we will supplement the current 1x5Tb RAID system with two identical systems. The raw data will be mirrored on two systems and one RAID system will be shipped to ANUSF per week.

11.3 ANUSF

Funding basis for ANUSF involvement: Currently we do not have the money for required tapes and data server to be placed at ANUSF. There is a LIEF grant currently under consideration to cover this expenditure.

Risk: Moderate

Action: Currently none available.

Availability of LC1 Linux cluster: This resource is allocated via a competitive application system.

Risk: Low

Action: Apply for time on LC1 for initial install of OPUS.

11.4 Staffing

1. The departure of Brian Schmidt and/or Stefan Keller represents a significant one-point failure risk due to their overall system management.

Risk: Low

Action: None

2. Availability of additional staffing resources (see Section 7). Resources must be acquired in early 2006 to make up current staff shortfall.

Risk: Low

Action: Ensure recruitment commences as soon as possible following CDR approval.