

VISTA Data Flow System: Status

Jim Emerson^{*a}, Mike Irwin^b, Nigel Hambly^c

^aAstronomy Unit, Queen Mary University of London, Mile End Road, London E1 4NS, UK

^bInstitute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK,

^cInstitute for Astronomy, School of Physics, University of Edinburgh, Edinburgh EH9 3HJ, UK

ABSTRACT

Data from two IR survey cameras UKIRT's WFCAM and ESO's VISTA can arrive at rates approaching 1.4 TB/night for of order 10 years. Handling the rate, and volume of survey data accumulated over time, are both challenges. The UK's VISTA Data Flow System (for WFCAM & VISTA near-IR survey data) removes instrumental artefacts, astrometrically and photometrically calibrates, extracts catalogues, puts the products in a curated archive, facilitates production of user-specified data products, and is designed in the context of the Virtual Observatory. The VDFS design concept is outlined, and experience in handling the first year of WFCAM data described. This work will minimize risk in meeting the more taxing requirements of VISTA, which will be commissioned in 2007. Tools for preparing survey observations with VISTA are outlined.

Keywords: surveys, data quality, pipeline, science archives, survey area definition tool, VISTA, WFCAM, VDFS

1. INTRODUCTION

The Visible and Infrared Survey Telescope for Astronomy (VISTA) is a wide-field survey telescope with a 4-m primary, feeding a 1.65 deg diameter field of view near-IR camera equipped with a 4x4 array of 2048x2048 pixels Raytheon VIRGO detectors which will be operated in queue scheduled mode and is due for on sky commissioning at the European Southern Observatory's (ESO) Cerro Paranal Observatory in early 2007 (Emerson et al 2004³). Handling the very large data rates (the mean data rate for VISTA is estimated ~315 GB per operational night depending on the observing modes used and their overheads) and volumes produced by such surveys over periods of many years requires careful preparation. Pending data rates from VISTA's 16 detector near-IR camera its closest analogue is the 4 detector Wide Field Camera (WFCAM) on the United Kingdom Infra Red Telescope (UKIRT). Large surveys, for example the UK Infrared Deep Sky Survey (UKIDSS, Lawrence et al 2006¹³) with WFCAM, and ESO's surveys to be made with VISTA, require the construction of complete, reliable, and documented products from the raw data. The very large volume of survey data means that to achieve these goals requires a uniform and automated approach to data processing. Likewise the large accumulated volume of products, many tens of Terabytes, means that as well as providing public data access, online querying and analysis facilities need to be provided as a service. The UK-side VDFS aims at (i) removing instrumental signature; (ii) extracting source catalogues on a frame by frame basis; (iii) constructing survey level products - stacked pixel mosaics and merged catalogues; (iv) providing users with both data access and methods for querying and analyzing the data.

Therefore from the outset it was planned that the system to process VISTA data would be reached through the intermediate step of a system to routinely process and archive WFCAM data. This approach gains us two years of experience of handling WFCAM data before VISTA data arrives, thereby helping to efficiently debug, derisk and improve future VISTA data processing, whilst simultaneously producing current science products from WFCAM. The system for WFCAM and VISTA is known as the VISTA Data Flow System (VDFS) and an overview of the VDFS was presented by Emerson et al (2004⁴), with details of the pipeline design by Irwin et al (2004¹¹) and of the science archive design by Hambly et al (2004a⁶). Of course VISTA and WFCAM data differ in more than just volume, but nevertheless the similarities are such that we believe our approach will greatly speed the time taken to get to a stable working system for VISTA. In this contribution we discuss the experience of the UK-side VDFS on WFCAM data. This experience feeds back into modules for VISTA data quality monitoring at the telescope, and into calibration of frames extracted from the VISTA raw data archive at ESO, as well as to the UK-side VDFS system for VISTA data. We also describe a new tool for ESO observers to use in preparing the survey observations with VISTA.

* j.p.emerson@qmul.ac.uk tel +44 207 882 5040

2. DATA FLOW

After conversion to FITS files at the UKIRT summit we make use of lossless Rice tile compression (e.g. Sabbey et al. 1998¹⁴) to reduce the data storage, I/O overheads and transport requirements. For this type of data (32 bit integer) the Rice compression algorithm typically gives an overall factor of 3–4 reduction in file size. Data are shipped roughly weekly from Hawaii using LTO tapes, one per detector channel, and combined to create the raw archived multi-extension FITS files on ingest in Cambridge, where they are available within a month of the observations being taken. The real-time status of the data transfer from JAC and of the pipeline processing in Cambridge is automatically monitored and made available online together with other general WFCAM information.

Raw data are then transferred via the Internet for ingest into the ESO archive system. All of the raw data is also available online through the WFCAM raw data archive centre in Cambridge. Access to the Cambridge archive is password protected. Registration is through the on-line interface. Note that although the raw data is stored using Rice Tile compression an uncompress on-the-fly option is available for retrieval - though we recommend transferring the compressed images due to the $\sim x$ 3 reduction in bandwidth required.

3. PIPELINE: INSTRUMENTAL EFFECTS REMOVAL

The general philosophy behind the pipeline is that all fundamental data products are FITS multi-extension files with headers describing the data taking protocols in sufficient detail to trigger the appropriate pipeline processing components, and that all derived information, quality control measures, photometric and astrometric calibration and processing details, are also incorporated within the FITS headers. Generated object catalogues are stored as multi-extension FITS binary tables. These FITS files thereby provide the basis for ingest into databases both for archiving and for real time monitoring of survey progress and hence survey planning.

The data processing strategy is to minimise the use of on-sky science data to form "calibration" images for removing the instrumental signature. By doing this we also minimise the creation of data-related artefacts introduced in the image processing phase. To achieve this we have departed somewhat from the usual NIR processing strategies by, in particular, making extensive use of twilight flats, rather than dark-sky flats (which potentially can be corrupted by thermal glow, fringing, large objects and so on) and by attempting to decouple, insofar as is possible, sky estimation/correction from the science images.

Each night of data is pipeline processed independently using the master calibration twilight flats (updated at least monthly) and a series of nightly generated dark frames covering the range of exposure times and readout modes used during that night. A running sky "average" in each passband is used for sky artifact correction. After removing the basic instrumental signature the pipeline then uses the header control keywords to produce interleaved and/or combined (stacked) image frames for further analysis. This includes generation of detected object catalogues, and astrometric and photometric calibration based on 2MASS. A more detailed description of the WFCAM processing is given in Irwin et al. (2006¹²).

The following processing steps are applied to all images:

3.1. Linearity correction

Measurements from dome flat sequences indicate the system is linear to <1% up to saturation regime. We therefore do not currently apply a linearity correction. Although it is feasible to measure and correct correlated double sample mode non-linearity, it would add ~ 30 s extra CPU overhead to the data processing per multi-extension FITS file.

3.2. Dark correction

Darks are routinely computed from the daily observations, by combining as many darks as are generally available for each exposure time and readout mode. If a particular combination is not available the nearest suitable calibration dark frame from nearby nights is used instead. If this still does not produce all the required darks to process a night's data, a suitable combination of closely related dark frames is created and used instead. The dark current is low and darks are stable over nights.

3.3. Twilight flatfield

Weekly dawn twilight flatfield sequences are normally taken (typically 9-point jitter sequences in Y,H one night and Z,J,K the next) and used to form master flats. The flats give good dark sky correction (gradients are at the ~1% level at most), and show no fringing or measurable thermal emission.

3.4. Decurtaining

There is a pseudo-periodic ripple on all frames at the +/-5 ADU level which we call a curtain. The curtain effect is removed by robustly estimating two 1D 1024 additive correction functions making heavy use of iterative clipped medians and the 4-fold 90 degree quadrant symmetry of the detectors. Note that decurtaining also corrects for the majority of any residual reset anomaly. The curtaining effect, and the efficacy of the correction for it is best illustrated from the difference between two dark frames of the same type. Figs 1 show such a difference image where the darks have not been corrected for curtaining, whereas Fig 2 shows the difference between the same two darks after they have each had the correction applied.

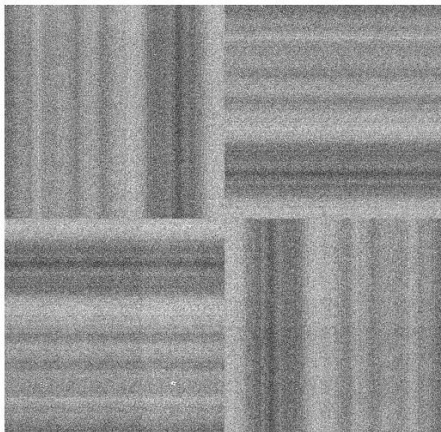


Figure 1 Difference of two darks showing the curtaining effect

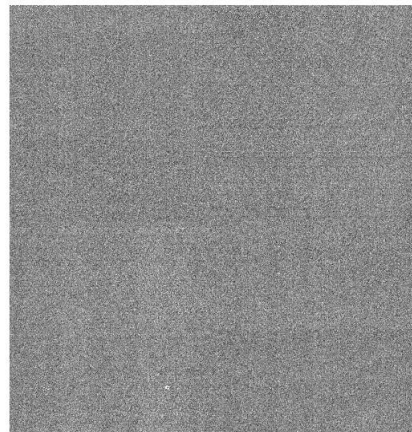


Figure 2 Difference between the two darks after they have each been 'de-curtained'

3.5. Sky Subtraction

After the decurtaining step images are temporally linked to produce master sky frames which are used in the sky (actually "crud" and illumination-dependent detector artefacts) correction step in an additive sense after rescaling to the frame sky level. Most of the sky correction is additive in nature and caused by illumination-dependent reset anomaly and pedestal offsets. Residual gradients introduced by the twilight flatfielding strategy are negligible across detectors and are well below the 1% level.

3.6. Cross-talk

All the detectors show similar cross-talk artefacts which are essentially time derivatives of saturated stars with either a doughnut appearance from heavily saturated regions (+/-128xm pixels either side, where m is the interleave factor m x m from microstepping e.g. 1x1 2x2 3x3, plus progressively weaker secondaries further multiples on), or half-moon-like plus/minus images from only weakly saturated stars. Adjacent cross-talk images have features at ~1% of the differential flux of the source, dropping to ~0.2% and ~0.05% further out. Beyond three channels the effect is negligible.. The good news is these artefacts are non-astronomical in appearance and do not "talk" across the detector quadrant boundaries, or between detectors; the bad news is that they do not generally jitter stack out. Significant cross-talk only appears to be generated by saturated images as is evident in the field of the open cluster shown in Fig 3 to illustrate the overall nature of the problem. Figs 4 illustrates the result of applying the cross talk removal algorithm to the same data as in Fig 3.

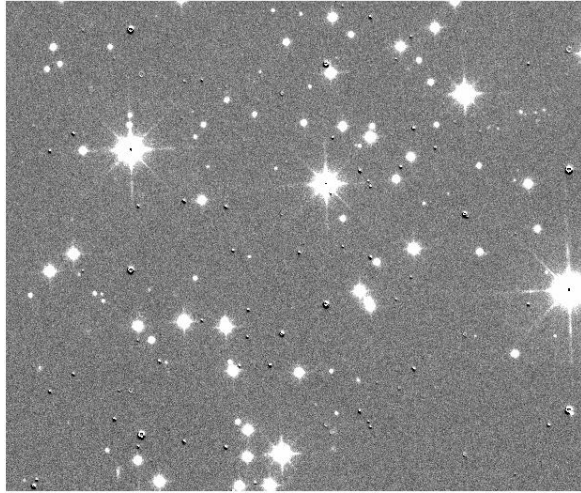


Figure 3 Open cluster showing that significant cross-talk only appears to be generated by saturated images. (see the cross talk 'bumps' symmetrically above and below the brightest stars).

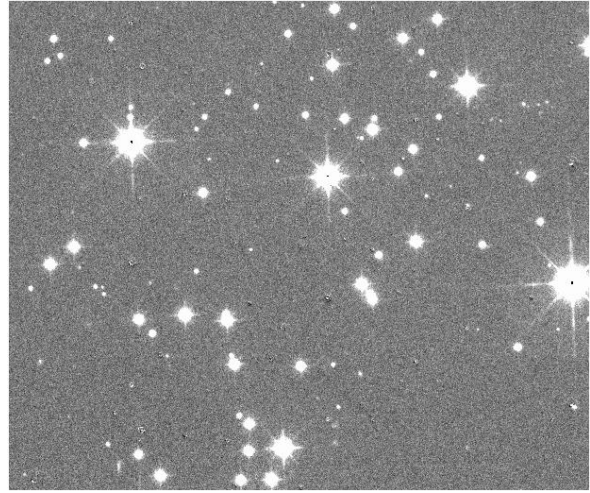


Figure 4 Result of applying cross-talk removal software to the previous Open Cluster image.

3.7. Persistence

Persistence images of a saturated star can appear over several subsequent frames and are not currently treated in the pipeline. The brightness of the persistence image is a function of source count rate, filter, and number of reset/reads. Persistence images are characteristically the size of the saturated part of their progenitor image. Images persisting across a number of frames will follow any offset pattern employed and in such cases the effect usually stacks out. However the artefacts look astronomical as can be seen in Fig 5. The images mimic fuzzy looking galaxies and in this image peak at typically 5-20 counts above sky. Note how long the persistence lasts - these were persistence artefacts from a previous part of a tile 9-point jitter sequence of the Open Cluster field in Figs 3 & 4.

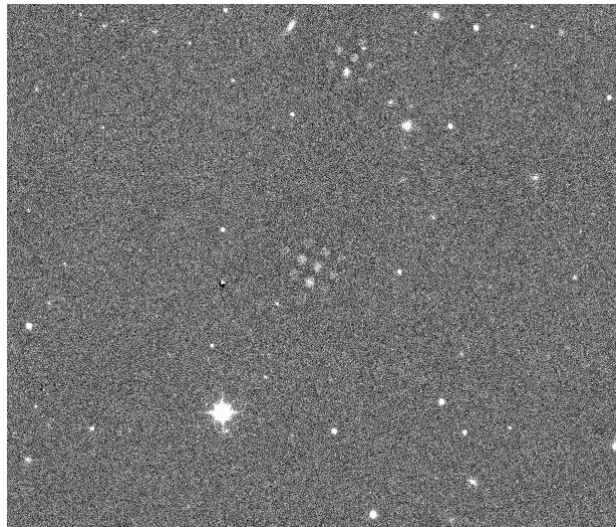


Figure 5 Image showing Persistence Artefacts

3.8. Cataloguing

After interleaving (if microstepped) or dither/jitter stacking (if dither/jittered) frames have full catalogue generation carried out and an accurate WCS computed and back-propagated through the component images.

3.9. Confidence Maps

All image products have an associated confidence map that tags the progress of bad pixels, sensitivity variations, stacking etc. through the various processing steps. The raw frames typically have between 0.1% and ~1% bad pixels, which generally are evenly distributed apart from one individual detector which has a lot around the edges.

3.10. Quality Control

Quality Control (QC) parameters are routinely computed for each generated catalogue: these include measures of seeing, ellipticity of stellar images, sky brightness and noise, mag ZP and extinction trends. A QC database for the pipeline products has been developed for monitoring data quality; following up reported problems with the processed data; providing feedback on status to the survey observation system; and as a prototype survey progress monitoring tool for the UKIDSS (http://apm15.ast.cam.ac.uk/docs/wfcam/science-verification/ukidss_progress.png/view)

4. CALIBRATION

Astrometric: The astrometry for all frames is based on 2MASS and globally is good to ~100mas and much better than that internally i.e. less than ~50mas.

Photometric: Photometry is currently based on 2MASS, via colour equations to convert to the WFCAM instrumental system. 2MASS solutions for every catalogued frame are generated and allow monitoring of effective Zero Points at the ~few % level. A series of tests are being carried out, using observations of the UKIRT Faint Standards (Hawarden et al 2001⁹) to monitor the WFCAM 2MASS calibration with very promising results. Analysis suggests that the 2MASS calibration is indeed delivering product frame-by-frame photometric zero-points (with factored-in extinction tracking) at the +/-2% level and looks highly likely to meet the WFCAM photometric calibration requirements. This is mainly due to the huge effort the 2MASS survey team made in ensuring a reliable all-sky calibration.

5. SCIENCE ARCHIVE

Data processing (in the VDFS pipeline in Cambridge) delivers standard nightly pipeline processed images and associated single passband catalogues, complete with astrometric and first-pass photometric calibrations and all associated 'meta' (descriptive) data in flat FITS files. After checking, the processed data products: images, confidence maps and catalogues, are automatically transferred to Edinburgh for ingest in the Science Archive using multi-threaded scp-like protocols with transfer speeds of around ~10Mbyte/s. To date some 25 Tbytes of raw WFCAM data are held in store (~8 Tbytes with Rice compression) with roughly 40 Tbytes of processed products created (~10 Tbytes with Rice compression). The processed data are then released to the public at periodic intervals.

To produce survey products however three more processes are needed - image stacking, source merging, and Quality Control (QC) filtering. Stacking and merging are the responsibility of the VDFS team and are described in Irwin et al (2006¹²) and Hambly et al (2006⁸). The QC process is a joint responsibility of the survey consortium and the VDFS project. The process for the Early Data Release of the UKIDSS surveys with WFCAM is described in Dye et al 2006¹. WFCAM image data volume is typically 200 GBytes per night, with catalogue and descriptive data being typically 10% of that figure. Hence, over the course of several years of observations it is anticipated that 10s/100s of Tbytes of catalogue/image data will be produced by survey operations with WFCAM alone. In order to enable science exploitation of these datasets, the concept of a 'science archive' has been developed as the final stage in the systems-engineered data flow system from instrument to end-user (Hambly et al. 2004a⁶).

The VDFS WFCAM Science Archive (WSA) is much more than a simple repository of the basic data products described previously. A commercial relational database management system (RDBMS) deployed on a mid-range, scalable hardware platform is used as the online storage into which all catalogue and metadata are ingested. This RDBMS acts as the backing store for a set of curation applications that produce enhanced database driven data products

(both image products, e.g. broad-band/narrow-band difference images; and catalogue products, e.g. merged multi-colour, multi-epoch source lists). Moreover, the same relational data model is exposed to the users through a set of web-interface applications that provide extremely flexible user access to the enhanced database driven data products via a Structured Query Language interface. The primary purpose of the WSA is to provide user access to UKIDSS datasets and a full description, along with typical usage examples, is given in Hambly et al. (2006⁸).

The design of the WSA is based, in part, on that of the science archive system for the SDSS (Thakar et al. 2003¹⁵). In particular, we have made extensive use of the relational design philosophy of the SDSS science archive, and have implemented some of the associated software modules (e.g. that for the computation and use of Hierarchical Triangular Mesh indexing of spherical coordinates – see Kunz et al. 2000¹⁰). Scalability of the design to terabyte data volumes was prototyped using our own existing legacy Schmidt survey dataset, the SuperCOSMOS Sky Survey (Hambly et al. 2001⁵). The resulting prototype science archive system, the SuperCOSMOS Science Archive (SSA) is described in Hambly et al. (2004b⁷) and provides an illustration of the contrast in end-user experience of an old-style survey interface (as described in Hambly et al. 2001⁵) and the new. Extensive technical design documentation for the WSA is maintained online at www.roe.ac.uk/~nch/wfcam.

The WSA receives processed data from the pipeline component of the overall data flow system in the form of FITS image and catalogue binary table files. No raw pixel data are held in the WSA. Processed data consists of instrumentally corrected WFCAM frames, associated descriptive and calibration data (including calibration images, e.g. darks and flats) and single-passband detection lists derived from the science frames. Calibration information also includes astrometric and photometric coefficients. Metadata are defined by a set of descriptor keywords agreed between the archive and pipeline centres, and include all information propagated from the instrument and observatory, along with additional keywords that described the processing applied to each image and catalogue in the pipeline. Single passband detections catalogues for each science image have a standard set of 80 photometric, astrometric and morphological attributes along with error estimates, and a variety of summary quality control measures (e.g. seeing, average point source ellipticity etc.). For more details, see Irwin et al. (2006¹²).

The design of the WSA was based, from the outset, on a classical client-server architecture employing a third-party back-end database management system (DBMS). This followed similar but earlier developments for the SDSS science archive, and reflects the great flexibility of such a system from the point of view of both applications development and end-user querying. The WSA has been based on a relational model from the start, bringing many advantages for astronomy applications (indeed, for any scientific discipline) where related sets of tabular information are familiar. Step-by-step examples of WSA usage are included in Dye et al (2006²) and the archive is at <http://surveys.roe.ac.uk/wsa/pre/>

6. SURVEY PREPARATION TOOLS

Depending on the observing strategy of a given survey there may be requirements to execute a series of observations of the same area at a given frequency (for variability), or to complete an area in several filters before moving to the next area (to get accurate colours at a single epoch), or to cover an area in one filter before moving to the next filter, or to ensure an area is covered within a given time interval etc etc etc. For a queue scheduled systems, such as VISTA, carrying out multiple surveys which may continue over many years it is necessary to provide the means to provide such information to the automated scheduling system, so it is in a position to best meet the scientific requirements. All these facilities could of course be useful in the course of shorter less extensive observations, but large scale surveys would be very hard to carry out efficiently without a means of dealing with these priorities. This is being addressed by changes to ESO's Phase 2 Preparation Program (P2PP) (Chavan et al 2005¹)

6.1. Time-linking of Observations

In addition to the case of absolute timing (which is simple to implement) two types of time-linkage are needed for surveys:

- 1) *Relative time links* where an observation must be executed within a time interval after the execution of a previous observation, but not necessarily at a fixed date. Examples of this are monitoring observations of a variable source at roughly constant intervals
- 2) *Concatenated observations* where two or more observations are executed consecutively. For example one may want to complete an area of sky at more or less the same epoch before moving to the next filter or area, or one

may want to complete a given area in several filters before moving on to the next area. Given that for scheduling purposes individual observations should generally not last more than about an hour.

The need for a more robust way of dealing with such dependencies is well recognized and a facility for the users to specify them will be available in an upcoming release of ESO's P2PP. This tool is provisionally called the Observation Block (OB) Linker and Execution TimER (OBLETER). OBLETER will provide users with an interface where OBs existing in the P2PP local cache can be loaded and organized in a chain. The user will then be able to define the earliest and latest time when a given OB in the series must be executed with respect to the preceding OB. A part of this interface will allow the user to deal with the special case of a zero time interval with respect to the preceding OB, corresponding to the case of concatenated OBs. Some basic chain manipulation capabilities are foreseen, such as the possibility of adding or removing OBs from chains, of continuing work on a chain partially defined in a previous P2PP session, or of modifying chains that had been already submitted to the ESO Database for execution. The time-related information will be stored in a database, from where it will be retrieved by scheduling tools available to the operator on the mountain in order to build up a short-term schedule that properly takes these constraints into account.

6.2. Definition of groups of OBs

It is possible at present for ESO to assign an execution priority to each OB, so that the operator is aware of the ones that have a higher scientific importance at the time of deciding on observations to execute for a given programme. It has been recognized nevertheless that such simple priority scheme is sometimes insufficient to deal with programmes containing large number of OBs, and especially for surveys containing large numbers of target fields observed in a number of instrumental setups. In such cases the need for a prioritization scheme above the individual OB level, which can take into account the past execution history of the programme, becomes clear. One can consider for instance the case of a survey of several target fields to be observed through several different filters, with each field and filter specified in a single OBs. Depending on the science goals of the programme it may be desirable to complete the observations of a given field in all filters before proceeding to the next field, or conversely to observe all the fields in a given filter before proceeding to the next filter, or even ensure that contiguous coverage among the fields takes priority.

The approach adopted to deal with such cases is the definition of *groups* of OBs, in which internal priorities within each group are reflected in the form of a contribution of each OB to the total group *score*. The short-term scheduling tools available on the mountain will take into account the current scores of each group of OBs, and will then apply a number of rules in order to prioritize the possible OBs to be executed according to them. Such rules will for instance give the highest execution priority to those OBs that set a new maximum of the score among the existing groups; and among those, the highest priority will be given in turn to those that produce the largest increase in group score. By assigning to the OBs the appropriate contributions to the scores of their respective groups, the users can make sure that the progress in the execution of the programme will take place in a way that is consistent with the scientific priorities of the observations. In addition, it will be possible to assign different priorities to each group.

A new facility deployable from P2PP will be available to users in order to properly define and manipulate the groups. Provisionally called the Tool for OB Organization in Groups and Networks (TOBOGAN), this facility will allow users to define groups, assign or de-assign OBs to them, define the contributions of OBs to the scores of their groups, and to assign relative priorities to the groups that have been defined. Like in the case of time constraints described above, TOBOGAN will also be used to manage the interaction between the P2PP local cache where OBs and groups are defined, and the ESO Database where they are checked-in and stored for execution once prepared. It will also give the user the possibility of continuing work on the definition of groups by retrieving information stored either in the P2PP local cache or in the ESO Database from a previous P2PP session. The group and priority information stored in the ESO Database will be used by short-term scheduling tools available on the mountain in order to determine the relative priorities among the OBs available for execution at any given time, thus providing a robust way for the operator at the telescope to select OBs to be executed in a manner consistent with the scientific preferences of the programme.

6.3. Import of target fields produced by the Survey Area Definition Tool

The Survey Area Definition Tool (SADT) is a utility currently undergoing development that will allow users to define survey areas to be covered by surveys executed with either VISTA or OmegaCam at the VST. The SADT will then determine the central coordinates of the different pointings required to cover the field according to the specifications, as

well as ancillary guide and wave front sensor star information to allow acquisition and guiding. The output produced is a file to be ingested from P2PP containing all the target information needed for the preparation of the OBs with which the survey will be executed. Together with OBLETER and TOBOGAN the basic tools to efficiently make survey observations and schedule them will be in place for the start of ESO's surveys with VISTA and VST..

7. CONCLUSIONS

Whilst we have concentrated here on data from WFCAM all the experience gained helps test the system for VISTA in preparation for the flood of VISA data expected in 2007.

ACKNOWLEDGMENTS

The VISTA Data Flow System is funded by grants from the UK Particle Physics and Astronomy Research Council.

REFERENCES

1. A.M. Chavan, F. Comeron, M. Peron, T. Canavan, D. Dorigo, and P. Nunes "From Handicraft to Industry: Supporting Surveys at ESO Telescopes", in *ADASS XV, ASP Conference Series*, eds. G. Gabriel, C. Arviset, D. Ponz, and E. Solano, 2005
2. S. Dye, S.J. Warren, N.C. Hambly, N.J.G. Cross, S.T. Hodgkin, M.J. Irwin, A. Lawrence, A.J. Adamson, O. Almaini, A.C. Edge, P. Hirst, R.F. Jameson, P.W. Lucas, C. van Breukelen, J. Bryant, M. Casali, R.S. Collins, G.B. Dalton, J.I. Davies, C.J. Davis, J.P. Emerson, D.W. Evans, S. Foucaud, E.A. Gonzalez-Solares, P.C. Hewett, T.R. Kendall, T.H. Kerr, S.K. Leggett, N. Lodieu, J. Loveday, J.R. Lewis, R.G. Mann, R.G. McMahon, D.J. Mortlock, Y. Nakajima, D.J. Pinfield, M.G. Rawlings, M.A. Read, M. Riello, K. Sekiguchi, A.J. Smith, E.T.W. Sutorius, W. Varricatt, N.A. Walton, S.J. Weatherley, "The UKIRT Infrared Deep Sky Survey Early Data Release", *astro-ph/0603608*
3. J.P. Emerson, W.J. Sutherland, A.M. McPherson, S.C. Craig, G.B. Dalton, A.K. Ward, "VISTA: The Visible & Infrared Survey Telescope for Astronomy" *ESO The Messenger* **117**, p27-32, 2004
4. J. Emerson, M. Irwin, J. Lewis, S. Hodgkin, D. Evans, P. Bunclark, R. McMahon, N. Hambly, R. Mann, I. Bond, E. Sutorius, M. Read, P. Williams, A. Lawrence, M. Stewart, "VISTA Data Flow System: Overview", in *Optimizing Scientific Return from Astronomy through Information Technologies*, P.J. Quinn & A. Bridger eds., *Proc SPIE* **5493**, 401-411, 2004.
5. N.C. Hambly et al *MNRAS*, **326**, 1279, 2001
6. N. C. Hambly, B. Mann, I. Bond, E. Sutorius, M. Read, P. Williams, A. Lawrence, J. Emerson., "VISTA Data Flow System: survey access and curation: The WFCAM Science Archive", in *Optimizing Scientific Return from Astronomy through Information Technologies*, P.J. Quinn & A. Bridger eds., *Proc SPIE* **5493**, 423-431, 2004a.
7. N. C. Hambly, et al., In: *Proceedings of the 13th Astronomical Data Analysis and Software Systems (ADASS)*, eds. F. Oschenbein, M.G. Allen & D. Egret, *ASP. Conf. Ser.*, **314**, 137, 2004b
8. N.C. Hambly et al 2006 in preparation
9. T. G. Hawarden, S.K. Leggett, M.B. Letawsky, D.R. Ballantyne, M.M. Casali, *MNRAS*, **325**, 563, 2001
10. Kunszt, P.Z., Szalay, A.S., Csabai, I., Thakar, A.R., In: *Proceedings of the 9th meeting on Astronomical Data Analysis and Software Systems (ADASS)*, eds. N. Manset, C. Veillet, & D. Crabtree, *ASP Conf. Ser.*, 216, 141, 2000
11. M. J. Irwin, P. Bunclark, D. Evans, S. Hodgkin, J. Lewis, R. McMahon, J. Emerson, S. Beard, M. Stewart, "VISTA Data Flow System: pipeline processing for WFCAM and VISTA", in *Optimizing Scientific Return from Astronomy through Information Technologies*, P.J. Quinn & A. Bridger eds., *Proc SPIE* **5493**, 411-422, 2004.
12. M.J. Irwin et al 2006 in preparation
13. A. Lawrence, S.J. Warren, O. Almaini, A.C. Edge, N.C. Hambly, R.F. Jameson, P. Lucas, M. Casali, A. Adamson, S. Dye, J.P. Emerson, S. Foucaud, P. Hewett, P. Hirst, S.T. Hodgkin, M.J. Irwin, N. Lodieu, R.G. McMahon, C. Simpson, I. Smail, D. Mortlock, M. Folger, "The UKIRT Infrared Deep Sky Survey (UKIDSS)", *astro-ph/0604426*
14. C. Sabbey, P. Coppi, and A. Oemler. *PASP* **112**, 867, 1998
15. A.R. Thakar, A.S. Szalay, J.V. Vandenberg, J. Gray, In: *Proceedings of the 12th meeting on Astronomical Data Analysis and Software Systems (ADASS)*, eds. H.E. Payne, R.I. Jedrzejewski & R.N. Hook, *ASP Conf. Ser.*, **295**, 217, 2003
16. S. Warren, "Scientific goals of the UKIRT Infrared Deep Sky Survey", in *Survey and Other Telescope Technologies and Discoveries*, J.A. Tyson & S. Wolff eds., *Proc SPIE*, **4836**, pp. 313-320, 2002