VISTA DATA FLOW SYSTEM

Document Title: **Report to Grid Steering Committee Feb 03**

Document Number: **VDF-MEM-GSC-00004-0001 = GSC(03)05**

Issue: **Final**

Date: **07 February 2003**

Document Prepared By:	**Jim Emerson (VDFS)**	**Signature and Date:**	*James P Emerson* *05/02/03*

Table of Contents

1 Introduction

The Data Flow system in the UK consists of a 'pipeline' run at the Cambridge Astronomical survey Unit (CASU) which archives the raw data, removes instrumental effects, extracts catalogs of objects and transfers the calibrated data and catalogues to the (science) 'archive' at the Wide Field Astronomy Unit (WFAU in Edinburgh).
As the CASU pipeline is essentially a pipeline driven by each nights data it can apply processing that requires memory but not that which requires foresight. Thus, for example, it can coadd all observations taken on one bit of sky during one night, but it cannot coadd all observations on one bit of sky taken over many nights – the latter would be done at the science archive. The same applies to detecting variables. The science requirements on the pipeline are at www.jach.hawaii.edu/~adamson/wfplrq.html

The science archive at WFAU provides the portal through which users access, search and manipulate the products. The archive must also provide many functions to enable the data taken on many different nights, over many years, to be combined, differenced, cross calibrated, compared with images at other wavelengths, visualised etc etc. The archive will include routinely produced products, and capabilities for user to invoke special processing on their specified selections of data. The science requirements on the archive are at www.jach.hawaii.edu/~adamson/wfarcrq.html.

The purpose of this paper is both to report on progress and to seek confirmation of funding for the full project to Dec 2004, the deliverables for which are now clearly specified, following extensive discussions amongst ourselves and most importantly with the science user community. There are also requests for modest additional funding to deal with gaps identified during our studies.

Sec 2 reports on progress since the last GSC meeting, by work package, Sec 3 notes areas of concern, Sec 4 lists milestones, Sec 5 addresses the concerns expressed by GSC on the vagueness of the deliverables stated in the original proposal and revises these and the effort and costs, including hardware in 2003-4.

Finally Sec 6 summarises the requirements and list the actions requested of GSC.

2 Progress since Last report

2.1 General

In the overall VISTA Data Flow System (VDFS) project, the WFCAM data flow system provides the project `test-bed' or `phase A'. This is a natural result of WFCAM, which will see first light three years earlier than VISTA, being a similar camera to the one that is being built for VISTA (which has 4 x as many detectors), and hence all the associated processing, archiving, curation and end-user requirements are similar. Being smaller in scope, the WFCAM project is a natural stepping stone to a VO-compatible VISTA Science Archive. Therefore work has concentrated on the equivalent modules for the WFCAM system as the VISTA test-bed.

Discussions with users in the form of the UKIDSS Consortium have clarified the science deliverables required from both the pipeline (Cambridge) and the archive (Edinburgh).

Analysis of the science requirements, and required dates for WFCAM (and later VISTA) functionality, and resources available have led us to adopt a phased approach so that a pipeline and archive with `standard' functionality, v1, is available at first light, enabling immediate science exploitation, and subsequently a fully functioning archive system is made available one year after WFCAM survey operations begin, v2.

This approach is both desirable, to allow an evolutionary cycle of coding, testing and enhancement, and pragmatic given that WFCAM data starts to arrive in only 8 months time in September 2003.

The WFCAM software Critical Design Review was held in Oct 2002. The response is reported below by work package.

To ensure quality control of products it has been decided that:

- User access to VDFS material at CASU will be restricted to raw pixel data only.
- User access to all other VDFS products will be only through WFAU.

The period Oct-Dec coincided with the Project Leader's teaching responsibilities. He is now free from teaching load, and is able to visit Cambridge and Edinburgh more often.

At the request of the Director of e-science PPARC's VISTA Board (Chair R Wade) agreed to act as the oversight committee for the VISTA Data-Flow System work on behalf of the Grid Steering Committee. The Board meets monthly and receives a report on VDFS progress.

The remainder of the progress report is divided up by the work packages used in the original VDFS bid, although it should be realised that some work is applicable to more than one WP.

2.2 WP-1 Survey Planning & Progress Tools

This will be pursued with ESO in Spring 2003 when the necessary manpower is identified.

2.3 WP-2 (Quality Control and Calibration Pipeline Module - for ESO)

This work is the responsibility of staff at the Cambridge Astronomical Survey Unit (CASU).

WP-2 is eventually deliverable to ESO and is therefore a separate WP from things that are deliverable to the UK, even though most of the material delivered to ESO will be reused in WP3 & some in WP-4. The Design Document is at `ftp://www.ast.cam.ac.uk/~mike/wfcampipedoc_v2.ps.gz`

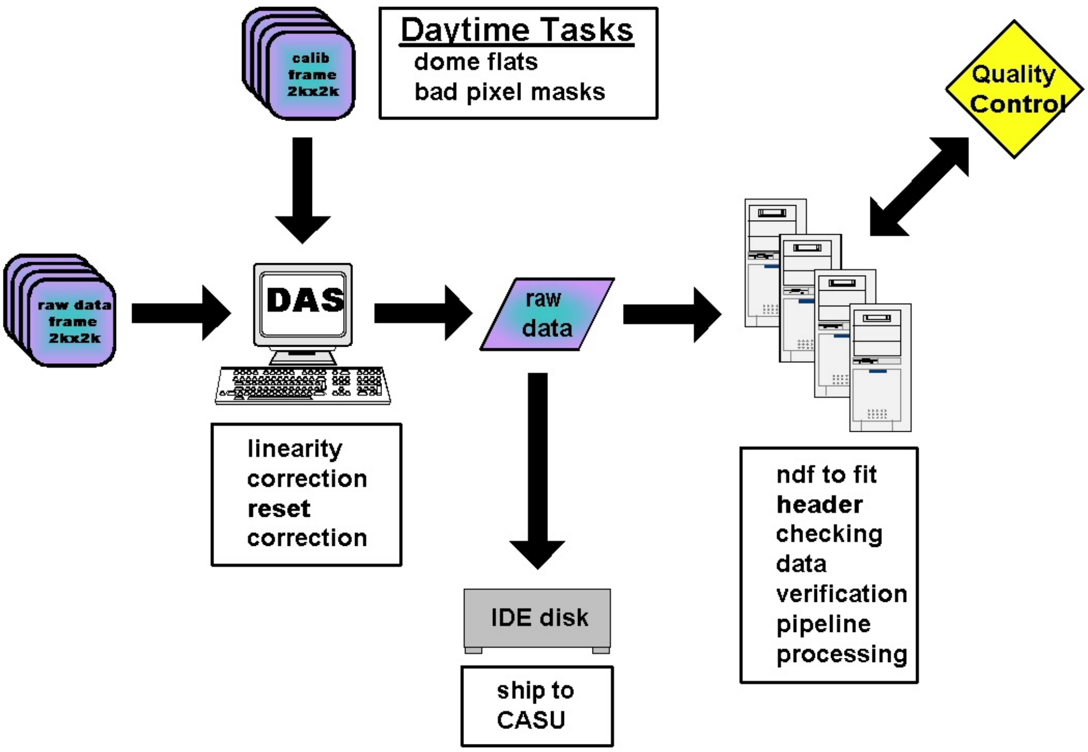

Figure 1: An overview of the processing at the telescope.

2.3.1 Data Quality Monitoring

Discussions and visits to the Joint Astronomy Centre in Hawaii have led to clear definition of the data quality control (DQC) measures desirable and to a scheme for producing such measures.

The following DQC measures greatly enhance the usefulness of data products for the end user:

- Night-time observing conditions/weather
- Instrument status values: detectors/filters/controllers

Individual frames will have the above information written into their headers; the quick-look telescope pipeline will then derive further DQC measures and insert them into headers:

- sky brightness - saturated? Too bright? No signal?
- sky noise - detectors OK? Pattern noise?
- average stellar ellipticity - trailed images?
- average seeing measure - for each detector, in focus?
- stellar aperture correction - weird PSF?
- no. of spurious images - corrupted data?
- astrometric errors - wrong RA, Dec?
- first pass photometric calibration - system throughput?

All processing done by the UK-side pipeline will record progress in image FITS headers, possibly derive further DQC measures, and version and time stamp the processing stages.

2.3.2 Test (quick look) Calibration Pipeline

A pipeline to run on UFTI data at UKIRT, using their ORAC-DR environment, has been constructed to test the integrability of the CASU code with a ‘foreign’ environment. This has gone well.

2.3.3 Review

The data quality monitoring and calibration pipeline plans were well received at the WFCAM software CDR in October 2002.

2.3.4 Documentation for ESO

Production of documents for ESO review has not yet started.
This is due to lack of manpower, because of non-receipt at Cambridge of the grant from PPARC. This problem has been highlighted many times, with little result from PPARC until the last few weeks. (and is discussed under 3.1)

A further consequence is that it has not been possible to maintain the schedule for reviews of this package. It had been intended to eventually bring them into phase with the overall VISTA software reviews, but the delay has pushed back the time when this can occur. Although it is some sort of consolation that the code itself is progressing well it remains a concern that the process of ensuring ESO acceptance of the software has not yet been able to begin. Once the new individual is in post we will try to force the pace of work.

2.4 WP-3 (Calibration Update Pipeline - in UK)

This work is the responsibility of staff at the Cambridge Astronomical Survey Unit (CASU). The input is raw data frames and calibration data frames. The output is calibrated data frames. Both the input and output frames are stored off line, with a database holding the description of each frame and its processing history.

The calibration update work builds on the (ESO deliverable) products in WP-2 but allows the products to be controlled by the UK, and tailored to the interests of the UK community. VISTA raw data will inevitably need to be reprocessed in response to updated understanding of instrument, sky and products, or changing science needs, via this Pipeline.

The following standards have been adopted.

- Pipelines delivered in the form of C library modules under perl/XS using FITS as the primitive transport.
- Code stored in CVS repository.
- Use of the revised FITS standards
- Use of CFITSIO and WCS libraries

A report on first tests at ATC with one of the science grade WFCAM (Rockwell) detectors has been provided to CASU to give an idea of actual detector performance. Persistence does not appear to be a problem.

A prototype pipeline has been built and the major steps successfully run on CIRSI data (CIRSI is a multi-detector IR imager built by Cambridge).

As mentioned in 2.3.2 the Pipeline Design presented at the October WFCAM Software CDR was approved by the Review Panel.

2.5 WP-4 (Single Observation Survey Products)

This work is the responsibility of staff at the Cambridge Astronomical Survey Unit (CASU). The input is calibrated data frames. The output is source catalogues, mosaiced frames, and stacked frames for the night.

The Pipeline science requirements include definition of single observation survey products.

Much of the work done here was on defining the deliverables and the Design of the system. This is covered in section 5.2.

The Design Document is at
`ftp://www.ast.cam.ac.uk/~mike/wfcampipedoc_v2.ps.gz`

The plans for implementing this were well received at the WFCAM Software CDR in October 2002.

2.6 WP-5 (Advanced Survey Product Development)

This work is the responsibility of staff at the Cambridge Astronomical Survey Unit (CASU). The scheme is that the necessary algorithms are first developed and tested at Cambridge then, when working, packaged as grid services and tested in that form. Ultimately they are made available at Edinburgh as grid services.

The lack of grant to Cambridge and the consequent lack of the expected new staff members (see 3.1) has so far restricted this work mainly to interacting with the user community to define and prioritize the scientific requirements, and to planning the tools required. This necessary work is reflected in the plans in 5.3. However four CASU team members, together with Guy Rixon of Astrogrid, have produced a prototype browser and portal for data grids (see www.ast.cam.ac.uk/~agrid/publications/adass-gtr-20010930/) which has given us experience in the work required.

We recognise that the combination of urgency of preparation for the arrival of WFCAM data, and the effort limitation, has resulted in less grid specific work being undertaken than had been hoped. However we believe that this was the correct balance, given the constraints, the need to remain science driven, and the work that was going on elsewhere to define VO standards.

2.7 WP-6 (Science Archive: Survey Products Access & Curation)

This work is the responsibility of staff at the Wield Field Astronomy Unit (WFAU) in Edinburgh. The input is the output of WP-4 (source catalogues, mosaiced frames, and stacked frames). The outputs are catalogue and pixel data, selected by user-specified criteria, and/or products derived by the users from these, either based on IR camera data alone, or using that data in conjunction with catalogues or pixels from other sky surveys. No raw pixel data is held in the Science Archive (raw pixel data is at CASU).

Proposals for the VDFS test bed products to be available from the archive respectively were discussed by the users in the form of the UKIDSS Consortium. The resulting changes have been documented and propagated into the plans.

WFAU have focused on detailing the functional requirements and design choices for the archive and `20 Usages' have been developed to illustrate likely functional requirements (www.roe.ac.uk/~nch/wfcam/misc/wsausage.html).

2.7.1 (First) CDR

Although material was presented at the WFCAM software CDR by the Science Archive it was not in a form that was appropriate at CDR level, and so a mini-CDR of the science archive only is being scheduled for 15-16 April 2003. Whilst the need to revisit a CDR was worrying (and correct) it is felt that the fundamental design is in a fair state. The problem at the CDR was more due to a misunderstanding of the detail that needed to be presented than that the work itself had not actually been done. A greater level of specification of what must be presented at reviews will be provided to WFAU presenters in future, and dry runs of content presentations will be held at an early stage to avoid recurrence of such misunderstandings.

2.7.2 Requirements

In consultation with the VDFS Project Leader Jim Emerson, UKIRT Director Andy Adamson and the WFCAM user community (i.e. the UKIDSS Consortium) via the Consortium Survey Scientist Steve Warren, the archive functional requirements (available at www.roe.ac.uk/~nch/wfcam/srd/wsasrd/wsasrd.html) were finalized. These functional requirements are being further developed through the design process; most relevant here is the Data Flow Document (DFD; available in draft form at www.roe.ac.uk/~nch/wfcam/dfd/wsadfd/wsadfd.html).

2.7.3 DBMS Investigations

A mirror for the SDSS has been established (see www-wfau.roe.ac.uk/sdss). This has allowed us to gain experience in the issues surrounding setting up such a database of a few 100 GB of catalogues, images and spectra using a commercial DBMS (in this case Objectivity) running on Linux PCs connected to SCSI disks. Developments using Objectivity have now halted, and we are now following the SDSS Archive team and using SQLServer to set up high-performance, flexible archives at WFAU (see below); we are further investigating the other two large data warehouse DBMS products, namely IBM's DB2 and Oracle, and are keeping our options open as to the final DBMS choice (see below).

2.7.4 Trial Datasets

Our priority is now working with existing datasets towards the functional requirements of the SA (i.e. rapid prototyping and proof-of-concepts to develop the necessary experience, at the same time supplying useful archive services to the user community with current datasets). We have used SQLServer to provide a flexible archive for the 6dF database (see http://www-wfau.roe.ac.uk/6dFGS).
We have designed a schema for an SQLServer implementation of our own SuperCOSMOS Sky Survey (SSS; see http://www-wfau.roe.ac.uk/sss). We have followed the `20 queries' approach in specifying the requirements for this implementation; the query set includes joint queries with the SDSS as well as several SA-like usages. We have successfully ingested and exercised an SQLServer database of the SSS in the SDSS-EDR regions (consisting of 4×10^7 records, or 1% of the final SSS size).

2.7.5 Need for Experimentation

WFAU's general DBMS work has included working closely with Jim Gray of MicroSoft research on our SSS/SQLServer implementation. We are maintaining our links with Alex Szalay and the Johns Hopkins University SDSS Science Archive team with a view to mirroring the next release of the SDSS (using the new `SkyServer' system which is implemented using SQLServer). We have also established contacts with Ian Carney (Oracle) and Andy Knox (IBM/DB2) for technical advice concerning their respective DB Management Systems. Three members of the SA team attended a course on relational database design, supported by Oracle and hosted by NeSC in December 2001. A key point to note about this is that the advice from the DBMS experts is that *experimentation is vitally important* when prototyping a large database.

2.7.6 Hardware Configurations

Given the complexity of hardware configuration, we have begun to investigate different hardware solutions using loaned hardware prior to purchase. For example, we have compared response times for our standard 20 queries between our own PC system running Windows/SQLServer (Pentium II processors and SCSI disks) and a loaned system comprising a Xeon quad processor, also running W2K/SQLServer and attached to SCSI disks. We obtained identical performance on each system (limited by the slowest link in the I/O chain - in this case the disk controllers) in terms of the table scan rate of 10^7 SSS records per minute. Currently we are investigating fibre-coupled SCSI and IDE RAID systems with the same dataset and queries to increase this performance rate by more than a factor of 10.

3 Areas of Concern

3.1 No issue of grant

The uncertainties resulting from the continued failure to make grant announcements to CASU & WFAU have continued to be disruptive, and in the case of CASU continued to delay getting the team up to complement.

Since the last report various timescales for announcement to CASU and WFAU of the results of their grant application have systematically moved later and later. PPARC should have been in no doubt at the problems that this caused which were vigorously reported to them on several occasions. In early January a grant by the end of January was promised. An action was placed at the January VISTA Board meeting to see what could be done to expedite the grants. The latest from PPARC is that the grants will be out by April 2003. The situation remains unsatisfactory.

As far as the VDFS is concerned it would have been much better to have issued grants in Apr 02, rather than wrap them up with other work by CASU and WFAU, thereby causing avoidable delay.

At CASU the effect is that 1 FTE of effort that should have been available since Oct 02 is still not available. To some extent this has been compensated by some of the (nominally 30%) research time of some CASU staff being expended on matters that were very urgent, but for the most part it has caused a month on month delay to the work of documenting the deliverables for ESO, and has left insufficient time to work on the more advanced processing options.

At WFAU, the effects have been a delay in the transfer of a University-funded archive/datamining post to the rolling grant and delay in the purchase of hardware urgently needed at this stage (see the hardware case). To minimise delay, the University funded individual is now (January 2003) working on WP6. Because the lead time for hardware purchase is shorter than that for staff recruitment, the delay implied by the April 2003 start date is not problematical if PPARC allows the WFAU to initiate purchase of the first tranche of equipment approved by the GSC, and charge the cost to their existing rolling grant and vire accordingly.

3.2 Adequacy of manpower & timing of grid enablement

Analysis of the task ahead has shown it is bigger than at first realised, the urgent scientific priorities and the limited manpower have made it difficult to justify diversion of much effort on implementing grid specific items before we have a system ready to handle WFCAM data. The urgency of preparing for WFCAM data makes it difficult to justify a different approach. As confidence builds in the software adopted, and progress in grid standards for astronomy stabilise we will begin to build some test grid applications. The very large astronomical database that VDFS is going to produce will be used as essential feedstock for science and is already attracting a lot of interest from database people. We must be sure that the product inside is one to be confident in, before going full throttle on providing capabilities to access and use it with escience tools. We request a modest increase in manpower to help this situation.

4 Milestones

The following top-level milestones have been identified so far.

Milestone	Scheduled Date	Actual date
WFCAM Software CDR	Oct 2002	21/22 Oct 2002
	2003	
WFCAM Science Archive CDR	15-16 Apr 2003	
VISTA Pipeline PDR	Jun 2003	
VISTA Science Archive PDR	Jun 2003	
WFCAM Pipeline v 1 complete	Sep 2003	
WFCAM Science Archive v1 complete	Sep 2003	
VISTA Pipeline CDR	Dec 2003	
VISTA Science Archive CDR	Dec 2003	
	2004	
VISTA Progress review	Jun 2004	
Phase B starts	Oct 2004	
WFCAM pipeline v 2 complete	Dec 2004	
WFCAM Science Archive v2 complete	Dec 2004	
	2005	
VISTA focal plane Test images available	Jan 2005	
VISTA Instrument test (Europe) starts	Sep 2005	
VISTA Progress Review	Jun 2005	
WFCAM Pipeline v3 complete	Dec 2005	
WFCAM Science Archive v3 complete	Dec 2005	
	2006	
VISTA European Instrument Integration	Feb 2006	
VISTA Instrument Commissioning Pt 1 starts	Mar 2006	
VISTA Instrument Commissioning Pt 2 starts	May 2006	
VISTA Final Instrument Delivery review	June 2006	
ESO Commissioning & Acceptance complete	July 2006	

VISTA scientifically operational.	Aug 2006	
VISTA data arrives	Sep 2006	
VISTA Debug phase with real data complete	Dec 2006	
VDFS Phase B ends		
	2007	
VISTA Operational Phase begins	Jan 2007	

5 Deliverables & Decisions Requested

5.1 Introduction

At its April 2002 meeting GSC made the following observations when approving full funding for the other parts of the VDFS proposal.

- Staff funding was approved to October 2002 for WPs 4, 5 & 6 for scoping, pending a more detailed proposal. ("The objectives of WP5 (advanced survey products) were ill-defined because there was a range of possible deliverables in this area but precisely defined deliverables would not be possible until more scoping had been done. Overall WP5 was potentially the most interesting part of the whole proposal but, together with WPs 4 and 6, needed more detailed planning and definition of deliverables before the committee could consider full approval.")
- Hardware funding was approved as requested except the 2003/04 hardware (£175k) which was withheld pending the more detailed planning requested above. ("There was currently no indication of how any required database licenses would be acquired. This was a clear risk. There is no specific explanation given for the requirement for a 50 node Beowulf cluster.")

This Section aims to address these concerns and is grouped by work done mainly at CASU and work done mainly at WFAU.

In section 6.1.2 GSC are requested to make the following decisions, the background to each of which is detailed in the case that follows.

The updated plans and definition of deliverables for WP-4 (with 3) & 5 are addressed in Sections 5.2 & 5.3.

- ➢ GSC is requested to approve continuation of funding for WP4 (Pipeline) work, and to agree to fund WP5 (Advanced Survey Product Development) work.

The total effort needed for WP-4 & 5, with WP-2 & 3 also included to give the overall picture at CASU, is addressed in 5.4.

- ➢ GSC is requested to agree the 5.5 dsy level of manpower (and consequent funding) for WP4&5 work by CASU, which includes an additional (relative to Apr 02 VDFS bid) 1.5 FTE from March 2003.

The CASU hardware requirement in 2003-4 is explained in 5.5.

- GSC is requested to agree hardware funding of £96k for of 2003/4 for work by CASU.

The updated plans and definition of deliverables for WP-6 is addressed in Section 5.6.

- GSC is requested to approve to continuation of funding for WP6 (Archive) work.

The total effort needed for WP-6, WFAU, is addressed in 5.7.

- GSC is requested to agree the 4.75 dsy level of manpower (and consequent funding) for WP6 work by WFAU, which includes an additional (relative to Apr 02 VDFS bid) 0.75 dsy from March 2003.

The WFAU hardware requirement is addressed in 2003-4 in Section 5.8.

- GSC is requested to agree hardware funding of £65k for of 2003/4 for work by WFAU, and a further £48k for 2002/3.

Finally the date when v2 is to be delivered is 3 months after the Sep04 date when the Apr02 allocation nominally ceases.

- GSC is requested to approve to an additional 3 months of effort funding for the VDFS to take it to the Dec04 delivery of the v2 system. This should be at a rate that provides a further 0.25 x (Oct03-Sep04 effort) which gives £167.4k.

VDFS Milestones were listed in Sec 4.

5.2 WP4 Deliverables (Single Observation Survey Products)

The input to WP4 is the output of WP3. The output of WP4 is input to WP6.

To appreciate the WP4 deliverables they must be seen in the context of the preceding step of the calibration pipeline (WP3), which is first described. Some issues still requiring attention are also highlighted.

The approach adopted is throughout is for a test bed v1 delivery by Sep 2003 (to enable the system to be tested with real WFCAM data) followed by a test bed v2 delivery by Dec 2004.

5.2.1 WP-3 (Calibration Update Pipeline)

- The calibration pipeline in Cambridge will be a superset of the UKIRT/ESO pipelines such that when the same processing sequence is used identical output products will result and will operate on the observations taken during each night to produce the final individual instrument signature-free images and to derive object catalogues from them. The object catalogues will be used to provide the first pass astrometric and photometric calibration which together with the images forms the standard science survey products. (Parameters to include: image location, various flux and shape measures, and error estimates for them. The parameters provide the basic input to the object morphological classification scheme.)

- Image data products at this stage will be instrumental signature-corrected and sky variation-corrected: single frames; lossless interleaved superframes;

- The raw science data will be held in an online (spinning disk) database in Cambridge, both for backup and reprocessing purposes and to provide accredited access to the raw data for external users should they so wish.

In more detail:

In what follows we deal with each step in the approximate location they occur in the pipeline

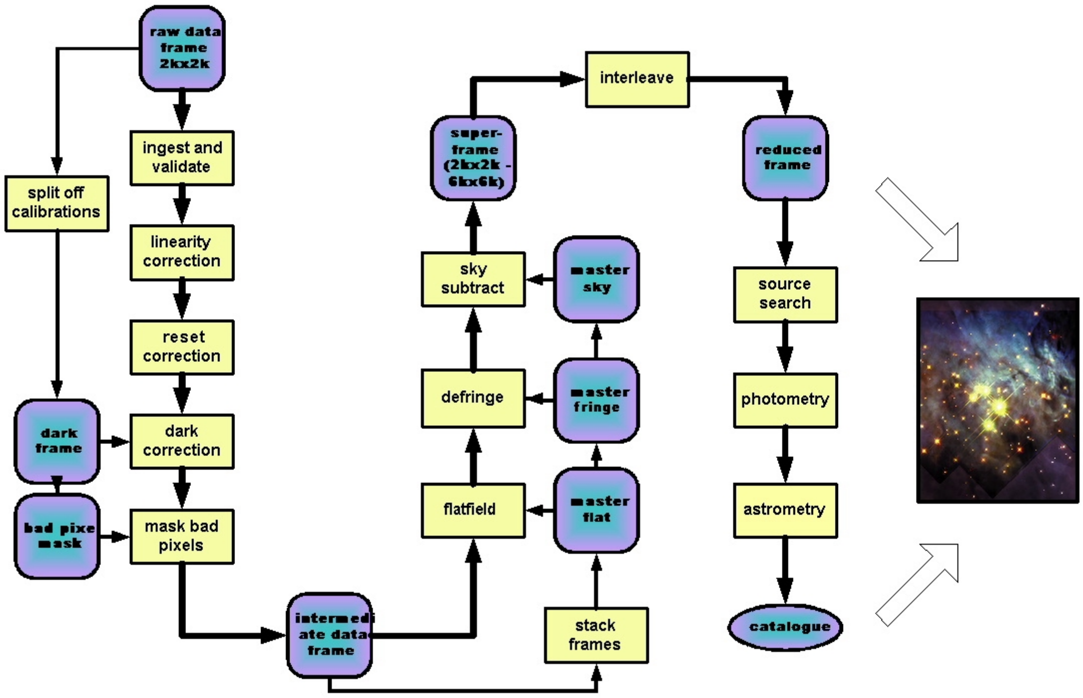

Figure 1 Overview of Pipeline Operations

Dark correction: as mentioned before this is not a completely resolved issue. A mean dark frame should come from several dark observations either from within the present or from earlier observations, suitably combined with rejection to get rid of any cosmic ray hits. In theory, dark correction involves scaling the mean dark frame to the exposure time of the program frame and then subtracting from the latter. In practice if other artefacts are present, dark frames of the same exposure time as the individual science exposures will be necessary.

Flatfield correction: Because WFCAM & VISTA are multi-detector cameras, internal calibration will also involve accounting for the variation in mean gain from chip to chip. If the sky "on average" uniformly illuminates each detector then the variation in the "mean" counts on the flatfield map for each chip is a measure of the variation of the mean gain. Flatfielding is usually defined such that the "mean" counts in the object frame remain constant. For situations where there is a gain difference between detectors, the mean flat for each detector will in effect be normalised by the ensemble average counts over all detectors thereby ensuring correct inter-detector gain normalisation.

To be more precise, as a function of position x,y and wavelength λ:

$$D_n(x,y) = \int_\lambda G_n(\lambda) g_n(x,y,\lambda) f(\lambda)\, S(x,y,\lambda)\, d\lambda \qquad (1)$$

where $D_n(x,y)$ is the data as observed on detector n, G_n is the gain correction for chip n, g_n is the relative gain map for chip n (the flatfield), f is the remaining system response curve and S is the source function. The obvious caveat with defining differential gain corrections this way is that they are potentially a function of source colour due to inherent variation of the QE curves of the detectors (indeed this can even be the case at the individual pixel level).

Form initial confidence map: the initial confidence map for each exposure is formed from the mean current flatfield, and will therefore be the same for all images in a given flatfield sequence. Pixels outside a specified tolerance range are given a confidence of zero. The confidence for unrejected pixels is defined on a percentile scale such that the median confidence of unrejected pixels is 100%. Confidence maps are carried through further processing stages in conjunction with the "average" data frame background level noise variance.

Defringe: if the fringe spatial pattern is stable and if flatfields can be generated without fringing present, it is possible to decouple sky correction and fringe correction and apply a defringing method similar to the one we have developed for optical imaging.

Sky subtraction: alternatives. Standard NIR processing often subtracts sky first and then flatfields. We can see why this can be advantageous compared with dark-correcting, flatfielding and sky-correcting by considering the following encapsulation of the problem,

$$D(x,y) = ff(x,y)\ [S(x,y) + F(x,y) + O(x,y) + T(x,y)] + dk(x,y) \qquad (2)$$

where D(x,y) is observed, ff(x,y) the flatfield function, S(x,y) is the sky illumination, F(x,y) is the fringe contribution, O(x,y) is the object contribution, T(x,y) is the thermal contribution, dk(x,y) is the dark current and without loss of generality we have dropped any explicit wavelength- and timedependence for clarity. Stacking with rejection produces an estimate of the terms

$$\hat{I}(x,y) = ff(x,y)\ [S(x,y) + F(x,y) + T(x,y)] + dk(x,y) \tag{3}$$

therefore,

$$D(x,y) - \hat{I}(x,y) = ff(x,y)\ O(x,y) \tag{4}$$

obviating the need for dark-correcting and fringe removal as both separate data gathering requirements and as separate data processing steps; and minimising the effect of systematic and random errors in the flatfield function by removing the largest potential error terms.

This of course glosses over the question of the spatial and temporal variability of the sky background. If the sky characteristics change significantly over an OBSERVATION then this method will still leave sky residual patterns exacerbating the problems, for example, of forming useful mosaic tiles.[5]

Reset anomaly: Many NIR detector systems have noticeable residual structure left in the background after the reset correction. For multi-sector reads, reset anomalies could leave a challenging background variation over the detector to deal with. This is analogous to the problem of dealing with rapid variations in sky background level/structure during stacking/mosaicing. We have had to develop ad hoc techniques to deal this problem for CIRSI data.

Image retention: ghosts of images/artefacts from preceding frames may be present. Strategies for dealing with this involve assessing the time decay characteristics and adjacency effects (*i.e.* ghost image spreading). Correcting for image persistence will either involve updating and maintaining a persistence mask, or accumulating a persistence map, running over observations if necessary, to subtract from the current image.

Interleave: the on-sky detector pixel scale for WFCAM will undersample typical seeing conditions on Mauna Kea. To recover some of this lost resolution, an observation at a particular pointing optionally consists of several microstepped exposures. Microstepping is done by shifting the telescope by a very precise non-integral pixel distance. In the main, WFCAM will probably use a 2x2 microstep sequence and the shift for this sequence will be an integer number of pixels plus a half. Interleaving is done for this sequence by creating an output image that is a regular interwoven pattern of **all** the input pixels. Therefore each output image pixel samples the sky on a finer grid that the input image pixels and helps recover some of the lost resolution for undersampled images. However, interleaving does nothing about removing bad pixels. In situations where the observation includes a dither pattern, this latter can be used to remove bad pixels and cosmic rays. Note that interleaving and/or dithering will result in a new confidence map for the output image. A further issue with interleaving is the potential short timescale variability of the PSF. This may lead to unacceptably "spiky" interleaved PSFs which will require specialised analysis routines to deal with.

Rough World Coordinate System (WCS): basic information in the FITS header as well as knowledge of the instrument will allow us to give each processed frame a WCS that is good to of order several arcseconds. The basic information needed is the RA and Dec of the pointing, a (stable) reference point on the detector grid for those coordinates (this is usually the optical axis of the instrument), the projected pixel size, the rotation of the camera, the relative orientation of each detector and the geometrical distortion of the telescope + camera optics, which defines the astrometric projection to use. Although most of this can be gleaned off-line, deriving accurate values through on-sky observations will be an important task at commissioning time.

First Pass WCS fit: once we have a rough WCS for the processed frames, we will fit a more accurate WCS using secondary faint astrometric standards (*e.g.* USNO V2). If we ultimately wish to mosaic (tile) several of these frames together then this WCS will be used to map the input frames onto the output grid. The procedure is standard and straightforward: extract a (shallow) catalogue from the frame; extract a section of the astrometric catalogue for this region and use the rough WCS to help match the celestial positions with the (x,y) positions on the detector; do iterative clipped least-squares solution; update WCS.

At this stage we have produced single photometricaly and astrometricaly calibrated frames, in principle similar to those produced at ESO in the case of VISTA. The distinction between the products is that the UK, the drivers of VISTA, control this pipeline and how it is run, and whether to reprocess data to improve quality and hence science. This puts the UK in a better position to capitalize on its investment in VISTA. The operation and use of the ESO pipeline is not under the control of the UK.

Before continuing to WP-4 we note a development in monitoring data flow which we also intend to progress, partly for its intrinsic usefulness and partly as it will help with driving WP5 work.

5.2.1.1 Monitoring the Data Flow

There is a need for automatic real time monitoring of the pipeline processing in Cambridge with minimal user input and with simple visualization of overall survey progress. We are developing a POSTGRES database for this purpose and have been trialing it using the Isaac Newton Telescope Wide Field Survey (INT WFS) as a test case.

The development includes interactive GUIs and CLI for querying survey DQC information, state of processing, progress of survey, and should also eventually be capable of feedback to observing strategy. As an example, measures summarising the seeing, image trailing, sky brightness and sky noise, astrometric accuracy and pipeline performance, are generated for each frame processed and stored in the FITS headers. This information is then automatically ingested into the POSTGRES database and used to update the survey progress web pages on-the-fly.

This database will also provide an interim driver for testing some of the advanced processing options (WP5) that require predefined (asynchronous) observations to have occurred in for example, multiple passbands or multiple visits to the same field for deeper stacking.

5.2.2 WP-4 (Single Observation Survey Products)

This part of the pipeline is not required to be delivered to ESO, but is essential for scientific exploitation. The input is the output from WP-3. The outputs are

- Image data products: stacked dither (super)frames; tiled contiguous mosaics combining four pointings with any of previous options; and confidence maps for all pipeline image products.

- Catalogue data products: lists of detected images with an agreed set of parameters (details in 5.2.3) summarising useful astronomical information and providing the necessary DQC information. (Note that the Cambridge processing will have access to an entire nights data and may derive different flatfields, and sky correction frames compared to the causal summit pipeline.)

- Data products will be made securely available for WFAU to transfer to Edinburgh from Cambridge directly via the internet. To download 100 GB per night we will need a rate of about 6 MB/s. Tests have found that this rate is not currently achieved in the route from Cambridge to Edinburgh. Investigations are continuing to identify, and remove or circumvent, the bottlenecks.

In more detail:

Tiling: the pros and cons of stitching interleaved/dithered images together to form a contiguous mosaic tile are:
pros: making better use of fuzzy "dither" edges round each of the 16 "detectors" in a contiguous tile - a large fraction of the total contiguous area is on more than one "detector stack"; generating "smooth" mosaic diagnostics and internal calibration statistics from combining many input detector stacks; reduces by factor ×16 later complexities of overlap cross-calibration; user expectations for contiguous tiled mosaics
cons: involves non-linear resampling and possible use of sub-pixel interpolation schemes; sky variation may make it difficult to achieve; not necessarily using individual integrations optimally.

An important aspect of tiling involves the output WCS. The input projection geometry will have no meaning in a tiled image as there is no unique centre of reference for the projection. The tiling algorithm is therefore free to assign a new projection for the output image and to map the input pixels directly to it. We suggest that output mosaic tiles should all have the FITS TAN projection with a reference tangent point defined to be the centre of the output image. Another possibility is to keep the same projection as used to describe the telescope+camera, Arc projection with Zenithal polynomial distortion. The disadvantage of the latter is the much higher radial distortion term ×200 and the uncertainty over support for the projection in image analysis/display packages.

Tile catalogue generation: generate object catalogues for the tile using the associated confidence map. Whether or not to use tile-generated catalogues as the primary science product is still an open question, so both sets of catalogues will be produced.

Second Pass WCS: If tiling is done, it is a good idea to refit the WCS using all the objects on the tile and using the new WCS geometry. This step is done in exactly the same way as the first pass, except that the input file is now a tile catalogue.

Basic catalogue generation: The basic catalogue generation software makes direct use of the confidence maps and produces the following global information:

- **DQC information:** includes mean sky brightness, sky noise, image detection threshold, average FWHM seeing, average stellar ellipticity and average saturation level;
- **Basic object descriptors:** fluxes, positions and shape information, the list of 75 parameters is defined in www.ast.cam.ac.uk/~dwe/wfcam/docs/newparams2.ps
- **Overlay file:** in addition to the object catalogue, the generation software produces a 'regions' file suitable for use with DS9 to overlay detection ellipses on an image;
- **Astrometric calibration:** as discussed previously, derived from all-sky astrometric catalogues;
- **Photometric calibration:** the generated catalogues will be matched automatically to whatever photometric standard fields are available. More details of this are this in the UKIDSS technical photometric calibration document.

Miscellaneous: these are all potential problems that we have encountered and had to cope with in processing other mosaic camera data:

- undersampling and intra-pixel sensitivity variation problems;
- how to deal with bright stars, ghost images, trails ? One possibility it to use input predictions for positions of bright stars, galaxies, globular clusters, solar system objects and so on;

- cross-talk between arrays and even between readout sectors". If present usually well-defined and dealt with by measuring and applying a cross-talk matrix to remove effect.

- pupil corrections to remove the effects of ghosting due to internal multiple reflections of the night sky. In the worst cases this can severely affect the ability to flatfield accurately.

5.2.2.1 Derivation of further parameters from raw data

The standard pipeline operates on a nightly basis in the sense that the processing requirements are observation driven and are independent of previous, or following, nights data. Further operations on this nightly data are still required, but need detailed knowledge of the Point Spread Function (PSF) and its variation as a function of position within the array. This information is most naturally derived from the output provided by the standard pipeline, thereby defining the architecture for these (nightly) further processing stages. The information would then be included in the Science Archive to enable driving of generation of products from the archive.

- The first step involves computing oversampled, spatially varying, PSFs, possibly using the independent components of the interleave if the seeing is found to vary rapidly on short timescales. This information is also required for several of the database-driven advanced processing options.

- At this stage accurate PSFs are needed for automated PSF fitting to improve the signal to noise ratio of point sources, particularly in crowded stellar regions, and for further input to the image classification scheme. The options for PSF fitting, in order of implementation and increasing complexity, are:
  - profile fit without centroid adjustment using standard multiple components;
  - profile fit with centroid adjustment using standard multiple components;
  - the equivalent of a DAOPHOT/CLEAN algorithm, i.e. fit multiple sources with full iterative refinement, including adding and removing components as necessary. This obviously will not work for galaxies but may be necessary in crowded regions such as in the GPS.

Seeing-deconvolved Sersic profile fits in 1D using the flux in apertures parameter set (with a goal of full 2D modeling later) will be used to quantify galaxy morphology. This also clearly requires accurate knowledge of the stellar PSF. Note that Sersic profiles include both exponential and de Vaucouleurs as special cases.

5.3 WP5 Deliverables (Advanced Survey Product Development)

Broadly speaking development of these tools is the responsibility of CASU, but they will be run in the science archive by WFAU.

The basic goal is to develop Grid enabled pixel processing facilities that are specific to large area Infra Red mosaic imaging systems such as WFCAM and VISTA but by using the emerging Grid and VO standards the software will be of wider generic use for optical and IR imaging astronomy.

The overall long term generic goals of WP 5 were outlined in the original VDFS proposal. Here we focus on the deliverables for the period to Dec 2004 which are tightly coupled with the needs of the WFCAM system (Test-bed for VISTA).

By the delay in the start of this work whilst, of necessity, we work on the prerequisites for producing the good quality data to be accessed, we are effectively able to 'catch the wave' and build on the extensive related work on evaluation of the required technology and standards such as that contained within the Astrogrid Phase A Report.

Version 1.0 of the SA (scheduled for Sep 2003) is effectively a classical static archive offered access to static data products. None of the WP 5 deliverables are scientifically required for the SA Version 1.0. (Data volumes of calibrated pixels and catalogues will be relatively small (cf. the whole problem) in the initial stages of data acquisition).

5.3.1 Database-driven advanced processing

Most of the requirements for an advanced dynamic processing system stem from the necessary combinatorial operations that will be required on images/catalogues derived from the standard calibrated science survey products. Since the time of acquisition of **all** the requisite data is unpredictable this is distinct from the standard processing and some of the further processing options and has to be database-driven.

Many of these advanced survey products are predicated on the ability to automatically generate detailed well-sampled PSFs as a function of position, which will happen at the nightly further processing stage (see 5.2.2.1).

Version 2.0 of the SA is deliverable no later than Dec 2004 (one year after WFCAM survey operations begin) and will include some functionality based on the advanced survey products that are deliverables of WP-5. These will be 'database driven' services and thus will be benefit from the OGSA-DAI work. A detailed set of Use Cases with related UML diagrams have yet to be devised for WP-5 due to lack of available effort. However initial considerations give us confidence that various elements of the recent Astrogrid work on the development of a Web Service that will generate source catalogues with arbitrary user defined input from remote images were a good training exercise in using the current web service protocols (this ran SeXtractor and made use of SOAP, WSDL & XML), and increasing or understanding of the tasks. Work by the US (NVO) Montage project (running on the US Teragrid) also provides useful guidelines on implementing some of the *basic* stacking and tiling preprocessing steps. Below we outline the specific functions that we aim to implement as Web Services and/or Grid Services. These applications will be developed and tested at CASU but made available to users as grid services through WFAU.

- User-specified options for stacking pixel data, i.e. select images to be stacked, and the stacking algorithm from a choice of: i) unweighted; ii) sensitivity weighted; iii) psf matched

- Arbitrary sized, mosaiced images (across tile boundaries), blocked down as appropriate, with a multi-colour option. This will be initiated automatically from a database-driven list of survey fields and survey progress e.g. is it worth (re)stacking (re)mosaicing (re)calibrating this region?

- Generalised difference imaging (and subsequent source analysis) including adaptive kernel matching of PSFs for difference imaging for transient event detection

- Source extraction options on any specified subset or bespoke stack of pixel data: i) CASU standard source extraction; ii) SExtractor; iii) multiple simultaneous profile fitting (i.e. DAOphot-like)

- Ability to analyse archive pixel data (both WFCAM and other, e.g. SDSS) at arbitrary positions defined by an input list of positions, apertures and/or profiles types/models (i.e. list-driven photometry for *any* data to provide the ability to measure fluxes and limits to fluxes in a consistent way).

Based on the lessons learnt from the work on these we will be in a better position to scope the considerable effort needed to implement the more extensive goals identified in the April 2002 VDFS proposal.

5.4 Summary of Effort (work at CASU)

Bottom up estimates of effort (in staff weeks from Oct 02, where one DSY produce 45 staff weeks of work per year) required by task:

5.4.1 V1 Delivery

Interaction with ESO	22.5
Pipeline infrastructure and database progress monitoring development	22.5
Development and testing of standard 2d processing	22.5
Development and testing of standard catalogue products	22.5
Further development of DQC measures at telescope and Cambridge	4.5
Tuning and science verification during commissioning and after	22.5
Development and testing of further processing options 1,2 psf fitting	9

Master catalogue generation, co-located list driven photometry	13.5
Further development of benchmark simple stacking/mosaicing algorithm	9
Continuum subtraction and basic difference imaging	4.5
Interpolation techniques, implications for different stacking methods	9
Modelling and simulations for assessment of parameter reliability	13.5
TOTAL	176

Total is equivalent to 3.9 DSY.

5.4.2 V2 Delivery

Interaction with ESO	22.5
Setting up and managing raw science data archive	9
Ingestion of data, verification, running standard pipeline	45
Setting up and trialing of further processing pipeline	13.5
More modeling and simulations parameter reliability assessment	9
Develop oversampled spatially varying PSF characterisation	9
Further development and assessment of psf fitting options 1,2	13.5
Develop PSF deconvolved Sersic profile fits for galaxies	13.5
Simple methods for LSBGs, nebulosity detection and parameterisation	22.5
Cross photometry on non-WFCAM 2d datasets	22.5
User-specified options for stacking pixel data	40.5
Arbitrary sized, mosaiced images	40.5
Generalised difference imaging (and subsequent source analysis)	18
Source extraction options on any specified subset or stack of pixel data	18
Analysis of archive pixel data at arbitrary positions defined by an input list	18
TOTAL	315

Total v2 is equivalent to 7.0 DSY (to Dec 04 – it is split pro rata into 5.6dsy for Oct03-Sep 04 and 1.4dsy for Sep 04 to Dec 04 in rows 2 & 4 of the overview below).

5.4.3 Effort increase request

The original (Apr02) estimates of staff effort required for WPs2+3, 4 & 5 were 2, 0.5 & 1 dsy in 2002-3 and 2, 0.5 & 2 dsy in 2003-4. The studies leading to specification of deliverables required, and translation into corresponding effort have shown a significant under-resource for WP4. The weeks of effort tabulated above, when turned in dsy (45weeks = 1 dsy) and apportioned by WPs 2+3, 4 & 5, becomes 2, 0.9 & 1 in 2002-3 and 2, 1.6 & 2 in 2003-4. Thus our WP4 requirement of 0.9dsy in Oct02-Sep03 exceeds our Apr02 estimate of 0.5, and our requirement of 1.6 for Oct03-Sep04 exceeds our Apr02 estimate of 0.5.

In view of the imminence of WFCAM data, and the desire to maintain sufficient effort in WP5 to implement some functionality as web/grid services, we request an increase (relative to the Apr 02 estimates for those years) in allocation for WP4 of 0.5dsy for Oct02-Sep03 and 1.0dsy for Oct03-Sep04. Note that the work in WP4 underpins the scientific exploitation of all the data in the archive and its scientific strength in the context of the emerging VO.

V2 of the VDFS-testbed system is due for delivery in Dec 2004, which is 3 months after the monies currently funded by GSC expire. We therefore request pro rata WP4&5 funding for the effort needed in the period Oct-Dec 04 to allow us to complete v2 of the VDFS.

5.4.4 Effort Overview (CASU)

Row		**Phase length Months**	**Effort in Staff weeks**	**Man years Dsy**	**Original VDFS bid (Oct 02-Sep 04)**	**Current request dsy**
	v1 from Oct 02					
1	WP2+3		90	2	2	2
2	**WP4**		40.5	0.9	0.5	1.0
3	**WP5**		45	1	1	1
4	V1 (CASU)	12	176	3.9	3.5	4.0
	v2 to Sep 04					
5	WP2+3		90	2	2	2
6	**WP4**		72	1.6	0.5	1.5
7	**WP5**		90	2	2	2
8	V2 (CASU)	12	252	5.6	4.5	5.5
9	Sum **Oct02-Sep04**	24	428	9.5	8	9.5
10	V2 WP4&5 **Oct04-Dec04**	3	41	0.9	0	0.9
Current Request Summary						
11	Total Oct02-Sep04 Oct04-Dec04	24 | + 3	428 +41	9.5 +0.9	8 + 0	**9.5 +0.9**

5.5 Hardware & Licenses (work at CASU)

CASU require disks for pixel storage, CPU for running the pipeline and accessing the observations database and raw pixel data archive.

5.5.1 WP 2-4 Hardware (CASU)

5.5.1.1 Pixel data (Raw)

Precisely which system to use for data transfer between UKIRT and Cambridge is still under discussion but a strong possibility is to ship using pluggable IDE disks to minimise production and ingestion overheads (*e.g.* foreseeable next generation IDE disk will be 320 GB and should hold ≈ 2-3 nights of data each).

The raw science data will be held online for direct access for three main purposes in order of priority:-
1. to enable automated reprocessing for standard products on major code upgrades/bug fixes
2. to provide a raw science pixel archive for interested users to download unprocessed data
3. as a one-step (reprocessing) removed backup for the science archive at WFAU

Images will be stored (mainly) as 2-byte pixel data. (We assume the data will predominately be in 2 byte form as output from the A/D converter – but if it turns out that, in practice a significant fraction of the data is coadded in the data acquisition feature and is in 4 byte from we would compensate by speeding up our disk purchases).

Costs of suitable scale commodity RAID arrays (≈10TB capacity for each full year of operation) have already fallen and will continue to do so sufficiently that by 2004 when the first tranche of bulk storage will need to be in place) this will be a cost effective and viable (scalable) solution. It is prudent to start with a system that will hold 2.5 years of data (this also protects us from the possibility that raw data from more observations than anticipated arrive in 4 byte form).

5.5.1.2 Pixel data(calibrated) &Catalogue products

At first we will obviously store the processed images online while system shakedown occurs. Later, there is no requirement to keep a processed image repository, but depending on costs of storage, it may turn out to be trivial costwise to do so.

Although we will therefore not usually hold a backup of the calibrated frames sent to WFAU, if recovery of calibrated pixel data from the raw data archive is required this would be feasible as we can re-produce them from the raw data and thereby effectively provide backup for the main on-line science archive at WFAU.

With on-line storage in place at two sites (three if we include raw data in Hawaii) we do not foresee the need for further (expensive) near-line "tape" storage systems for pixel data providing that a high density transfer medium is used from Hawaii, whereby the ≈ 100Gbytes of data expected per night is stored on one "unit" or shipped disk.

We need for disk storage of 25TB to Dec 04 at CASU for raw data and copies of the calibrated data and catalogues during the shakedown phase. We will purchase the hardware at roughly quarterly intervals to maximize return on the spend (Moore's law).

Pixel storage: 25TB for incoming WFCAM data
Total cost: £62,500

5.5.1.3 Transfer of products to Science Archive (WFAU)

Transfer of processed data between Cambridge and Edinburgh will be via the internet. CASU currently share a 1 Gbit/s link into the LAN. Assuming WFAU have a similar level of access by 2003-4, 10% of that bandwidth would suffice to transfer each days processed data in a few hours.

5.5.1.4 CPU

We have sufficient computing power in place as part of the Cambridge Data Processing Centre activities to handle the development issues of the pipeline for WFCAM, including trialing different processing methods and benchmarking simulated WFCAM data processing.

We have estimated the processing requirements for the pipeline itself using our prototype pipeline as developed to process IR data from the IOA CIRSI camera and the WHT INGRID camera. We found (in 2002) a pixel processing rate of 0.2 to 1.0 * 10^5 pixels/sec depending on the level of sophistication of processing used. This was timed on a dual processor 800Mhz P3 system configured with a SCSI-IDE raid array. Such a system would take up to 400 hrs to process a single 8hr night of UKIRT WFCAM data. We would like to be able to process 2 nights of UKIRT data per 24 hr day. At 2002 processing rates WFCAM would require a 32 node Beowulf but assuming Moore's law a 16 node system should be sufficient for WFCAM (2003-Q3), and any spare capacity would be used for testing out some of the advanced developments. In 2003-Q3 we will purchase a 16 node production Beowulf system (effectively 4 nodes per chip and there are four chips).

- CPU: 2x8 node cluster of >2GHz P4 +>2GB memory to deal with the real WFCAM data flow:
  Total cost: £48,000

5.5.1.5 Licenses

Sybase DBMS License: ~£1k/yr
Total cost: £2,000

Other licenses are freeware since we are going to use standard Linux PCs and commodity RAID arrays.

5.5.2 WP5 Hardware (CASU)

For V1/2.0 WP-5 will need some hardware additional to that needed for the development of the standard products.

Hardware specifications and costs including VAT are as follows:

5.5.2.1 Pixel Data

- pixel storage: 5TB Raid Array as a dedicated datastore for the input data and the results **(**not to interfere with the standard processing system)
  Total cost: £15,000

5.5.2.2 CPU

- CPU: a Windows PC to host a local copy of the DQC database.
  Total cost: ~ £2,000

- CPU: two Linux PC's to be the prototype compute grid running the Web and Grid services that we want to integrate and test.
  Total cost: ~ £4,000

- CPU: a further Linux PC for the development work.
  Total cost: ~ £2,000

- 1GB 3COM Managed Hub
  Total cost: ~ £3,000

5.5.2.3 Licenses

WP-5 will need some hardware additional to that needed for the development of the standard products.

£1K per annum
Total cost: ~ £2,000

5.5.3 Summary of Hardware (CASU)

The costings (inclusive of VAT and rounded up to nearest £1K) and schedule (the spend is split over the remaining period to 2004 so as to maximise value for money by delaying purchase as long as possible – for example we will purchase the new disk space required on a quarterly basis) are summarised in the following table:

Item	Description	Unit	Total	Date of
No.		cost	cost	spend
V1	Oct 02-Sep 03			
1	16 nodes of >2GHz P4 +>2GB memory)	£3K	£48K	Q3 2003
2	V1 pixel storage; master node 25 TB disk space	£2.5TB/k	£62.5K	Immediate
3	Windows PC for database	£2K	£2K	
4	2 Linux PCs for running the Web Service's	£2K	£4K	
5	1 Linux PC for development work	£2K	£2K	
7	5 TB Raid Array for datastore	£2.5TB/k	£12.5K	
8	1GB 3COM Managed Hub	£3K	£3K	
9	Licenses (2/yr)	£1k	£4k	
10	Sum items 1-8		£138 k	
11	Less GSC allocation (for CPU & pixel storage) Apr02-Sep03		-£42k	
	Total net (new) request for CASU		**£96K**	
N.B. Apr 02 request was for £110.5				

5.6 WP6 Deliverables (Science Archive: Survey Products Access & Curation)

The input is the output of WP-4 (source catalogues, mosaiced frames, and stacked frames). The outputs are catalogue and pixel data, selected by user-specified criteria, and/or products derived by the users from these, either based on IR camera data alone, or using that data in conjunction with catalogues or pixels from other sky surveys.

The approach adopted is for a testbed v1 delivery by Sep 2003 (to enable the system to be tested with real UKIDSS survey data) followed by a testbed v2 delivery by Dec 2004.

The deliverable of WP5 will function on the science archive at WFAU.

5.6.1 Science Archive Version 1

Version 1 of the Science Archive is deliverable by September 2003 (ready for the start of WFCAM surveys).

5.6.1.1 Contents

SA v1 will contain the following information in a relational DBMS:

1. *Observations Information* containing details of observations contained in the archive and their generic properties;
2. *Image Information* containing details of all images (stored as flat files) in the archive along with housekeeping data (from stripped FITS headers);
3. *Observations Catalogue Information* containing the object catalogues, generated by the CASU standard pipeline, associated with each image, and list-driven source catalogues between the different passbands in any given field;
4. *Merged Catalogue Information* for each of the accumulating UKIDSS subsurveys LAS, GPS and GCS (merged in the sense that the `same' objects observed in different colours and/or at different times will be merged into one multi-colour, multi-epoch record);
5. A *Survey Progress Catalogue*, containing for each of the 5 UKIDSS subsurveys information on observations taken to date;
6. Catalogues for 2MASS, SDSS DR1, SSS, USNO-B, FIRST, IRAS and ROSAT-ASS surveys;
7. Image data (pixels with confidence maps; default stacks) for the deep surveys (flat files)
8. Difference images for the GPS K band (flat files)
9. Large reserve (scratch) workspace for use during querying
10. Online documentation and `cookbook' style worked examples to aid users.

5.6.1.2 Functionality

SA v1 will have the following access points:

1. A web interface allowing searching of individual (or all) UKIDSS survey catalogues on the following criteria:
    - position rectangle expressed in spherical co-ordinates: RA,Dec (J2000); l,b (Galactic); and λ,η (the SDSS spherical co-ordinate system)
    - circular sky area within specified radius from given spherical co-ordinates: RA,Dec (J2000); l,b (Galactic); and λ,η (the SDSS spherical co-ordinate system)
    - circular sky patch within a specified radius of a resolvable source name (using the CDS/NED name resolver)

   and additionally the same searching functions on a user-specified ASCII (space separated) of centres (sexagesimal or decimal degrees) and search radii (i.e. a batch mode search). This interface will also produce ellipse plots for use as finder charts.

2. A web form interface allowing *querying* of individual (or all) SA catalogues (e.g. UKIDSS survey catalogues, housekeeping data, details of archived images) via Structured Query Language (SQL), with push-button options for the format of output data:
    - VOTable format;
    - FITS binary tables;

- ASCII (space, tab or comma-separated);
- uncompressed or lossless compression (e.g. gzip);

combinatorial queries with the 2MASS, SSS, SDSS-DR1 and USNO-B catalogues will be provided for.

3. A web form interface that returns pixel data (images and confidence maps) given an arbitrary input position (as in 1 above) and size up to 0.8^{o} (one WFCAM tile) as follows:
   - mosaiced across any frame boundaries as necessary;
   - FITS format, with user-specified options for lossless or lossy compression;
   - with corresponding *merged* catalogue in the formats specified in the previous item.

`Remote server' functionality for web-based browsing tools (e.g. SkyCAT/GAIA/Aladin) will be provided for some of the above image/catalogue servers, along with a command line interface for remote user non-interactive web access. Archive response time for catalogue queries will be rapid for indexed quantities as follows: position, magnitude, colour, and image class.

5.6.2 Science Archive Version 2

Version 2 is deliverable no later than Dec 2004 (one year after WFCAM survey operations begin), and will include more `database driven' products and features. In addition to contents and functionality provided in V1, the following specifies the V2 contents and functionality to be provided by WFAU. WP5 at CASU will contribute further functionality as detailed above.

5.6.2.1 Contents

In addition to the contents in V1, V2 will contain:

1. A database of open-time observations;
2. Enhanced UKIDSS catalogues containing derived information (e.g. proper motions, dereddened colours, catalogue parameters from placing apertures on SDSS pixels at WFCAM detection positions) where possible using available data;
3. Externally provided catalogues and pixel data (UKIDSS complementary imaging, and SDSS data release as available at that time)

5.6.2.2 Functionality

In addition to the access tools provided in V1, GUI(s) or web/grid services (see WP5) will be provided that have the following functionality:

1. Enhanced output format options to include any new Virtual Observatory standards available at that time;

2. Persistence of multi-stage usage/query; storage of intermediate user-generated results sets
3. User-specified options for stacking pixel data, i.e. select images to be stacked, and the stacking algorithm from a choice of: i) unweighted; ii) sensitivity weighted; iii) psf matched.
4. Arbitrary sized, mosaiced images (across tile boundaries), blocked down as appropriate, with a multi-colour option
5. Generalised difference imaging (and subsequent source analysis)
6. Source extraction options on any specified subset or bespoke stack of pixel data: i) CASU standard source extraction; ii) SExtractor; iii) multiple simultaneous profile fitting (i.e. DAOphot-like)
7. Data exploration/interaction facilities: simple XY plotting; histogram plotting; simple model fitting routines (generalised least-squares with robust outlier rejection);
8. Ability to analyse archive pixel data (both WFCAM and other, e.g. SDSS) at arbitrary positions defined by an input list of positions, apertures and/or profiles types/models (i.e. list-driven photometry for *any* data);
9. Automatic user-supplied catalogue ingest facility for joint querying with existing catalogues;

The access tools in V1 will be supplemented with a `web service' interface (e.g. a non-interactive access tool employing XML format data transferred using Simple Object Access Protocol) to provide, where appropriate, non-interactive access to pixel and catalogue data. Archive response time is to be ~100 sec for wholesale catalogue trawls on non-indexed quantities.

5.7 Summary of Effort (work at WFAU)

5.7.1 V1

Bottom up estimates of v1 effort (in staff weeks from Oct 02, where one DSY produce 45 staff weeks of work per year) required by task:

Design V1 DB (SQLserver schemas) to hold specified information from Surveys (observation info, image info, object catalogues, merged survey catalogues, progress catalogue) and open-time WFCAM observations	31.5
Design DB schemas for external survey catalogues specified e.g. 2MASS, USNO B, IRAS	4.5
Implement spatial and other indexing for catalogues	9
Test DB hardware, install DBMS, prototype and tune DB system, initially with SSS data and other external catalogues specified.	9
Design, build, test tools for ingesting image and object data received from CASU	9
Design, build, test tool for producing merged catalogues from newly ingested data	9
Design, build, test tool for photometric or astrometric recalibration	4.5

Design, build, test WWW interface and tools for searching UKIDSS by position (equatorial, galactic or SDSS spherical) and returning finder-chart (10′) images	9
Design, build, test WWW interface and tools for searching all UKIDSS catalogues and in combination with external (2MASS, SDSS DR1, SSS etc.) catalogues via SQL, returning tables in specified format, e.g. ASCII, FITS table, VOTable	13.5
Design, build, test WWW interface and tools for searching UKIDSS catalogues by position to return mosaiced images (up to 0°.8) with matched catalogue	22.5
Provide remote server (SkyCAT/GAIA, Aladin) access for Search Tools above	4.5
Implement and test tools and interfaces to generate stacked images given a user-selected list of input images	9
TOTAL	135.0

Total v1 effort is equivalent to 3.0 DSY.

5.7.2 V2

Benchmark SQLserver, Oracle and DB2 DBMS using SSS data on different hardware to measure performance and scalability for V2.	22.5
If new DBMS selected, revise schemas to hold specified information from Surveys and open-time WFCAM observations.	4.5
If new DBMS selected, install DBMS, and migrate data from V1 archive	9
If new DBMS selected, revise tools for ingesting image and object data received from CASU and producing merged catalogues	9
If new DBMS selected, revise WWW interface and tools for searching UKIDSS by position (equatorial, galactic or SDSS spherical) to return finder-chart (10′) images.	4.5
If new DBMS selected, revise WWW interface and tools for searching all UKIDSS catalogues and in combination with external catalogues via SQL, returning tables in specified format, e.g. ASCII, FITS table, VOTable	4.5
If new DBMS selected, revise WWW interface and tools for searching UKIDSS catalogues by position to return mosaiced images (up to 0°.8) with matched catalogue	4.5
Extend WWW interface and tools for searching UKIDSS catalogues by position for finder-chart images to search data from open-time observations, to produce colour-pixel plots and to allow ecliptic and super-galactic coordinates	9

Extend WWW interface and tools for searching all UKIDSS catalogues and in combination with external catalogues to allow user-supplied catalogues and searches on parameter ranges and arithmetic functions of parameters e.g. optical-IR colours	13.5
Implement and test tools to generate optical-IR colours for IR objects using SDSS image data where necessary.	13.5
Extend WWW interface and tools for searching UKIDSS catalogues to return mosaiced images larger than in V1, blocked down as specified by user.	4.5
Extend production of stacked images to allow use of user-chosen stacking algorithms	4.5
Extend interface and tools to allow complex and multi-stage searches	22.5
	126

Total v2 is equivalent to 2.8 DSY (to Dec 04 – it split pro rata into part of Sep 04 and part Sep 04 to Dec 04 in rows 2 & 4 of the overview below).

5.7.3 Effort increase request

The original estimate of staff effort required for WP6 was 2.0 dsy in 2002-3 and 2003-4. The studies leading to development of functional specifications and translations of these into effort show the 2.0 dsy to be a significant underestimate. We have identified ways of transferring some effort from other work, some temporarily funded by the University of Edinburgh, but cannot make up the whole shortfall this way. We request additional support of 0.5 FTE (0.5 person working) from Oct 2003., to help bridge the gap and deliver the agreed functional specifications. (This translates 0.25 dsy (to start in March 03) in Oct 02-Sep 03, 0.5 dsy in Oct 03-Sep 04). We have a model for acquiring this effort: an individual with substantial database experience and astronomical knowledge currently funded 50:50 by Starlink and AstroGrid. Subject to the agreement of the GSC (Starlink management are content), we propose transferring the Starlink half of this person's effort to the WP6 work.

We also request funding for the effort (0.6dsy) needed in the period Oct-Dec 04 to allow us to complete v2 of the VDFS.

5.7.4 Effort Overview WP6 WFAU

Row		Phase Months	Staff weeks from Oct 02	Man years	Original VDFS bid (Oct 02-Sep 04)	Current request
1	V1 from Oct 02	12	135	3.0	2.0	2.25
2	V2 WFAU to sep 2004	12	100.8	2.24	2.0	2.5
3	Sum of row 1+2 Oct02-Sep04	24	235.8	5.24	4.0	4.75
4	V2 WFAU from Oct 2004	3	25.2	0.6	none	0.63
Current Request Summary						
5	Total Oct02-Sep04 + Oct04-Dec 04	24 +3	235.8 +25.2	5.24 +0.6	4 + 0	**4.75 +0.63**

5.8 Hardware & Licenses (work at WFAU)

5.8.1 Introduction

5.8.1.1 Development

SA developments are documented at http://www.roe.ac.uk/~nch/wfcam, where much more detail and plans can be found. The aim of this document is to present specific plans for hardware, operating system and DBMS software for the period to September 2004. This document will also mention relevant work that has been done to date to justify our choices. The reason for considering hardware, OS and DBMS together is that the three are intimately linked; optimal DBMS operation is not always possible for a given hardware and OS configuration.

5.8.1.2 Dilemma

In our developments for the SA so far, we have had to deal with some tensions. The need for timely development work (e.g. hands-on experience) has to be balanced against the general best practice of delaying hardware purchases as long as possible (e.g. Moore's law). Further, the timescale for significant developments in computer hardware technology implies that over the SA lifetime the archive hardware solution is likely to change and a migration from one system to another will almost certainly be needed; this is even more so for the overall VDFS when it ingests VISTA data. Hardware configuration has many complicated variables, and it is only via experimentation that most questions can be answered.

5.8.1.3 Good News

However, the good news is as follows. Hardware manufacturers (e.g. Sun MicroSystems, IBM, Compusys) are open to donation of discounts, money and/or hardware for what they see as big projects in R&D. Furthermore, many hardware suppliers are open to loan of high-specification kit, meaning that experimentation is possible with no outlay on hardware. The hardware/software `data warehouse' big players (i.e. MicroSoft/SQLServer, IBM/DB2 and Oracle) are open to supplying expertise and advice to large database developers and more importantly to supplying licenses that are heavily discounted or even free-of-charge. Finally, many of the VDFS Phase A hardware issues overlap with similar ones in other IT projects in the UK (e.g. initiatives at the Edinburgh National e-Science Centre, within AstroGrid, etc.) and farther afield in Europe (e.g. ASTRO-WISE: http://www.astro-wise.org ; NGAST: http://archive.eso.org/NGAST).

5.8.2 Plan

5.8.2.1 Phase and Split

The clear requirement to arise out of the functional requirements is for a phased approach so that a SA with `standard' functionality is available at first light, enabling immediate science exploitation, and subsequently a fully functioning archive system is made available one year after survey operations begin in earnest. Furthermore, there is a clear split in the requirements for volume and access speed for catalogue and processed pixel data (hereafter, `pixel data' will mean processed pixels; note that there is no requirement on the SA to store the raw pixel data). The SA usages require user interaction with *catalogue data* (i.e. complex queries returning results in as close to real time as is feasible) for data mining and data exploration while *pixel data* usages are less time-critical (users would be prepared to wait for the results of operations on large pixel volumes since these would be executed relatively rarely).

5.8.2.2 Different Hardware for catalogues and Pixels

The phased approach and volume/speed split is reinforced in the light of the hardware considerations above: we will implement two distinct hardware solutions to satisfy the user requirements. If one subsequently becomes a viable solution for both, but at the same time at reduced cost, then migration from one solution to the other will be straightforward. Furthermore, the phased approach maximises the possibility of exploiting the most recent advances in computer technology since we will not be wedded to one hardware solution from the start and will delay as long as possible the final hardware acquisition (for example, we present a revised hardware plan here because there have been significant hardware developments since submission of the original VDFS proposal).

5.8.2.3 Hardware Requirement

The outline hardware requirements are as follows:

- V1 system at WFCAM first light (currently Q4 2003);
- V2 system at one year after survey operations begin (currently end of Q4 2004, but note acquisition of the hardware takes place in Q2/Q3 2004);

- WFCAM pixel volume: 20 TB/year; speed of access is not a critical issue for pixel data (calibrated data from CASU is in 4 byte form so calibrated volume exceeds that of the 2byte raw data);
- Other pixel data requiring storage will be dominated by the SDSS pixels: DR1/DR2/final releases will be 15%/50%/100% complete where the total SDSS pixel volume is ~ 10 TB. V1 will include DR1 and V2 will include DR2; SDSS pixel volumes are 1.5 and 5 TB respectively for DR1 and DR2.
- WFCAM object catalogues/ancillary data: 2 TB/year; `real-time' access is required (i.e. allow users to interact with and explore the data and we suggest ~ 100s response time is therefore necessary);
- Other catalogues requiring storage: again, see the FRD; they will be dominated by the SSS & SDSS each of which is of order 1 TB.

Hence the split is V1/V2 and pixels/catalogues. After one year of operation, the V1 catalogue volume is 5 TB; the V1 pixel volume is 22 TB. Data accumulation is then 20 TB/year (pixels) and 2 TB/year (catalogues), so after one year of operation of the V2 archive the pixel volume will be larger by 25 TB (including SDSS DR2) and the catalogue data volume will be larger by 2.5 TB (including SDSS DR2).

5.8.2.4 Phasing of Purchases

In order to have the V1 catalogue system running in good time and to implement SA-like services with an existing dataset, we require to purchase the V1 catalogue hardware <u>now</u> and implement the TB SSS on it as a first step. We have opted for SQLServer/W2K as the DBMS/OS choice so as to take advantage of the development work done for the SDSS `DR1' system (known as `SkyServer'). Hence at first light, the SA system will be functional and contain the required external catalogue data specified in the requirements along with the necessary WFCAM data schemas - we will be ready to ingest WFCAM data. For the V1 pixel system, we intend to follow ESO and implement an NGAST-like system (see previously for the NGAST URL) which comprises relatively low-cost bulk storage employing IDE disks. We propose to store the pixel data in a flat-file Linux file system; a database of the contents, details (e.g. header information) and location of given files within that system will be part of the queryable catalogue database.

5.8.2.5 Load Server

One further point, given advice from our SDSS colleagues and based on our own experience of loading data, is that we will require a dedicated `load server' for that catalogue system. This is because the loading process (including ingest, indexing and other miscellaneous curation tasks) has significant CPU and IO overheads that must not interfere with user access to the system. Additionally, it is important to have a secure load server that is not generally accessible to the outside world, and that load server should have sufficient storage to `mirror' the database tables being updated.

5.8.2.6 Scalability

The V2 catalogue system must be scalable to data volumes ~ 10 TB and beyond. Currently, there is some uncertainty as to the suitability of SQLServer for this; maintaining rapid response for such large data volumes requires high aggregate I/O and `threading' of processor tasks coupled via high bandwidth interconnect to the disk store. Oracle or DB2 may be a more suitable choice, but this needs more investigation. However, our V1 catalogue solution is modular and scalable, and provides enhanced CPU and storage with each additional unit of 2.4 TB. We propose acquiring a further 2.4 TB server unit (dual processor) for V2, but note that for the same price we expect to acquire a quad processor server in place of the dual processor; this enhanced CPU power will likely be needed for V2 and beyond. The V2 pixel store is more easily specified: the V1 solution is essentially infinitely scalable and we require an additional 25 TB of storage for the first year of V2.

5.8.2.7 Summary

We believe the schedule for hardware acquisition presented here is pragmatic and sensible given the requirements and likely evolution of computer technology. The costs are based on explicit quotes from a competitive supplier; the exact specification of the hardware we acquire may not exactly be as listed, but at no significant change in cost. In further rounds of e-science money we expect to be requesting funding to fulfill the running requirements (especially the storage requirements for accumulating data).

5.8.3 Costs

5.8.3.1 Hardware

Hardware specifications and costs excluding VAT [VAT is included in the summary Table in 5.8.4] are as follows:

V1

- catalogue servers: two dual Pentium IV Xeon systems with 2.4 TB fibre IDE RAID each:
  Total cost: £25,560

- catalogue load server: one of the above systems:
  Total cost: £13,280

- pixel storage: 2.4 TB IDE RAID file server plus 9× 2.4 TB IDE RAID file nodes yielding total storage of ~ 22 TB
  Total cost: £61,336 (ex VAT)

V2

- catalogue server: additional 2.4 TB server as specified for V1:
  Total cost: ~ £13,000

- pixel storage: another 10 IDE RAID file nodes to expand capacity a further ~ 24 TB; cost as for V1 but noting that 300 GB disks will likely replace 200

GB devices for the same price (i.e. an overall price reduction of a factor 2/3):
Total cost: ~ £40,000 (ex VAT)

V1/2.0

- catalogue backup system: LTO 2 tape library backup system and media with capacity 6 TB :
  Total cost: £12,625 (ex VAT)

Note: a backup system for the catalogues is essential to fulfill security requirements. Backup on removable media is important to be able to quickly recover the catalogue database without having to repeat the ingestion process. On the other hand, backup of the pixel data is not required. This is because copies of the processed pixels will be recoverable from the processing centre (CASU); the pixel system will run RAID-5 for fault tolerance against individual disk failure. Note also that all the above hardware solutions are modular and expandable and can be incorporated into new hardware configurations as the archive system evolves towards the VISTA era.

5.8.3.2 DBMS Licenses

Licencing is required for development work and for deployment. These need to be considered separately since the former is contained within the WFAU research group while the latter is running on a multi-node system freely accessible over the internet to an unspecified number of users, and as such is sometimes licenced at high cost by the licence vendors.

We have development licences for SQLServer and DB2 provided under the various companies educational and academic licencing schemes. For MicroSoft, we are a member of the MicroSoft Development Network Academic Alliance, see http://www.msdnaa.net; for DB2, we have a development licence under IBM's Scholar's Programme, see http://www-3.ibm.com/software/info/university. Development licences for Oracle can be downloaded free-of-charge from their website. We have been advised by Microsoft that the MSDNAA scheme will cover a deployment system and have also been provisionally advised by IBM that this is also the case for DB2 and our Scholar's Programme licence. Oracle deployment licencing is potentially more costly. The University of Edinburgh has an Oracle Campus licence; these campus licences are heavily discounted (95%) and then the remaining cost spread amongst the departments requiring use. We have been advised that WFAU's contribution for usage of the UoE campus licence, which would cover deployment, would currently cost £1338+VAT per processor per year; we require 8 processors in our V2 system (as detailed above). Hence, licencing for the first year of operation of V2 would cost £13K, in subsequent years this cost would be less as more contributions to the overall cost are made by increasing numbers of participating departments. We would like to point out the huge savings implied by the above licencing arrangements, irrespective of the DBMS finally chosen. The costs of a perpetual (i.e. one-off payment) commercial licence *per processor* are £30K, £3K and £1K respectively for Oracle, SQLServer and DB2, implying massive savings (> £100K for example for Oracle).

Moreover, such perpetual licences do not cover updated editions of the DBMS software product in future, unlike the arrangements for the Academic, Scholars or campus licences.

5.8.4 WFAU Hardware Summary

The costings (inclusive of VAT and rounded up to nearest £1K) and schedule (the spend is split over the remaining period to 2004 so as to maximise value for money by delaying purchase as long as possible) are summarised in the following table:

Item No.	Description	Unit cost	Total cost	Date of spend
V1	Oct 02-Sep 03			
1	V1 catalogue servers ×2	£16K	£32K	Immediate
2	V1 catalogue load server	£16K	£16K	Immediate
3	V1 LTO tape library (catalogue backup)	£15K	£15K	Immediate
4	SQLServer licence (development/deployment)	£0K	£0K	-
5	DB2 licence (development/deployment)	£0K	£0K	-
6	Oracle development licence	£0K	£0K	-
7	V1 pixel storage; master node	£12K	£12K	Immediate
8	V1 pixel storage; slave nodes ×9	£7K	£63K	Q4 2003
	Sum items 1-8		£138k	
9	Less GSC allocation (for pixel storage) Apr02-Sep03		-£90k	
10	**Net request for WFAU hardware enhancement Oct02-Sep03**		**£48k**	
V2	Oct 03-Dec 04			
119	V2 catalogue server	£16K	£16K	Q3 2004
120	Oracle deployment licence	£13K	£13K	Q3 2004
131	V2 pixel storage	£36K	£36K	Q3 2004
14	**Net request for WFAU hardware Oct03-Dec04**		**£65k**	
	Total net (new) request for WFAU (item10+item14)		**£113K**	
N.B. Apr 02 request was for £65.0k				

6 Review & Actions requested

6.1.1 Total Costs Summary (from Oct 2002)

		Apr 02 bid	Apr 02 bid	Current request	Current need
Effort		dsy	£k	dsy	£k
CASU WPs4&5 (see 5.4.4)	Oct02-Se03	1.5	97.5	2.0	130.0
WFAU WP6 (see 0)	ditto	2	130.0	2.25	146.3
CASU WPs4&5 (see 5.4.4)	Oct03-Se04	2.5	162.5	3.5	227.5
WFAU WP6 (see 0)	ditto	2	130.0	2.5	162.5
WPs4&5 (25% x Oct03-Sep04)	Oct04-De04	0	0	0.9	56.9
WP6 (25% of Oct03-Sep04)	ditto	0	0	0.63	40.6
WP0-3&7 (25% of Oc03-Se04)	ditto	0	0	1.1	69.9
Total Effort (see note)	Oct02-Se04 Oct04-De04	12 +0	£520.0 +0	14.25 +2.6	**£666.3** **+£167.4**
The 2.25 extra effort compared to Apr 02 estimates represents, 1.5 increase for CASU, and 0.75 increase needed for the archive at WFAU.					
Hardware 2003-4					
CASU Hardware see 5.5.3	2002-3 extra	-	no extra	-	£0.0
"	2003-4	-	£110.5	-	£96.0
WFAU Hardware see 5.8.4	2002-3 extra	-	no extra	-	£48.0
"	2003-4	-	£65.0	-	£65.0
Total Hardware 2003-4		-	£175.5	-	**£209.0**
The £33.5k extra hardware represents about 7 of the 9 v1 pixel storage slave nodes.					
Grand Total (£thousands)	Oct02-Sep04 Oct04-Dec04		£695.5 +£0		**£875.3** **+£167.4**

Note: As in the original VDFS bid we have assumed an average cost of £65k/dsy based on actual mean costs corresponding to a senior PDRA @ £32.5k +22%NIS +46% overhead + 2.2k 8% secretary with overhead + 3.9k travel + 1k workstation/laptop provision.

6.1.2 Actions

Actions requested of GSC are:

- GSC is requested to approve continuation of funding for WP4 (Pipeline) work, and to agree to fund WP5 (Advanced Survey Product Development) work.
- GSC is requested to approve to continuation of funding for WP6 (Archive) work by WFAU.

- GSC is requested to agree the 5.5 dsy level of manpower (and consequent funding) for WP4&5 work by CASU, which includes an additional (relative to Apr 02 VDFS bid) 1.5 FTE from March 2003.
- GSC is requested to agree the 4.75 dsy level of manpower (and consequent funding) for WP6 work by WFAU, which includes an additional (relative to Apr 02 VDFS bid) 0.75 dsy from March 2003.

- GSC is requested to agree hardware funding of £96k for of 2003/4 for work by CASU.
- GSC is requested to agree hardware funding of £65k for of 2003/4 for work by WFAU, and a further £48k for 2002/3.

- GSC is requested to approve an additional 3 months of effort funding for the VDFS to take it to the Dec04 delivery of the v2 system. This should be at a rate that provides a further 0.25 x (Oct03-Sep04 effort) which totals £167.4k.