Data Flow System

Document Title: **VISTA Data Flow System Pipeline –
Revised Plan**

Document Number: **VIS-***-IOA-*****-******
Issue: **1.0**

Date: **2005-09-05**

Document Prepared by:	**Mike Irwin Simon Hodgkin** (CASU)	**Signature And date:**	
Document Approved by:		**Signature And date:**	
Document Reviewed by:	**Jim Emerson**	**Signature And date:**	
Document Released by:	**Jim Emerson**	**Signature And date:**	

Change Record

Issue	Date	Sections Affected	Remarks
0.1	2005-08-31	All	Draft document
1.0	2005-09-05	All	Reformat by JPE. Add VEGA text. Add Introduction. Remove ESO parts.

Notification List

The following people should be notified by email that a new version of this document has been issued:

IoA	Will Sutherland
QMUL	Jim Emerson
ATC	Malcolm Stewart
Exeter	Tim Naylor

Contents

1 Introduction

Section 2 reproduces the original project plan from the VEGA application in Feb 2004 which described five versions (VDFSv1-v5) of the Pipeline, to be released annually at the ends of 2003-2007. Versions VDFSv1-3 were for WFCAM data, offering progressively more functionality, while VDFSv4 is the WFCAM pipeline converted for VISTA. The final release, VDFSv5, is fully shaken down with VISTA data.

Owing to the delay in delivery and commissioning of WFCAM, the time-scale between WFCAM coming into operation and the scheduled VISTA delivery has been greatly compressed, necessitating revision of the plan. Now that WFCAM science verification data are being used by the community, it is timely to revisit the plan and we give our new proposals in Section 3 which specifies the pipeline version deliverables and gives an updated estimate of the timeframe for delivery.

Section 4 goes through each of the UK URD requirements and indicates which will need more development effort, which are already covered, and also provides some additional comments where appropriate.

As an appendix Section 5 gives the VEGA proposal plan for the Science Archive.

2 VEGA 2.1.3.1 VDFS Pipeline releases (UK part)

The UK pipeline processing for VISTA and WFCAM has to fulfill a series of requirements that lead to the following overall design and implementation strategy.

VDFS-v1(WFCAM) delivered end Q4 2003, originally to be ready for WFCAM commissioning
- standard pipeline - run in Cambridge and operates on individual nights of data, main deliverables are: DQC information; astrometric and initial photometric calibration; standard catalogue detection and parameterisation; final instrumental signature-free 2D science images
- database driven products - asynchronous processing capability with deliverables trial versions of: general purpose list-driven photometry package; difference imaging and continuum subtraction software for narrow band work; stacking and mosaicing software

VDFS-v2(WFCAM) to be released end of Q4 2004
- as for v1 but with final versions of standard pipeline tuned for WFCAM data as a result of commissioning characterisation; raw WFCAM data archive set up and transfers to ESO data archive begun
- further processing - trial version of pipeline running on individual nights of data that delivers: automatic PSF generation; PSF profile fitting for stellar images; general purpose Sersic profile fitting for galaxy images

- database driven products - final version of general purpose list-driven photometry package including photometry driven from externally derived lists; final version of difference imaging and continuum subtraction software; enhanced functionality for stacking and mosaicing software; trial software version for assessment of nebulosity and low surface brightness (LSB) object detection and parameterisation

VDFS-v3(WFCAM) to be released end of Q4 2005
- as for v1,v2 but with final version of further processing pipeline fully tuned for WFCAM data as a result of science data characterisation and delivery of full iterative PSF model fitting
- optimised stacking and mosaicing software designed to address a range of problem-specific functionality including variable PSFs
- dependent on v2 outcome, if feasible, more general purpose nebulosity and LSB feature detection and characterisation
- detection of transient phenomena using full adaptive kernel difference imaging software, including re-running image detection algorithms
- begin development of VO-enabled pipeline processing based on AstroGrid deliverables

VDFS-v4(VISTA) released Q4 2006 (ESO deliverables in Q3)
- as for v1,v2,v3 but on VISTA data

VDFS-v5(VISTA) released Q4 2007, final development release for VISTA
- final shakedown and tuning of pipelines for VISTA based on experience of running VISTA pipeline for a year
- completion of interfaces to VO including enabling of remote configuration and running of pipeline components

The pipeline design philosophy together with a more detailed presentation of the components passed a CDR in October 2002.

The advanced pipeline functionalities required for the database-driven products will be provided to WFAU for integration within the Science Archive.

3   Updated VDFS pipeline deliverables timetable

Note there is a 1 quarter shift between CASU & WFAU version delivery dates (CASU after WFAU) – this needs removing after further discussion.

3.1   VDFS-v1.x delivered Q2 2005, tuned with WFCAM semester 05A data

- standard pipeline (some catalogue parameters still to be added)
- pre-WFCAM version advanced processing software (stacking, mosaicing, difference imaging, list-driven photometry)
- raw WFCAM archive setup and transfers to ESO archive initiated
- trial versions of automatic PSF generation and PSF fitting (1,2) running on test nights of data

3.2   VDFS-v2 by end of Q4 2005 - WFAUv2 is 1Q earlier

- as for v1 but with final versions of standard pipeline including full standard catalogue parameter set tuned using 05A data
- further processing PSF fitting stage 1,2 (tested by completing for 05A data)
- trial version of Sersic profile fitting for galaxy images
- database driven software updated for full WFCAM compliance and delivered to WSA including stacking, mosaicing, difference imaging, list-driven photometry and catalogue generation software assessment of feasibility of LSB object detection and nebulosity characterisation (possible descope item)
- photometry characterised and secondary photometric standards setup
- VST data products design requirements assessment and schema for implementation
- assessment of extra processing requirements from UK URD for VISTA

3.3   VDFS-v3 by end of Q2 2006 - WFAUv3 is 1Q earlier

- as for v1,v2 but with final version of further processing pipeline fully tuned on 1st year of WFCAM data (expect to descope by not attempting full iterative PSF model fitting)
- optimised stacking and mosaicing software coping automatically with variable seeing and extinction/throughput
- detection of transient phenomena using optional adaptive kernel difference imaging software or aperture-based list driven lightcurve analysis
- LSB object detection and nebulosity characterisation (if not descoped by reality)
- trial versions of some of the enhanced software modules to meet surviving extra UK URD requirements (mainly stacking and mosaicing at this stage) tested on WFCAM data

3.4 VDFS-v4 by end of Q4 2006 - as for WFAU

- As for v1,v2,v3 but handling VISTA data volumes

3.5 VDFS-v5 by Q4 2007- WFAUv5 is 1Q earlier when grant ends

- final shakedown and tuning of UK pipelines for VISTA based on real science data
- photometry characterised and secondary photometric standards setup

4 UK URD point-by-point

4.1 General Requirements

5.1	covered by the existing design
5.2-3	WFAU, though there is a worry about agreeing to standards that don't exist yet
5.4	covered if buy enough hardware
5.5a	covered by the existing design
5.5b	the comment is covered by 5.5a, *i.e.* there will be a traceable series of processing steps recorded in the headers that provide enough information to do a comparison given the relevant technical description (TBC)
5.6	covered by the existing design
5.7	we need to discuss who is going to do this
5.8	covered by existing design
5.9-17	WFAU

4.2 Astrometric Requirements

6.1	covered by existing design (WFCAM)
6.2	covered by existing design (WFCAM)
6.3-4	covered by existing design (iff sources have high enough signal:noise and iff intra-pixel effects and colour-dependent and other atmospheric effects are not an issue)

4.3 Photometric Requirements

7.1-2	covered by existing design (WFCAM) but 1% may be impossible
7.3	covered by existing design (WFCAM)
7.4	the goal is hard since if you defocus too much it is difficult to measure the aperture corrections accurately and automatically.

4.4 Tiling, Stacking, Microstepping

8.1	covered by existing design
8.2	would require modifications of existing software, but we have severe doubts about the usefulness of this, surely better to do variable objects from either catalogues or difference imaging, since at the pixel level all objects vary due to seeing differences
8.3	from a pipeline POV the same stacking for all is preferred, otherwise as we have noted previously, the information about which stacking algorithm to use has to be propagated all the way through the observing system (the choice of stacking method would have negligible impact on derived QC parameters). Drizzle, SWARP, MONTAGE *etc* are available as external packages so cost-wise it would be better to run different stacking/tiling choices on the archived products
8.4	would require modifications of existing software, however these are already planned and needed elsewhere
8.5	only a modest change required to handle 2x1 interlacing, though the 45 deg implicit rotation requires more thought
8.6a,b	if a choice is really needed this generally moves tiling to the archive end. But why would you want to "trim", since you are throwing information away. Which contributing pixels would you choose ? Since the associated confidence maps convey all this information anyway what's the point ? [its effectively no different to taking the normal stacked pawprints and adding noise to higher confidence pixels]. Another possibility is to modify the catalogue generation software to deal with this (in fact it already does) Option b. is covered by existing design.

4.5 Variable Objects

9.1	could use catalogues or difference imaging for this - see comment to 8.2
9.2	WFAU
9.3-4	WFAU time series analysis is best done as a list-driven photometry application from a master catalogue at the archive end

4.6 Object Catalogues

10.1	covered by existing design
10.2	covered by existing design
10.2	covered by existing design
10.4	WFAU
10.5	WFAU: not all asteroids are catalogued and the surveys will turn up many more than are currently known about which implies this type of detection/flagging is best left until later catalogue matching stages

10.6	probably covered by existing design though would need assessing with real VISTA data
10.7	probably requires a more sophisticated detection filter, this is specialist LSB galaxy detection, feasibility would need assessing with real VISTA data. [10-sigma at 4.5x is probably currently met]
10.8-9	radioactive detectors and large patches of bad pixels notwithstanding
10.10	in planned design enhancements for WFCAM
10.11	WFAU this may be tricky if the system does not have reasonably uniform sensitivity (cf. WFCAM)
10.12	same as 9.3?
10.13	same as 9.4?

5 Appendix VEGA 2.1.3.2 VDFS Science Archive

5.1.1 Science Archive Releases

The main feature of the VDFS Science Archive plan is a phased approach, based on delivery of a science archive for the UKIRT WFCAM in the first instance (the VDFSv1 science archive) followed by development of that system through subsequent versions to a fully-fledged, grid-enabled archive system for VISTA. The major milestones and deliverables in this plan are:

- a Critical Design Review (CDR) of the design of the WFCAM Science Archive (WSA) at the beginning of Q2 2003;
- delivery of the V1 WSA (VDFSv1) at the end of Q4 2003;
- a Preliminary Design Review (PDR) of the VISTA Science Archive in Q3 2004;
- begin operation of V1 WSA following commissioning of instrument in Q3 2004
- delivery of the V2 WSA (VDFSv2) one year after V1, and V3 WSA (VDFSv3) one year after V2;
- an Final Design Review (FDR) of V1 of the VISTA Science Archive in Q3 2005;
- delivery of the VSA V1 (VDFSv4) science archive at Q4 2006, ready for commissioning with real VISTA data
- begin VSA (VDFSv4) science operations Q1 2007
- release VDFSv5 system based on experience and user feedback of VDFSv4, end Jan 2008.

VDFS science archive development is planned to end one year after delivery of the final archive system to allow for a probationary maintenance period. Operation of the science archive begins in Q3 2004, immediately after installation of WFCAM, and continues in parallel with development to the end of VDFS project and beyond.

VDFS science archive development is comprehensively documented at the project homepage   http://www.roe.ac.uk/~nch/wfcam . For an overview of the science archive project, see http://www.roe.ac.uk/~nch/wfcam/VDF-WFA-WSA-001-I1/VDF-WFA-WSA-001-I1.html . The following section describes development work done to date.

VDFS-v1(WFCAM) Delivered end Q4 2003; ready for WFCAM commissioning

Will allow access to the basic data products, will have a web-based interface, and documentation. It will include access to 2MASS, SSS, SDSS-DR1, and USNO-B catalogues as well as WFCAM data. It will allow the user to :

- Browse the database of calibrated observation frames and associated catalogues, and download any one of these
- Produce small images and associated multi-colour merged catalogues, from either the observation frame collection, or the accumulating UKIDSS surveys
- Produce mosaic images on demand up a width of 0.8 degrees
- Search UKIDSS surveys on a position rectangle or circle in Celestial or Galactic co-ordinates, and within a radius of a resolvable source name
- Allow joint position searches as above, and produce merged catalogues, with the combination of UKIDSS plus any of 2MASS, SSS, SDSS-DR1
- Make SQL-like queries in any sensible arithmetic combination of the catalogue parameters, for example colours
- Allow general queries on the combination of UKIDSS, 2MASS, SSS, SDSS-DR1

VDFS-v2(WFCAM) released end Q4 2004

Will be compliant with any relevant protocols and formats produced as part of the various Virtual Observatory initiatives, and will work especially closely with AstroGrid. It will allow the user to :

- Search open-time data as well as UKIDSS data
- Plot returned parameters in histograms
- Plot pairs of returned parameters as (x,y) graphs
- Generate on request images and catalogues across observation frame boundaries
- Generate mosaic and stacked images and source catalogues from any area, from user specified observation frames
- Choose a range of stacking and source extraction algorithms on request, with a menu of tunable parameters

Probably other simple user tools will be available, such as the SED tool developed for the AVO first-light demo with the GOODS dataset. v2 will also :

- o Return catalogues and similar data in VOTable format, and images using binary data access protocols still under development
- o Recast all the V1 services as web services for publication in the AstroGrid Registry

VDFS-v3(WFCAM) released end Q4 2005

- o Develop (or install) a suite of advanced visualisation tools, e.g. pannable large area imaging, multi-dimensional catalogue plotting and rotation
- o Develop (or install) a suite of data analysis tools, e.g. cluster analysis, principal component analysis, etc
- o Implement both the above as server-based tools
- o Develop a system to allow uploadable user-specified analysis algorithms
- o Recast our web services as grid services, in order to allow queries and analysis using shared managed resources with other data centres

VDFS-v4 released Q4 2006 (ESO deliverables in Q3) to operate on VISTA data

- as for v3 except handles the much greater VISTA data rates

VDFS-v5 released end Jan 2008 final release for VISTA after 18 months shakedown period

- incorporates changes from experience with running actual VISTA data, and completion of interfaces to the VO