VISTA DATA FLOW SYSTEM (VDFS)

for VISTA & WFCAM data

Science Archive Science Requirements Analysis Document

author
N.C. Hambly (WFAU Edinburgh)
Science Archive Project Scientist

number
VDF-WFA-VSA-002
issue
1.0
date
September 2006

co-authors
P.M. Williams, M.A. Read

Contents

1 SCOPE

The large format VISTA and WFCAM infrared cameras will have an unprecedented data rate. Ultimately, successful science exploitation of VISTA and WFCAM will depend on user access to the large data volumes generated by the instruments. Data volumes are far in excess of those that users can expect to hold and process on their own facilities. This leads to the concept of pipeline processing and the establishment of a centralised 'science archive'. The VISTA Data Flow System (VDFS) is a systems engineered project to design and implement a pipeline and archive for WFCAM and VISTA data. The project to develop the VDFS Science Archives – the WFCAM Science Archive (hereafter WSA) and VISTA Science Archive (hereafter VSA) – is to be driven by the science requirements analysis presented here.

This document analyses the externally specified science requirements on the VDFS science archives to act as the starting point in the design process. Since the VDFS design philosophy is to develop a science archive system first for WFCAM, and then to scale the same system for VISTA, in this document the WFCAM requirements are analysed first, and then only those VISTA requirements on scope and functionality not already covered in the WFCAM analysis are considered further (differences in scale between WFCAM and VISTA are discussed extensively in AD01). Any requirements relevant only to pixel–level pipeline processing are not discussed here.

2 OVERVIEW

Standard, top–level analysis for complex digital systems consists of:

1. definition of requirements and specifications,
2. undertaking analysis and design,
3. code development and debugging,
4. unit and integration testing,
5. deployment and maintenance;

this sequence usually being iterative as scope/specifications change and feed back to modify the system requirements.

This science requirements analysis document (SRAD) details the basic requirements for the WSA and VSA, and represents item 1 in the above sequence. The intention is to state the top–level science requirements being placed on the science archives as a whole; analyse science usage examples of the WSA and VSA; and finally to discuss in more detail those requirements pertaining to them in order to produce a specification for their design. The approach taken in this document is to distil the external, top–level requirements and the usage examples, through analysis and implication, to an explicit statement of the science archive contents and functionality. Hence, the SRAD is structured as follows:

- Section 3 references the external requirements;
- Section 4 restates the WFCAM 'top–level', 'contents & functions', and 'security requirements', along with analysis and notes and further refines these above, along with the externally specified 'detailed requirements', to provide a basic list of the detailed requirements
- Section 5 examines new VISTA requirements over and above those already considered for WFCAM;

- Sections 7 and 8 summarise the WSA and VSA contents and functionality in a more concise and explicit way for the perusal of interested parties and to enable archive developers to progress the design.

Subsequently, we follow the sequence above and undertake a design for the archives including documentation of data products, data flow, and hardware and software architecture. The detailed specification for the WSA and VSA is developed further in these following documents rather than in the SRAD.

The WFAU VDFS science archive development homepage is at http://www.roe.ac.uk/˜nch/wfcam.

3 EXTERNAL REQUIREMENTS

3.1 WFCAM

The specified requirements are available online at http://www.jach.hawaii.edu/˜adamson/wfarcrq.html Usage examples are available[5] and have been developed in collaboration with the UKIRT Infrared Deep Sky Survey[3] (UKIDSS) consortium, which constitutes the primary user community for the WSA.

Table 1 summarises the major surveys currently being undertaken with UKIRT/WFCAM. The product of the area, number of filters and typical number of visits per filter gives some idea of the relative scale of the surveys. Note that at the time of writing, the UKIDSS surveys are undergoing a review for continuation after two years of survey operations, and that a northern hemisphere analogue of the VISTA Hemisphere Survey (see the next Section) is being proposed.

Name	Area (sq.deg)	Filters	Depth (K; Vega)	Synoptic?	Visits per field
Large Area[1]	4028	YJ(x2)HK	18.2	No	1
Galactic Plane[1]	1868	JHK(x3)H2(x3)	18.8	No	1
Galactic Clusters[1]	1067	ZYJHK(x2)	18.5	No	1
Deep Extragalactic[1]	35	JHK	20.8	No	14–20
Ultra Deep[1]	0.77	JHK	22.8	No	300–1000
Transit Survey[2]	3.08	J	J $\sim$ 17.0	Yes	$\sim$ 1000

Notes:
[1]UKIDSS Survey definitions taken from[4]
[2]UK PATT application U/06A/52

Table 1: *Major surveys being undertaken with UKIRT/WFCAM. The K band depth quoted is for S/N=5 for the combination of epoch visits to a given field except for the Transit Survey. The fifth column refers to whether the survey requirements specify any time–domain science goals, e.g. variables, transits etc, over and above a small, fixed number of distinct epoch visits.*

3.2 VISTA

Top–level requirements and usage examples are specified in AD01, which is available online at http://www-star.qmw.ac.uk/˜jpe/vdfs/.

Table 2 summarises the major public surveys being considered for implementation on VISTA, correct at the time of writing. Note that the final suite of surveys may consist of an altered set as a process of negotiation between the PIs and the ESO Public Surveys Panel is currently underway.

Name	Area (sq.deg)	Filters	Depth (K; Vega)	Synoptic?	Visits per field
VMC[1]	420	YJKs	20.3 (10σ)	Yes	$\sim$ 12
VUUDS[2]	1	ZYJ	J=25.7 (5σ)	No	400–950
VVV[3]	360	JHKs	18.0 (4σ)	Yes	50–200
VIKING[4]	1500	ZYJHKs	19.5 (5σ)	No	1
VIDEO[5]	30	ZYJHKs	21.7 (5σ)	No	12–150
VHS[6]	20,000	ZYJKs	18.1 (5σ)	No	1
ELVIS[7]	1.6	$\mathrm{JJ_{DARK}}$	$\mathrm{J_{AB}} = 25.5$ (5σ)	No	$\sim$ 500
VGPS[8]	1550	ZYJHKs(x3)	18.4 (5σ)	No	1
ULTRA–VISTA[9]	$\sim$ 1	ZYJHKs	$\mathrm{K_{AB}} = 25.8$ (5σ)	No	$\sim$ 1000

Notes:

[1] VISTA near–infrared survey of the Magellanic System

[2] VISTA–UKIRT Ultra Deep Survey

[3] VISTA Variables in the Via Lactea

[4] VISTA Kilo–Degree Infrared Galaxy Survey

[5] VISTA Deep Extragalactic Observations survey

[6] VISTA Hemisphere Survey

[7] Emission–Line galaxies with VISTA survey

[8] VISTA Galactic Plane Survey

[9] ULTRA–VISTA: Observing Beyond Reionisation

Table 2: *Major surveys being considered for implementation with VISTA. The Ks band depth is that for all combined exposures except for VVV. Note that at the time of writing, reconstituted proposals are being prepared that merge VUUDS and ELVIS into ULTRA–VISTA and VGPS into VVV.*

4 DETAILED REQUIREMENTS ANALYSIS: WFCAM

Note: this section is as originally prepared in Q1 2003, although some nomenclature in management documents now refers to 'phases' rather than 'versions'.

In the following analysis, we discuss the top–level requirements referenced in the previous Section in more detail (requirement labels refer to those used in the documents referenced). Each item has an associated *Rationale*, *Implications* which discuss the implications for the WSA design, an optional *Note*, and finally a concise statement of the requirement to be developed in later Sections. It is intended that the requirements cannot be changed without consultation (primarily with JAC and UKIDSS).

4.1 Top–level requirements

T1:
Science archive shall provide the maximum possible potential for capitalizing on the UKIDSS surveys.
Rationale: UKIDSS will absorb the greater fraction (75%) of *all* WFCAM time on UKIRT and so is the top priority for WSA usage.
Implications: The UKIDSS programme must be the prime science driver for the WSA. Archive development needs to be an open process, with as much UKIDSS involvement as possible. Hence, full and up-to-date documentation needs to be available in web–browsable form as well as hardcopy. The tight schedule for WFCAM, the competition from CFHT WIRCam, and the need for timely release of data for competitive and high–impact science place a correspondingly tight schedule on delivery of the WSA. Resource/time constraints imply a phased approach to WSA development, with a committment to producing a basic working archive system by instrument first light, followed by development to a fully functioning archive system thereafter. To expedite delivery of the WSA, design should be based on existing archive solutions and code where appropriate.
Requirement:
A basic working science archive (hereafter 'Version 1.0') *must* be in place at WFCAM first light. A fully functioning archive system (hereafter 'Version 2.0'), as defined by the requirements herein, must be available as soon as possible after WFCAM first light, and no later than 1 year after survey operations begin in earnest.

T2:
Science Archive must contain and serve *pipeline processed* data (processed pixels, object catalogues and housekeeping data) from both UKIDSS and other usage (e.g. open time, commissioning time).
Rationale: Even small PATT programmes (for example) may produce large amounts of data that are problematic for the user's home institute resources. Moreover, non–survey data will be a valuable datamining resource (see later).
Implications: WSA data accumulation must take into account non–survey usage. Database schema design must be flexible to allow for non–survey data. Proprietary rights need to be protectable in the WSA.
Note: Pipeline processing and subsequent archiving cannot be undertaken for frames taken in non–standard observing modes. For non–survey data that *are* taken in standard modes, limited standardised schemas will be set up and the data will be archived; it will not be possible to develop individual schemas on a case–by–case basis.
Requirement:
Science Archive (all Versions) must contain and serve *pipeline processed* data (pixels, object catalogues and housekeeping data) from both UKIDSS and other usage (e.g. open time, commissioning time).

T3:
Science Archive must be flexible to cope with alterations to UKIDSS survey design over time.
Rationale: The UKIDSS observing allocation and programme are subject to change by the Board on a 2 yearly rolling review.
Implications: WSA design must not preclude changes in design of the major surveys. Again, database design must be sufficiently modular and flexible to cope with this.
Requirement:
Science Archive (all Versions) will match UKIDSS survey requirements as they are currently specified, but will be flexible enough to follow changes in survey design.

T4:
Science Archive design must facilitate usage from 'Grid clients' and inclusion in the Virtual Observatory (VO).
Rationale: Given the legacy aspect of the UKIDSS surveys (especially the LAS and GPS) it is expected that the WSA will form a substantial element in the 'datagrid' of the VO (indeed, WFCAM is a prime science driver in the UK's AstroGrid project).
Implications: WSA access tools, data product formats and transfer protocols must conform to internationally agreed VO standards.
Requirement:
Version 1.0 Science Archive will conform to *existing* standards and will be designed such that new standards can be easily incorporated, but must not be delayed by waiting for new developments to crystalize. Ultimately, the Science Archive must conform to internationally agreed VO standards in access tools, data product formats and transfer protocols.

T5:
Science Archive must allow, for example, *simple* and *complex* queries, with appropriate interfaces.
Rationale: Many users will query the WSA, from the Grid–client 'power user' to the casual, non–expert interactively browsing astronomer. Both are important from the science exploitation point of view.
Implications: Different levels of user interface will be needed for the WSA, from interactive web forms through remote–client GUIs to Grid–enabled clients.
Requirement:
Version 1.0 Science Archive will allow *simple* (see later) queries. Version 2.0 Science Archive will allow usages at varying levels of complexity (as defined later).

T6:
Science Archive must be simple to use for PR purposes.
Rationale: UKIDSS is the next development in the UK's Wide Field programme. High profile science will emerge from UKIDSS, and as the first point of contact with the data, the WSA must be designed appropriately.
Implications: Again, the WSA must be user–friendly to the casual, browsing user. 'Aesthetic' data products (e.g. pseudo–colour images) must be available, in addition to 'serious' science products.
Note: The SDSS has good examples of entry points for PR purposes as well as scientist access points. However, while the production of individual images as a requirement of the WSA, the responsibility of designing and maintaining a 'gallery' website of publicity images lies elsewhere (eg. with JAC and/or UKIDSS).
Requirement:
Science Archive (all Versions) must have interfaces that are open to simple, intuitive use by the non–expert.

T7:
Science Archive must allow access to survey data before all observations are complete, and must not be disrupted by regular ingest of new survey data.
Rationale: Rapid exploitation requires immediate access. The full UKIDSS programme will take up to 6 years or more, and users will want to undertake preliminary analysis after months of data accumulation rather than wait until the full survey datasets are released.
Implications: WSA design must allow for constant data ingest and regular data releases (e.g. interim survey products). WSA must allow for updates to calibrated quantities. WSA must allow for archiving of catalogues from 'reruns' of the processing pipeline, as well as catalogues from previous runs, over pixel datasets in the event of bug fixes and/or enhancements of processing algorithms.
Note: The approach taken with the WFAU's SSS database is to locally mirror the entire released dataset so that two versions are held: a static online version, and another online (but inaccessible from the outside) version for updates. At a release point, the update version becomes the network online version, is copied back to mirror the latest updates, and the whole procedure is so cycled.
Requirement:
Version 1.0 Science Archive must be operable in time for WFCAM first light. Interim survey products must be released to the community on timescales determined by WFCAM observing periods (i.e. a survey 'release' will occur as soon as possible after each observing period, and before the end of the following period).

T8:
Science Archive must allow requests for arithmetic operations, and options from an advanced processing toolkit, on pixel data.
Rationale: Pixel data volumes will be too large for efficient transfer to users home institute for manipulation.
Implications: WSA needs sufficient online storage for pixel data, and sufficient CPU, temporary storage and appropriate software toolkits for pixel manipulation.
Note: The astronomy community in general, and the VDFS pipeline processing centre at CASU for example, are developing pixel processing algorithms. Not all routines will need coding from scratch.
Requirement: Version 2.0 Science Archive must allow requests for arithmetic operations, and options from an advanced processing toolkit (see later), on pixel data. There is no requirement on the Version 1.0 Science Archive to allow this advanced functionality, since we do not anticipate any demand for this immediately after first light.

T9:
Science Archive must be scalable to VISTA data volumes.
Rationale: The WFCAM and VISTA cameras (and science programmes being pursued with them) are similar enough that it makes sense to produce a scalable solution from WFCAM to VISTA for cost effectiveness.
Implications: WSA developments must be open to scrutiny by, and must receive input from, the VISTA project.
Note: VISTA first light is currently scheduled for Q4 2006.
Requirement:
Despite the need to expedite delivery of the WSA, development will be made *at all times* with due regard to scalability to VISTA data volumes.

T10:
Science Archive must be able to merge reduced frames taken in non–photometric conditions with other data from the same survey.
Rationale: Rapid progress may require acceptance of sub–optimal observations in lieu of better, later repeated observations.

Implications: WSA must be able to cope with sub–optimal data and their subsequent displacement by better, repeat observations.
Requirement:
Science Archive (all Versions) must be able to cope with sub–optimal survey observations, and their subsequent displacement by better, repeated observations.

T11:
Science Archive must have some capability for the remote user to carry out data exploration and interaction in real time.
Rationale: The UKIDSS programme contains many instances (e.g. see the specific usage examples) where the remote user will want to manipulate and visualise large amounts of data quickly (i.e. without transfering the large dataset to their own machine).
Implications: Remote client GUI tools (at the most basic level, web browsers) will need to be exploitable for the WSA to enable such interactive data exploration and manipulation. 'Real time' interaction has implications for WSA response time when trawling Tbyte–sized datasets. Clearly, $\sim$ 1000s response time is unacceptable for interactive use, while $\sim$ 10s response time is unrealistic given current technological and financial constraints (such a fast response time may be feasible with a very high degree of parallelism, with consequent complexity and cost implications). For these purposes, a figure of $\sim$ 100s response time seems reasonable.
Note: Of course, for queries on *indexed* quantities (position, image class, brightness and other commonly used attributes), WSA response time will be fast but ultimately limited by factors beyond the control of WFAU (eg. user network connectivity).
Requirement:
Version 2.0 Science Archive must have some capability for the remote user to carry out data exploration and interaction in real time, where 'real time' is understood to mean a timescale of $\sim$ 100s for wholesale trawls. No requirement on Version 1.0 Science Archive system to provide this speed; the ultimate goal should be a response time of $\sim$ 10s.

4.2 Science archive contents and functions (minimum)

C1:
Contains calibrated object catalogues resulting from the pipeline, for both UKIDSS and open–time observations
Rationale: These are obvious, basic science archive functions.
Implications: Database schemas must be set up for various tables of object catalogues. Catalogue ingest software and procedures will be required. Software will be required for 'post–processing' type operations, for example, merging routines and recalibration routines.
Requirement:
Science Archive (all Versions) must contain calibrated object catalogues resulting from the pipeline, for both UKIDSS and open–time observations

C2:
Ingests and stores pipeline output frames for later online processing, generates compressed pixel images on the fly for rapid web–based access, carries out immediate cross–referencing with existing UKIDSS survey data and produces consolidated UKIDSS catalogue in a given field
Rationale: Again, basic science archive functionality.
Implicattions: Database schemas must be designed to track between object catalogue tables and pixel data files. Pixel manipulation software will be required.
Requirement:
Science Archive (all Versions) must ingest and store pipeline output frames, allow rapid web–based

access to images, and produce merged UKIDSS catalogues in a given field.

C3:
Is able to recalibrate a given field or fields in the event of revised calibration information (specifically, photometric and astrometric), and allow database queries on the recalibrated quantities
Rationale: Changes in calibration information are frequently encountered in survey operations, and the science archive itself may lead to such changes.
Implications: Database schema must allow provision for recalibration – e.g. stores positions as pixel co-ordinates plus and astrometric solution (consisting of specified model and coefficients); stores photometry as flux measures plus calibration data. Calibrated quantities will also be required to be stored in tables, since inverting calibration models to translate queries in calibrated units to uncalibrated ones will be difficult in general. The archive must be able to replace calibrated quantities when new ones become available. Calibration version control within the archive is required.
Requirement:
Science Archive must be designed from the start to enable astrometric and photometric recalibration.

C4:
Is able to cross-calibrate photometric information using areas of overlap between processed frames, where available.
Rationale: This is not a sensible function of the pipeline, which is required only to produce results on a night–by–night basis. The science archive will have all photometric information and calibrations for all frames, and is where cross–calibration should happen.
Implications: Calibration tools will be required to homogenise photometry over surveyed areas using overlap information and photometric zeropoints.
Requirement:
Version 2.0 Science Archive must be able to cross-calibrate using areas of overlap between processed frames, where available (no requirement on Version 1.0 Science Archive to cross–calibrate).

C5:
Allows public access to subsets of survey data on a variety of different search criteria (specified below)
Rationale: Basic science archive functionality.
Implications: For versatility, SQL–like querying is required, even if this is transparent to the user (e.g. simple access via web–form interface).
Requirement:
Science Archive (all Versions) must be designed to allow public access to subsets of survey data on a variety of different search criteria (specified later).

C6:
Allows rapid on–line cross–referencing of search results with other catalogues.
Rationale: consistent with T1, this requirement is expanded on later.
Implications: The Science Archive must undertake to store commonly used catalogues locally for combination queries in a queryable database.
Requirement:
Science Archive (all Versions) must have available *commonly used catalogues* (see later) stored locally. Version 2.0 Science Archive may additionally hold SDSS (and other survey) pixel data for joint querying (see later).

C7:
Allows generation of finder charts via a web form
Rationale: Simple to provide and useful when observing at a site remote from the UK.
Implications: Software will be required for generation of pixel and/or ellipse plot finder charts. A web

form will be required as the user interface.
Requirement:
Science Archive (all Versions) must allow generation of finder charts via a web form.

C8:
Holds housekeeping information for all archived data.
Rationale: It is essential to propagate all available data description (e.g. FITS header data) through to the Science Archive, to enable users to query those data
Implications: The Science Archive must be able to track between object catalogue records, image data files and the housekeeping data. For example, to protect proprietary data rights the Science Archive will need to validate queries against the source of any particular image subset (e.g. UKIDSS, PATT time, etc.)
Requirement:
Science Archive (all Versions) must hold housekeeping information for all archived data.

4.3 Security

A1:
Archived *data* must be accessible only by validated users
Rationale: The WSA will contain data resulting from internationally competitive science proposals. Proprietary rights of the UKIDSS consortium and open–time PIs/CoIs must not be compromised by data being freely available through the online archive.
Implications: The Science Archive must have security systems in place that prevent unfettered access by opportunistic users, but at the same time must not become so protected that access by valid users is hampered (e.g. by constantly asking for usernames/passwords). Security systems must be able to cope with various proprietary periods, and allow unfettered access after appropriate time intervals. All of this in turn implies user registration with username/password login and/or 'digital certification'.
Note: *Any* user (not just proprietors) should be able to derive information on what is in the archive without being given access to those data.
Requirement:
Science Archive data (all Versions) must be accessible only by validated users; archive *content* information should be available without restrictions.

A2:
Archived data must be uncorruptable by Science Archive users.
Rationale: Scientific exploitation will be compromised if data are corrupted.
Implications: Constant data ingest, recalibration of photometry/astrometry, and functionality enhancements imply a 'living' archive that is subject to change. This opens up the possibility of accidental corruption, especially by local archive managers with read/write access to filesystems. Archive design must minimise the possibility of accidental corruption, and also insure against data loss and minimise reconstruction times by invoking an appropriate backup policy.
Requirement:
Science Archive (all Versions) must be uncorruptable by Science Archive users.

A3:
Science Archive must allow data protection on the basis of proprietary data (per frame)
Rationale: Proprietary periods will be different for different observations (survey/non–survey).
Implications: Security systems must be able to cope with various proprietary periods, and allow unfettered access after appropriate time intervals.

Requirement:
Science Archive (all Versions) must allow data protection on the basis of proprietary data (per frame)

A4:
Science Archive must be quickly recoverable in the event of corruption by hardware/software faults etc.
Rationale: Clear need to ensure against data loss.
Implications: Science Archive will require backup on removable media and/or 100% redundant storage with data striping (i.e. fault tolerant hardware/software).
Requirement:
Science Archive (all Versions) must be quickly recoverable in the event of corruption by hardware or software faults etc.

4.4 Detailed requirements

The following requirements form the baseline for the WSA; they are an expansion of the top–level requirements above and items D in the 'Detailed Requirements'. Following T1 above, we have divided the requirements into those that must be in place for WFCAM first light and those that need fulfilling after a significant amount of data have accumulated. There are several reasons for this: i) the timescale for the delivery of WFCAM is short, so there is limited time for R&D concerning a large, scalable archive system; ii) such a system is not required at first light anyway since data volumes will be of limited size initially; iii) a phased approach means that the final large hardware purchase can be delayed as long as possible. So, we have grouped these into 'Version 1.0 requirements', and 'Version 2.0 requirements'; some requirements appear in the earlier version with limited scope, and in the later versions with full–blown functionality. We include some more long–term goals which may or may not be delivered, contingent on implementation and resource constraints, and delivery of appropriate tools/knowledge from related e–science projects (e.g. Astrogrid).

4.4.1 Version 1.0 requirements

T1/T7: The 'Version 1.0' working science archive *must* be in place in time for WFCAM first light.
T2: Science Archive must contain and serve pipeline processed data (pixels, object catalogues and housekeeping data) from both UKIDSS and other usage (e.g. open time, commissioning time).
T3: Science Archive will match UKIDSS survey requirements as they are currently specified, but will be flexible enough to follow changes in survey design.
T4: Science Archive will conform to any *existing* 'Virtual Observatory' standards and will be designed such that new standards can be easily incorporated, but must not be delayed by waiting for new developments to crystalize.
T5: Science Archive will allow *simple* (see below) queries.
T6: Science Archive must have an interface that is open to simple, intuitive use by the non–expert.
T9: Despite the need to expedite delivery of the WSA, development will be made *at all times* with due regard to scalability to VISTA data volumes.
T10: Science Archive must be able to cope with sub–optimal observations, and their subsequent displacement by better, repeated observations.
C1: Science Archive must contain calibrated object catalogues resulting from the pipeline, for both UKIDSS and open–time observations
C2: Science Archive must ingest and store pipeline output frames, allow rapid web–based access to images, and produce merged UKIDSS catalogues in a given field.
C3: Science Archive must be designed from the start to enable astrometric and photometric recalibration.

C5: Science Archive must be designed to allow public access to subsets of survey data on a variety of different search criteria (specified below).
C6: Science Archive must have available *commonly used catalogues* (see later) stored locally.
C7: Science Archive must allow generation of finder charts via a web form.
C8: Science Archive must hold housekeeping information for all archived data.
D1: Science archive must allow searching individual (or all) UKIDSS surveys on the following criteria (or combination of them):

- Position rectangle expressed in spherical co-ordinates: RA/Dec (J2000); l, b (Galactic) and λ, η (SDSS system)
- Circular sky patch within specified radius from given spherical co-ordinates: RA/Dec (J2000); l, b (Galactic) and λ, η (SDSS system)
- Circular sky patch within specified radius of a resolvable source name

D3: Science Archive must allow similar queries to be repeated for all objects in a user–supplied source catalogue.
D4: Science Archive must allow combinations of queries on UKIDSS data and the following other source catalogues:

- 2MASS
- SuperCOSMOS Sky Survey
- SDSS DR1 data release
- USNO–B
- FIRST source catalogue
- IRAS point source catalogue
- ROSAT All–Sky Survey catalogue

D6: Science Archive must have a simple interface for very quick searching on a given object name or position.
D8: Science Archive must return pixel images, confidence maps and catalogue data in gzipped FITS format, and must allow users to specify the output format of returned data as follows:

- FITS images with options for lossless and/or lossy compression
- ASCII (tab or comma–separated) or FITS table, for object catalogue and housekeeping data
- Space–separated ASCII with CDS–type descriptors for object catalogues
- VOTable – http://vizier.u-strasbg.fr/doc/VOTable – a proposed protocol for exchange of astronomical data embedded in XML.

D9: Science Archive must be able to return pixel data in any available passband, over a contiguous field up to one 'tile' (0.8°) across together with a matched catalogue.
D11: Science Archive must be able to generate and return stacked images given a user–selected list of input images and the standard stacking algorithm in the CASU basic pipeline.
D12: Science Archive must be able to generate and return merged multi–colour, multi–parameter catalogues with the best available photometric and astrometric calibrations.

D13: Science Archive must support federation with the source catalogues specified in D4 above
D14: Science Archive must be able to generate and return meaningful optical/IR colours for all objects in the overlap with the existing SDSS data where counterpart detections occur in the SDSS object catalogue.
D16: Science Archive must support the returning of only a subset of the entire possible array of object parameters.
D19: Science Archive must be able to produce a finder chart of size up to 10 arcmin for any region within which survey data exist, returning ellipse detection plot and/or a single colour pixel plot, as specified by the user.
D20: Science Archive must allow access to best or duplicate data for objects in overlapping survey data.
D21: Science Archive must allow general access to all housekeeping data – e.g. for a given survey area, what is currently available, how good it is, etc.
D22: Science Archive must store uncalibrated quantities, calibrated quantities and the calibration model/coefficients. Archive output must therefore include (in headers)

- Archive version identifier
- calibration version identifier

D23: Science Archive must allow a summary of data available to be generated for a given search region.
A1: Science Archive must be accessible only by validated users.
A2: Science Archive must be uncorruptable by Science Archive users.
A3: Science Archive must allow data protection on the basis of proprietary data (per frame).
A4: Science Archive must be quickly recoverable in the event of corruption by hardware/software faults etc.

User access is to be through web forms providing fill–in boxes and button clicks, and also via an SQL query form interface; a command–line interface for remote users to bypass interactive webforms will also be provided.

The summary in Section 7 gives an explicit statement of the Version 1.0 WSA contents and functionality.

4.4.2 Version 2.0 requirements

In addition to the Version 1.0 requirements:

T1: A fully functioning archive system, as defined by the requirements (and where possible, goals) herein, must be available as soon as possible after WFCAM first light, and no later than 1 year after survey operations begin in earnest.
T4: Science Archive must eventually conform to internationally agreed VO standards in access tools, data product formats and transfer protocols.
T5: Science Archive will allow usages at varying levels of complexity (as defined later).
T7: Interim survey products must be released to the community on timescales determined by WFCAM observing periods (i.e. a survey ‘release’ will occur as soon as possible after each observing period, and before the end of the following period).
T8: Science Archive must allow requests for arithmetic operations, and options from an advanced processing toolkit (see later), on pixel data.
T9: WSA solution must be scalable to VISTA data volumes.
T11: Science Archive must have some capability for the remote user to carry out data exploration and interaction in real time: the Science Archive response time should be $\sim$ 100s for wholesale trawl–type

queries.
C4: Science Archive must be able to cross-calibrate photometric information using areas of overlap between processed frames, where available.
C6: Science Archive must have the final SDSS catalogues (and, if possible, images) stored locally, in addition to the catalogues specified for the Version 1.0 Science Archive.
D1: Science Archive must allow searching individual (or all) UKIDSS surveys on the following criteria (or combination of them):

- Search positions specified at arbitrary equinox and time system, and additionally ecliptic and super–Galactic systems
- Source colour in any linear combination of those colours available for any given survey
- Source parameter ranges

D2: Science Archive must allow searching within open–time programme data using the same criteria as D1 (where possible), returning whatever data are available.
D4: Science Archive must allow combinations of queries on UKIDSS data and the following other source catalogues:

- most recent SDSS data release, as per availability
- User supplied catalogue for complementary imaging (at any wavelength) for any of the UKIDSS sub–surveys.
- any general user–supplied catalogue at any wavelength (eg. GLIMPSE, ASTRO–F)

D5: Science Archive must allow arithmetic functions to be used in setting up complex queries (e.g. for a colour index not stored in survey catalogue tables)
D6: Science Archive must have a remote GUI application for formulating queries (e.g. an interface analogous to the SDSS Java–based query tool).
D7: Science Archive access GUI must allow plotting of returned parameters, in selected (X,Y) pairs or histograms, and also provide basic fitting routines.
D10: Science Archive must be able to generate (on–the–fly) and return larger (than D9) areas from survey data traversing survey tile boundaries, blocked down as specified by the user, in formats specified in D8.
D11: Science Archive must be able to generate and return stacked images using user–specified (see later) stacking algorithm options.
D12: Science Archive must be able to generate and return merged multi–colour, multi–parameter catalogues with the best (or previous as specified by the user) photometric and astrometric calibrations.
D13: Science Archive must support federation with the source catalogues specified in D4 above
D14: Science Archive must be able to generate and return meaningful optical/IR colours for all objects in the overlap with the SDSS, whether or not detected in the SDSS data (i.e. it must be possible to place an aperture in and measure the flux from SDSS image data given the position of an IR source detection).
D15: Science Archive must support ANDing of one query with another, where both have already been executed.
D17: Science Archive must allow trial–and–error searches (e.g. return the number of source hits rather than the output results), for any valid query
D18: Science Archive must allow repetition of queries using previous versions of astrometric and photometric calibrations.
D19: Science Archive must be able to produce a finder chart for any region within which survey data

exist, returning a colour pixel plot, as specified by the user, generated from available single–passband images of the same field.
D20: Science Archive must allow access to best or duplicate data for objects in overlapping survey data, and must contain proper motion measures for objects where multi–epoch position measurements exist.

Section 7 gives an explicit statement of the Version 2.0 WSA contents and functionality.

4.4.3 Goals

T11: Science Archive response time should be ∼ 10sec for wholesale trawl–type querying.
C6: Science Archive will, insofar as external developments allow, be integrated into the 'Virtual Observatory' (VO) as a general solution to rapid, online cross–referencing with any published astronomical catalogues that are also contained within the VO.
D1: Science Archive may recast web services as 'Grid services' (a Grid–based solution to user access) in collaboration with AstroGrid.
D4/13: Science Archive may allow combinatorial queries with catalogues anywhere on the 'data–Grid', i.e. may allow database federation across the grid.
D7: Science Archive will aspire to the mantra 'ship the results, not the data', i.e. may allow remote procedure calls to advanced manipulation tools and may allow user upload of analysis codes.
D10/11: Science Archive may ultimately support advanced visualisation tools, e.g. large area, panoramic pseudo–colour images with panning in real time; three–dimensional catalogue parameter plotting and rotation.

5 DETAILED REQUIREMENTS ANALYSIS: VISTA

AD01 discusses the requirements on VDFS as a whole in terms of the following broad headings: general; astrometric; photometric; tiling, stacking, mosaicing; variable objects; object catalogues; and finally science usage examples. For brevity, and following the VDFS design philosophy of prototyping for WFCAM and scaling to VISTA, in what follows we simply note where AD01 requirements are already covered without further discussion (indeed, in some cases, requirements in AD01 are a restatement of those previously specified for WFCAM). Not all requirements specified in AD01 are relevant to the science archive, and as such do not appear in this analysis. Note that labels in bold face refer to requirements and reference numbers in AD01.

5.1 General Requirements (AD01 Section 5)

5.1.1 Archive requirements already covered by those for the WSA

The following requirements have been already discussed previously in the context of the WSA (Section 4 and references therein; relevant WSA requirement labels are noted in each case):

AD01 5.1 The VDFS shall process all VISTA science data. . . : T2
AD01 5.2 The release, and use . . . shall take account of any proprietary periods . . . : T2
AD01 5.3 The data output by the VDFS science archive . . . shall be . . . compliant with the Virtual Observatory . . . : T4, D8
AD01 5.5a For reduced data, summary information . . . shall be written . . . : C8
AD01 5.7 The VDFS shall provide means of tracking survey . . . progress . . . : D21
AD01 5.8 The VDFS shall provide "summary statistics" . . . : C8, D21
AD01 5.9 There shall be "query–driven" access to the VDFS Archive . . . : T5 and additional details

as follows: a:T6; b:D1; c:D5; d:D1; e:D3; f:D4 (but note additional external surveys); g:D16
AD01 5.11 The VDFS Archive shall have capability . . . for arithmetic image processing: T8, D11
AD01 5.12 The VDFS Archive shall comprise both a "living" internal database . . . and a small number of data releases . . . : T7, A3
AD01 5.15a The archive shall be capable of limiting access . . . : A1
AD01 5.15b Access shall allow for any proprietary periods: A3
AD01 5.15c . . . it shall be impossible for Archive users to corrupt the data or overload . . . : A2
AD01 5.17 The goals for VDFS Archive response time . . . : T11

5.1.2 Archive requirements not already covered

Following the analysis presented previously for WFCAM requirements:

AD01 5.4:
The VDFS shall be able to . . . archive data at a peak rate of 1000 GB/day for 10 days, and a sustained mean rate of 650 GB/day . . . long term . . .
Rationale: Archive ingest must keep pace with the data flow rate upstream in the system.
Implications: Network connection with the processing centre and archive ingest must have sufficient bandwidth to keep pace with these numbers.
Note: These figures specified in AD01 are 'worst–case' since they allow for retransfer/reingest of reprocessed data, and do **not** include reduction factors for lossless compression. Experience with WFCAM indicates that 4–byte integer IR data compressess by a factor of up to 4×. Note also that catalogue and image metadata are typically a few percent of the size of the uncompressed, processed pixel data.
Requirement:
The Science Archive (all Versions) must be capable of transfer/ingest of pixel data at a peak rate of 1TB/day (uncompressed) and ingest of catalogue and image metadata at a peak rate of around 20 GB/day

AD01 5.6:
The VDFS shall function correctly in the event of . . . non–operational or missing detectors . . .
Rationale: Given the number of detectors in the tiled focal plane, there is an increased chance of one or more detector failures but a continuation of operations in spite of this.
Implications: no archive ingest, curation or user interface application can assume presence of all detectors, all of the time.
Requirement:
No Science Archive application (ingest, curation or user interface in any version) may be dependent on all detectors being present all of the time.

AD01 5.9h:
Query returns shall be able to included a user–selected random sampled percentage . . .
Rationale: General sample selection (rare object searches for example) may require an iterative usage mode where the user refines the completeness/contamination trade–off in a query to obtain a viable sample with which to work.
Implications: The Science Archive must provide a flexible querying facility that includes this feature.
Note: Given a uniformly distributed attribute that is not correlated with any science attribute, for example a database unique ID key, this is straightforward in SQL.
Requirement:
The Science Archive (all versions) should provide a query facility capable of returning a user–selectable percentage sampling for a given selection to aid in optimising sample completeness/contamination at

query time.

AD01 5.10:
When selecting object by colour, it shall be possible to select using a cut on time–lag between observations in different filters.
Rationale: To avoid contamination by photometric variables.
Implications: The Science Archive must have a data model that tracks the required metadata, and a flexible querying facility that enables sample selection predicated on any image metadata attributes.
Note: The WSA Questions and Answers page at http://surveys.roe.ac.uk/wsa/pre/qa.html#catobsdates demonstrates how to do this employing the WSA data model with an appropriate SQL query.
Requirement:
Science Archive query selection (all versions) shall facilitate predicates on metadata attributes as well as catalogue attributes (e.g. maximum observation date difference between filters when querying on source colours).

AD01 5.13:
Once released, each data release shall remain indefinitely available, and the image and catalogue data shall not be modified.
Rationale: Continued access to old versions is required for cross–checking results.
Implications: storage requirements must be estimated for the data volumes likely to be needed. If large amounts of reprocessing are deemed necessary, then it could be that all pixel processing versions cannot be retained; similarly, it may not be possible to retain all released database products. The pragmatic policy should be to discard the oldest versions if storage space needs to be recovered.
Requirement:
Science Archive (all versions) shall retain each catalogue database product and pixel processed dataset indefinitely, insofar as storage restrictions allow.

AD01 5.14:
To facilitate follow-up spectroscopy, it shall be possible for a user to upload a list of target positions and reference star search radius, then the Archive shall return the update target position (based on the latest astrometric solution) and a list of reference stars within the search radius.
Rationale: Fibre spectroscopy in particular requires targets and fiducials on a single, precise and uniform astrometric system.
Note: This is more of a specific usage example than a general requirement.
Requirement:
The Science Archive shall provide updated astrometric information to the user in an easy and flexible manner.

AD01 5.15d:
Releases will require prior authorisation by the VISTA PI to ensure they are agreed by the necessary supervisory bodies.
Rationale: To ensure controlled release of data at appropriate times and to appropriate user communities.
Requirement:
Each Science Archive release shall be authorised by the VISTA PI.

Finally, we note that AD01 5.16 specifies the goals for VDFS Archive uptime, but no requirement is specified, reflecting financial constraints and the fact that many factors that are beyond the control of WFAU may impact archive availability (e.g. network infrastructure downtime outwith the immediate control of WFAU).

5.2 Astrometric Requirements (AD01 Section 6)

The astrometric requirements are mainly relevant to pipeline processing in VDFS; however any archive–end astrometric recalibration must note the following:
Requirement:
Absolute astrometric accuracy is required to be better than 0.3 arcsec; differential astrometric accuracy is required to be better than 0.1 arcsec within a tile and better than 0.03 arcsec within the area of sky covered by a single detector.

5.3 Photometric Requirements (AD01 Section 7)

AD01 7.1:
Absolute photometric accuracy ≤ 0.02 mag in J, H, Ks bands . . .
First–pass photometric calibration with respect to 2MASS is provided in the nightly processing pipeline. It is likely that for the filters in common with 2MASS (JHK) and in fields free from heavy extinction,the nightly pipeline will produce the photometric accuracy required (indeed, experience operating the VDFS with WFCAM data has shown this to be the case). However, for other filters, for fields with moderate to high extinction, and for the purposes of approaching the goal of absolute JHKs photometric precision of ≤ 1% (AD01 7.2) and the requirement on ZY precision of ≤ 3% (AD01 7.3), photometric recalibration at the archive end may be necessary, employing more constraint information (e.g. that from overlap regions and/or secondary photometric standards).
Requirement:
Any archive end photometric recalibration procedure must deliver absolute photometric precision of ≤ 2% in JHK and ≤ 3% in ZY.

5.4 Tiling, Stacking, Mosaicing Requirements (AD01 Section 8)

5.4.1 Archive requirements already covered by those for the WSA

AD01 Section 8 mainly concerns pixel processing pipeline functionality.

5.4.2 Archive requirements not already covered

AD01 8.3:
The image stacking shall provide a choice of options for pixel interpolation, including . . .
We note that the list of options (a to e) specified here may be required to be implemented at the archive end, provided the requisite toolkit codes are available for a given option.

5.5 Variable Objects (AD01 Section 9)

AD01 Section 9 specified that basic source variability should be considered, with a set of four example scenarios. It is a fundamental requirement on the VSA that sample searches for variable objects, or conversely samples free from contamination by variable objects, are to be possible.

5.6 Object Catalogues (AD01 Section 10)

5.6.1 Archive requirements already covered by those for the WSA

AD01 10.1: The catalogues shall include the . . . calibrations described . . . : C1, D12
AD01 10.2: The catalogued object parameters shall include . . . the list as specified . . . : C1
AD01 10.3: There shall be the following options . . . in multiple bands . . . : C1, C2, D12 (for b)

5.6.2 Archive requirements not already covered

AD01 10.4:
It shall be possible to identify solar system objects . . .
Rationale: Fast moving objects appear at different positions in non–contemporaneously observed passbands, potentially giving rise to unpaired objects in merged source lists that can be mistaken for very rare (and highly sought after) objects of extreme colour.
Requirement:
Fast moving solar system objects must be detected and flagged as such in final source catalogues to prevent them from contaminating user–selected samples of, for example, extreme colour.

AD01 10.5
Known asteroids which appear in the final VDFS catalogues shall be flagged as such
Rationale: Again, to prevent confusion when selecting samples of rare objects
Requirement:
Known asteroids which appear in the final VDFS catalogues shall be flagged

AD01 10.10:
VDFS shall provide a simple completeness estimate for each tile . . .
Rationale: To aid completeness estimates and corrections when analysing the statistics of selected samples. We note also the goal AD01 10.11.
Requirement:
VDFS shall provide a simple completeness estimate for each tile

AD01 10.12
VDFS shall be able to generate basic variability data by choosing a time spacing . . . group a series of frames . . . into 'epochs' . . . and then reduce . . . to the level of one catalogue per epoch.
Rationale: Such archive–end functionality will be useful for the VVV and VMC Surveys (Section 3.2). We note also the goal AD01 10.13.
Requirement:
VDFS shall be able to generate variability data from grouped and stacked 'epoch' imaging data via source catalogue extraction from those epoch stacks.

5.7 Science Examples (AD01 Section 11)

The science examples covered in AD01 Section 11 overlap with those already considered for WFCAM[5] as follows (AD01 examples listed first in each case): 11.1 ≡ U2; 11.2 ≡ U6; 11.4 ≡ U8; 11.5 ≡ U8 along with external catalogue join; 11.6 ≡ U5, U13 & U19; 11.7 ≡ U19; 11.8 ≡ U2 & U4; 11.9 ≡ U2; 11.10 ≡ U2; 11.11 ≡ U4; 11.12 ≡ U19; 11.13 ≡ U19; 11.14 ≡ U19; 11.15 ≡ U3.

AD01 11.3:
Select objects in VISTA Deep survey at J–band using Petrosian magnitudes . . . :
Rationale: Similar to several WSA usage modes, but with the additional requirement of provision of a completeness map, e.g. estimated detection probability at a given magnitude limit averaged in 1 arcmin pixels across the survey area.
Requirement:
The VSA shall have some means of provision of estimated detection probability as a function of magnitue.

AD01 11.16:
Provide luminosity function for Galactic cluster . . . :
Rationale: Again, the provision of data for the compution of an empirical LF is similar to several existing WSA usage modes, but the provision of a completeness function is not.

Requirement:
The VSA shall have some means of provision of estimated completeness as a function of magnitue.

AD01 11.17:
. . . i) take a list of LSB galaxies detected in a visible survey . . . ii) Detection in NIR. . . . :
Rationale: AD01 notes that such a specialist application may require upload of user code to the archive to analyse pixel data.
Requirement:
The VSA shall have a facility for upload of user–supplied pixel analysis code.

Finally, we note that AD01 usage examples 11.18 and 11.19 are outwith the scope of the VDFS/VSA.

6 QUALITY CONTROL

For both WFCAM and VISTA, the broad area of quality control is refered to obliquely in the external requirements and their analysis previously. Experience with WFCAM/UKIDSS has shown, moreover, that quality control is a major issue (in fact, one of the most important issues) in preparing survey releases. It is therefore vital that all aspects of the VDFS Science Archives allow provision for open–ended and flexible quality control procedures as required by the respective survey scientists involved in defining and implementing the survey release products.

Without pre-empting the design presented in subsequent documents, we note that:

- the use of a relational database management system is fundamental to flexibility;
- provision of a data model that allows quality control flagging is essential;
- provision of sufficient local operations effort to support survey consortium quality control is necessary;
- user interface flexibility in allowing and presenting data selections of unreleased data is essential to facilitate quality control.

These items are expanded on in the design documentation, most notably AD02.

7 WFCAM SCIENCE ARCHIVE SUMMARY

At its meeting on 2002 November 25, the UKIDSS Consortium discussed the requirements and usages along with the WSA development plan. The Consortium suggested several changes along with some issues for discussion. The results of these discussions were folded into this document, yielding the following specification (in as much detail as is possible at this time) for the WSA functionality and contents at Versions 1.0 and 2.0 (note: this specification will be developed in later documents). The V2.0 requirements can be considered 'goals' of V1.0.

7.1 Version 1.0

WSA Version 1.0 is deliverable at WFCAM first light. In addition to the following, WFAU undertakes to apply UKIDSS–specified algorithms, and import UKIDSS–supplied catalogues, to the WSA in lieu of automatic tools for such functionality (see Version 2.0).

7.1.1 Contents

The V1.0 WSA 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. Catalogues for 2MASS, SDSS DR1, SSS, USNO–B, FIRST, IRAS and ROSAT all–sky surveys;
6. A *Survey Progress Catalogue*, containing for each of the 5 UKIDSS subsurveys information on observations taken to date;

and also image data (pixels with confidence maps; default stacks for the deep surveys; and difference images for the GPS K band) in flat files, along with a large reserve (scratch) workspace for use during querying. The V1.0 WSA will also contain online documentation and 'cookbook' style worked examples to aid users.

7.1.2 Functionality

The V1.0 WSA 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. For an example of such an interface, see WFAU's SuperCOSMOS Sky Survey access page http://www-wfau.roe.ac.uk/sss, particularly the 'Get a CATALOGUE' interface.

2. A web form interface allowing *querying* of individual (or all) WSA 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:

   - ASCII (space, tab or comma–separated);

- FITS binary tables;
- VOTable format;
- uncompressed or lossless compression (e.g. gzip);

combinatorial queries with the 2MASS, SSS, SDSS–DR1 and USNO–B catalogues will be provided for. For an example of such an SQL interface, see WFAU's 6dF access interface at URL http://www-wfau.roe.ac.uk/6dFGS/SQL.html.

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° (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 (supplied as FITS binary extension).

    For an example of such an interface, see WFAU's SSS page (URL above), particularly the 'Get an IMAGE' facility.

'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.

7.2 Version 2.0

Version 2.0 is deliverable no later than one year after survey operations begin, and will include more 'database driven' products and features. In addition to contents and functionality provided in V1.0, the following specifies the V2.0 contents and functionality.

7.2.1 Contents

The V2.0 WSA will additionally contain:

1. Externally provided catalogues and pixel data (UKIDSS complementary imaging, and SDSS data release as avialable at that time);
2. A database of open–time observations;
3. 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.

7.2.2 Functionality

In addition to the simple access tools provided in V1.0, one (or more) advanced GUI(s) will be provided that have the following functionality:

1. 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; etc.

2. Arbitrary sized, mosaiced images (across tile boundaries), blocked down as appropriate, with a multi–colour option;
3. Source extraction options on any specified subset or bespoke stack of pixel data: i) CASU standard source extraction; ii) SExtractor; iii) mutiple simultaneous profile fitting (i.e. DAOphot–like); etc.
4. Data exploration/interaction facilities: simple XY plotting; histogram plotting; simple model fitting routines (generalised least–squares with robust outlier rejection);
5. Automatic user–supplied catalogue ingest facility for joint querying with existing catalogues;
6. Enhanced output format options to include any new Virtual Observatory standards available at that time;
7. 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 (ie. list–driven photometry for *any* data);
8. Generalised difference imaging (and subsequent source analysis)
9. Persistence of multi–stage usage/query; storage of intermediate user–generated results sets

Additionally, the web–based access tools in V1.0 will be supplemented with a 'web service' interface (eg. a non–interactive access tool employing XML format data transfered using Simple Object Access Protocol) to provide, where appropriate, non–interactive access to pixel and catalogue data. Archive response time is to be $\sim$ 100s for wholesale catalogue trawls on non–indexed quantities.

7.3 Later Versions

Later versions will be developed in tandem with further development for the VISTA Science Archive (see below).

8 VISTA SCIENCE ARCHIVE SUMMARY

Owing to a slip in the delivery of WFCAM and subsequent start of survey operations, the originally planned archive versions (Section 7) were recast in the most recent externally review VDFS resource/planning document[6]. Some of the functionality described previously in specific archive versions was moved forward while some was implemented ahead of schedule. The resulting recast overall VDFS Science Archive versioning is as follows:

- VDFS–v1 (also known as Phase 1 for WSA V1.0): largely as detailed in Section 7.1;
- VDFS–v2 (also known as Phase 2 for WSA V2.0): incorporating WSA V1.0 plus as described in Section 7.2; note however that open time non–survey data access was moved back to WSA V1.0 while VO functionality was mover forward to (a new) WSA V3.0;
- VDFS–v3 (also known as Phase 3 for WSA V3.0): a new version incorporating WSA V2.0 plus new functionality as follows:
  - advanced server–based visualisation tools, e.g. pannable large–area imaging, multi–dimensional catalogue plotting and rotation
  - server–based data analysis tools, e.g. cluster analysis, PCA etc.

  - a system to allow uploadable user–specified analysis algorithms
  - recast existing web services as grid services in order to allow queries and analysis using shared managed resources with other data centres

- VDFS–v4 (also known as Phase 4 for VSA V1.0): as for WSA V3.0 except handles the much greater VISTA data rates — this version to be ready for VISTA first light;

- VDFS–v5 (also known as Phase 5 for VSA V2.0): incorporates changes from experience with handling real VISTA data along with completion of integration into the VO.

8.1 Additional requirements from VISTA Public Surveys

As part of on–going review of user requirements for the VISTA phase of the VDFS Science Archive project, those PIs proposing "public surveys" with VISTA were contacted recently to get further input on archive usage modes that may not be covered already in the existing WFCAM system. Generally speaking, the community seems to view the existing WSA as being able to cover most (or in some cases, all) of their baseline requirements in terms of archive contents and functionality. The main extension required for VISTA from WFCAM science programmes is an increased emphasis in time–domain applications, with more synoptic surveys being proposed for VISTA. Otherwise, at the time of writing, the following suggestions for additional contents/functionalities have been received from the respective PIs:

- VMC:

  - more explicit use of bit–wise quality flagging, e.g. sources affected by dead pixels, proximity of bright neighbours, etc.;
  - more flexible catalogue products from both pawprints, tiles and combinations of epochs (where those exist);
  - inclusion of the DENIS southern infrared survey as an external catalogue joined to VISTA catalogues.

- VVV:

  - Extract from the archive the unphased light curve of any source (or list of sources), in all bands, with errors; make a light curve plot.
  - Convenient interface for cross correlation with 2MASS, GLIMPSE(2), DSS, etc. or user supplied catalog to obtaina a combined cross-catalog with colors and proper motions.
  - Find variables via image subtraction technique (or aperture photometry in a more pedestrian case). This requires an ability to define template image – the image of the area with the highest quality. The next step is matching the image quality of all availaible epoch images to identical level, via convolution. Finally, the template is subtracted from each image and the point source on the residual image are detected. The products are: a catalog of variables with their respective unphased light curve (with attached errors to each measurement), minimum amplitudes, mean and median magnitudes, r.m.s., second and third moments of the distribution of the measurements (sharpness and assymetry of the distribution; these are useful for separating difference classes of variables).
  - Period determination of variable sources: using the two basic methods FFT and PPM. The limits within which the period is explored will be determined from: (upper limit) the length of time spanned by the available observations. and (lower limit) the Nyquist sampling. This implies that the period range is from 0.1 to 100 days. The producsts are: periods with uncertainties, confidence estimates of the periods, phased light curves, Fourier moments of the light curves.

  - Variable classification based on pre-defined constraints on: colors, magnitudes, periods and amplitudes. The products are catalogs of various types of variable sources.
  - statistical studies of variable sources for stellar population and Galactic structure studies. These are also useful for detection of diffuse light echoes from novae and supernovae - these are found by looking for clustering of varibale sources in time and space as the light echo crosses the line of sight towards the sources. Product: numbmer density of variable sources of different types per bin with user-specified bin size, i.e. from a few arcmin to one degree.
  - Reddening estimate to every individual variable source based on the average color of a pre-defined feature on the Color-Magnitude Diagram, for example the Red Clump Giants, within a circle with a given radii (i.e. 30 arcsec). This can be generalized for any object or a list of objects. The product is reddening estimate, and it is included in the master list of all variable sources.
  - Incorporation in the archive of references to papers with additional information, comments, and data (for example more precise periods, optical colors, X-ray measurements, etc.)
  - Multi-epoch proper motion tools – find moving sources via image subtraction technique (or PSF fitting or centroid calculation) with proper motions above 5 times their estimated proper motion uncertainties. Products: catalog of moving sources with all entries plus the reduced proper motion; – For each epoch find objects with 3-sigma (or better) detection that have no analogs in the other epochs; produce a merged catalog for all epochs (useful for novae, supernovae, microlenses, near Earth asteroids, etc.). Product: catalog of transient sources. – Determine proper motions using an external catalog: 2MASS, GLIMPSE(2), DSS, etc. or user supplied list.
  - Semi-real-time triggers for time-critical events of astrophysical interests during the third year of the VVV survey. These events include microlensing, novae/CVs, supernovae, microqiuasar activity, etc. The acceptable triggering delay is 1-2 days. Ideally, the pipeline will issue a trigger, i.e. sending e-mail.
  - Star catalogs with user-defined magnitude, color, proper motion constraints.
  - Extended source catalogs with user-defined magnitude, color, proper motion constraints, and morphology parameters (i.e. Sersic index, etc.)
  - Star counts in user-defined 2-dimensional cells on a grid in Galactic longitude and latitude, with user-defined cell size and constraints on the stellar magnitude, color and proper motion. Peak detection in the 2-dimensional star counts, above the neighbouring peaks. Products: 2-dimensional histograms, catalog of peaks containing peaks strenght in number of stars and sigmas above the background level, value of the background in number of stars.
  - List of all VVV observations on a user-specified area.
  - 3-color image of a specified sky area.
  - source identification within a certain radius from a given position or a list of positions.

- VIDEO:
  - Cross–identification with exisitng UKIDSS catalogue releases;
  - inclusion of SWIRE, various (unspecified) radio catalogues and future Hershel and x-ray catalogues, either as external catalogues in the VSA or provision of archive–end facility for user upload.

Clearly, some of the above (notably for the VVV) goes far beyond what can be reasonably expected of a science archive – some of the required functionality is already provided in multifarious client–side astronomy applications and it is wholly sensible to rely on those when small data sets are concerned. It is of course reasonable to require the archive to provide data in appropriate formats to enable manipulation in those same client–side applications, however.

References

[1] WFCAM Science Archive overview document; http://www.roe.ac.uk/˜nch/wfcam/VDF-WFA-WSA-001-I1/VDF-WFA-WSA-001-I1.html

[2] WFCAM/VISTA Science Archive Development, http://www.roe.ac.uk/˜nch/wfcam/

[3] The UKIDSS Proposal; http://www.ukidss.org/sciencecase/sciencecase.html

[4] The UKIDSS Infrared Deep Sky Survey (UKIDSS); Lawrence, A. et al., 2006, MNRAS, submitted; astro–ph/0604426

[5] Usages of the WFCAM Science Archive (Hambly, N.C. & Bond, I.A., 2003); http://www.roe.ac.uk/˜nch/wfcam/misc/wsausage.html

[6] VEGA: UK Excellence in Data Processing and Archiving — providing VO Processing for VISTA and GAIA (a revised proposal submitted in response to the 2003 PPARC e–Science Announcement of Opportunity)

9 ACRONYMS & ABBREVIATIONS

ADnn : Applicable Document No. nn
CASU : Cambridge Astronomical Survey Unit
CDS : Stellar Data Centre (Strasbourg, France)
CFHT : Canada France Hawaii Telescope
FITS : Flexible Image Transport System
GPS : Galactic Plane Survey (UKIDSS)
JAC : Joint Astronomy Centre
LAS : Large Area Survey (UKIDSS)
SDSS : Sloan Digitial Sky Survey
SQL : Structured Query Language
SSS : SuperCOSMOS Sky Survey
UKIDSS : UKIRT Deep Infrared Sky Survey
VDFS : VISTA Data Flow System
VISTA: Visible and Infrared Survey Telescope for Astronomy
WFAU : Wide Field Astronomy Unit (Edinburgh)
WFCAM : Wide–field infrared camera for the UK Infrared Telescope
WIRCam : Wide–field infrared camera for the CFHT
VO : Virutal Observatory
VOTable : XML format developed for astronomical data for the VO
VSA : VISTA Science Archive
WSA : WFCAM Science Archive
XML : eXtensible Markup Language
2MASS : 2 Micron All–Sky Survey

10 APPLICABLE DOCUMENTS

AD01	UK VISTA User Requirements	VDF-SPE-IOA-00009-0001 Issue: 3.0, 5/07/2005
AD02	VDFS Science Archive Database Design	VDF-WFA-VSA-007 Issue: 1.0, Sept. 2006

11 CHANGE RECORD

Issue	**Date**	**Section(s) Affected**	**Description of Change/Change Request Reference/Remarks**
Draft 1	Nov 2005	All	New document based on old WSA SRAD
Draft 2	Sep 2006	3,5,6,7	In preparation of UK VDFS FDR
Issue 1	Sep 2006	Final sections	Completion for UK VDFS FDR

12 NOTIFICATION LIST

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

WFAU:	P Williams, N Hambly
CASU:	M Irwin, J Lewis
QMUL:	J Emerson
ATC:	M. Stewart
JAC:	A. Adamson
UKIDSS:	S Warren, A Lawrence