VISTA DATA FLOW SYSTEM (VDFS)

---

for VISTA & WFCAM data

Management and Planning

author
P.M.Williams (WFAU Edinburgh)
WFAU Manager
number
VDF-WFA-VSA-003
issue
Issue 1.0
date
13 September 2006
co-authors
A.Lawrence, N.Hambly

Contents

1 SCOPE **3**

2 INTRODUCTION **3**

2.1 Context 3

2.2 Stakeholders 3

3 PROJECT DEFINITION AND MANAGEMENT **4**

3.1 Definition 4

3.2 Schedule 4

3.3 Management 6

4 RESOURCES **7**

4.1 Staff 7

4.2 Equipment 9

5 APPENDICES **10**

5.1 Standards 10

5.2 Risk Assessment 10

6 ACRONYMS & ABBREVIATIONS **11**

7 APPLICABLE DOCUMENTS **12**

8 CHANGE RECORD **12**

9 NOTIFICATION LIST **12**

1 SCOPE

This document considers and addresses the issues posed by management of the development of the Science Archives for the WFCAM on UKIRT (WSA) and the VIRCAM on VISTA (VSA). It identifies the resources available for the work. It describes the mechanisms for reporting and monitoring progress. It also considers the standards adopted by those working on the project and assesses the risks faced.

2 INTRODUCTION

2.1 Context

This document is part of the reference documentation for the VDFS Project.

The *development* of the VISTA Date Flow System (VDFS), including that of the WFCAM and VISTA Science Archives considered here, is funded by PPARC through a specific VDFS grant. This work is part of the programme of Wide-Field Astronomy and e-science at the Institute for Astronomy funded by PPARC through a Rolling Grant. The programme was last reviewed in 2004–5, and funds were sought, and granted, for the *operation* of the Science Archives. Despite separate funding routes, the development and operational staff work very closely together.

The WSA work, in addition to its intrinsic importance, was seen as a step towards (and test-bed for) the VSA. For this reason, it was agreed that the two be planned and designed together so that we could become more confident that the planned Archive is scalable to the data volumes expected from the VISTA IR camera, and ensure that no design decisions were taken which militated against scaling for use for VISTA science. It was also necessary that the archive be VO-compatible as soon as possible.

2.2 Stakeholders

Although the WSA is managed as part of VDFS, other groups, to whose needs VDFS is sensitive, have a strong interest in the products. This group includes the Joint Astronomy Centre, which is responsible for the operation of UKIRT and WFCAM, and the UKIDSS Consortium, which planned and is currently undertaking a set of public infrared surveys using WFCAM for most of its time on the telescope. These groups were instrumental in generation of the Science Requirements leading to the first Science Requirements Analysis Document (SRAD, AD01) and, following the release of WFCAM Science Verification data and publication of Data Release 1, have provided valuable feedback on the data delivery tools and user-interface of the of the Archive.

The group of stakeholders has broadened during the progress of the Project, and now includes astronomers from the ESO community.

The project is overseen by the VISTA Data Users' Committee (VDUC), which was constituted to provide independent and scientifically informed advice to the VDFS Project on behalf of VDFS's UK user community.

The VDUC is chaired by a nominee of PPARC and the membership includes the UKIDSS Survey Scientist, the VISTA and AstroGrid Project Scientists, and four other representative active researchers nominated by the UKIRT Board, the VISTA Project Board, the UKIDSS Consortium and the VISTA Consortium. It first met in 2004 May.

3 PROJECT DEFINITION AND MANAGEMENT

3.1 Definition

The Science Requirements Analysis Document (SRAD, AD01), based on the UK User Requirements document (URD, AD02) developed by the VISTA Project Scientist, is the primary determinant of the work to be undertaken.

3.2 Schedule

The original project plan set out in the 2004 VEGA-VDFS proposal described five versions (VDFSv1–v5) of the Science Archive, 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 archive scaled for VISTA VIRCAM data volumes, required to be ready for VIRCAM commissioning in December 2006, and VDFSv5, is the Archive fully shaken down with VIRCAM science data. Owing to the delay in delivery and commissioning of WFCAM, the time-scale between WFCAM coming into operation and the scheduled VISTA VIRCAM delivery has been greatly compressed, necessitating revision of the plan. The following revised plan was presented (AD03) to the VDUC at its meeting on 2005 September 6.

The principal change schedules preparatory work for the VISTA-ready version, VDFSv4 = VSA1 (scalability of WSA design, OS, DBMS and hardware for VISTA data volumes) in parallel with further development of data access tools for the WFCAM Science Archive (WSA3). Some of the design work for the VISTA Science Archive (VSA), namely the database schemas for the public surveys, can only be completed when the surveys themselves have been defined and accepted by ESO. This process is currently underway.

The top-level project plan is divided into five phases (we use the term 'phases' to allow the term 'versions' to be used for processed data):

1. WSA1: WFCAM Science Archive ready for receiving the first WFCAM data.

   As described in the Project Overview (AD03), in order to provide experience with scaling SQL Server to databases larger than the 6dF Survey database developed in 2002, this phase of the programme included the building of a copy of the 4-TByte SuperCosmos Object catalogue on SQL Server (the 'SSA'). This provided time for tuning the DBMS and Catalogue Server hardware and also a platform for testing the data access tools as they were developed, which was valuable as there was some delay in delivery and commissioning of the WFCAM.

2. WSA2: WFCAM Science Archive with experience of WFCAM science data (2004 October – 2005 September).

   It was recognized that the 'Day 1' archive (WSA1) would not provide all the functionality required by the SRAD and that there would also be some issues arising from the instrument itself and pipeline that would require re-factoring of some of the data ingest, curation and delivery tools. WSA2 was envisaged to be available one year after survey operations began but, because of the delay in WFCAM commissioning and the need to adhere to the VISTA time-table, we proposed closing this phase at the end of September 2005. The work on WSA2 was envisaged to benefit from feedback from members of the UKIDSS consortium working on Science Verification of the system, as occurred at the Science Verification meeting on August 18.

3. WSA3: WFCAM Science Archive with advanced data curation and access tools (2005 October – 2006 March)

From October 2005, part of the WFAU VDFS team continued to work on the WSA, WSA3, operating the archive and ingesting re-processed data, re-factoring database schemas and code in the light of experience and feedback from users, and adding additional facilities for catalogue maintenance and data delivery. This phase was extended, leading to the public Early Data Release (EDR) on February 10 and Data Release 1 (DR1) on July 25.

4. VSA1: VISTA Science Archive ready for receiving the first VIRCAM data (2005 October – 2006 December).

   The first phase of work from October 2005, in parallel with work on WSA3, was benchmarking of the DBMS/OS (MS SQL Server running under Windows) and hardware adopted for the WSA to see whether they were still appropriate for the VSA. The outcome from these studies is given in AD04. If a hardware architecture different from that hosting the WSA had been indicated, this would have given us time to procure and install the first nodes of the servers required for the VSA and iron out any bugs.

   Following DR1, we proceed to the following design work, leading to documentation for the present Final Design Review (FDR):

   - Requirements capture from URD 3.0 and development of SRAD (AD01)
   - Specify and agree data transfer interface control document (ICD) with CASU
   - Draw up Curation Use Cases analogous to those for the WSA
   - Design survey database schemas in the light of VISTA public survey decisions
   - Adapt WSA data ingest tools (CU1–4) to fit the ICD and VSA database design
   - Adapt WSA catalogue maintenance tools to fit VSA database design
   - Design survey progress tools for VISTA surveys
   - Design VSA User interface and adapt WSA data delivery tools for VSA data

   Following the FDR, and in the light of its outcome, we will proceed with the implementation of the design. Further hardware for data servers will be acquired and installed as required. Data transfer between CASU and WFAU using UKlight will be implemented.

   Towards the end of the construction phase, we expect to receive specimen files from CASU based on engineering data to test the Curation Tasks. It is hoped that the data will include some on-sky frames, but this depends on the progress of the installation and engineering commissioning of VISTA and VIRCAM.

5. VSA2: VISTA Science Archive shaken down with experience of VIRCAM data. (originally 2007 January – September)

   We expect a period of shake-down of the data-flow system and VSA once real data flow from VISTA. In the light of operational experience, and user feedback, we may have to re-factor some of the tools for data ingest, catalogue maintenance and data delivery. We expected that there would be time for feedback and re-factoring before the current VDFS funding ceases on 30 September 2007.

   Given that the scheduled date for handover of VISTA + VIRCAM to ESO, and the beginning of Science Commissioning is scheduled for August 2007, there is practically no time for this work and, as described below, funds are being sought from PPARC to extend the VDFS grant to allow it to be done.

   In any event, during this phase, we will add further data access tools as specified in the UK VISTA User Requirements (URD 3.0) and upgrade our VO-compatibility.

The programme is summarized in a Gantt chart (Fig. 1), which marks the significant milestones.

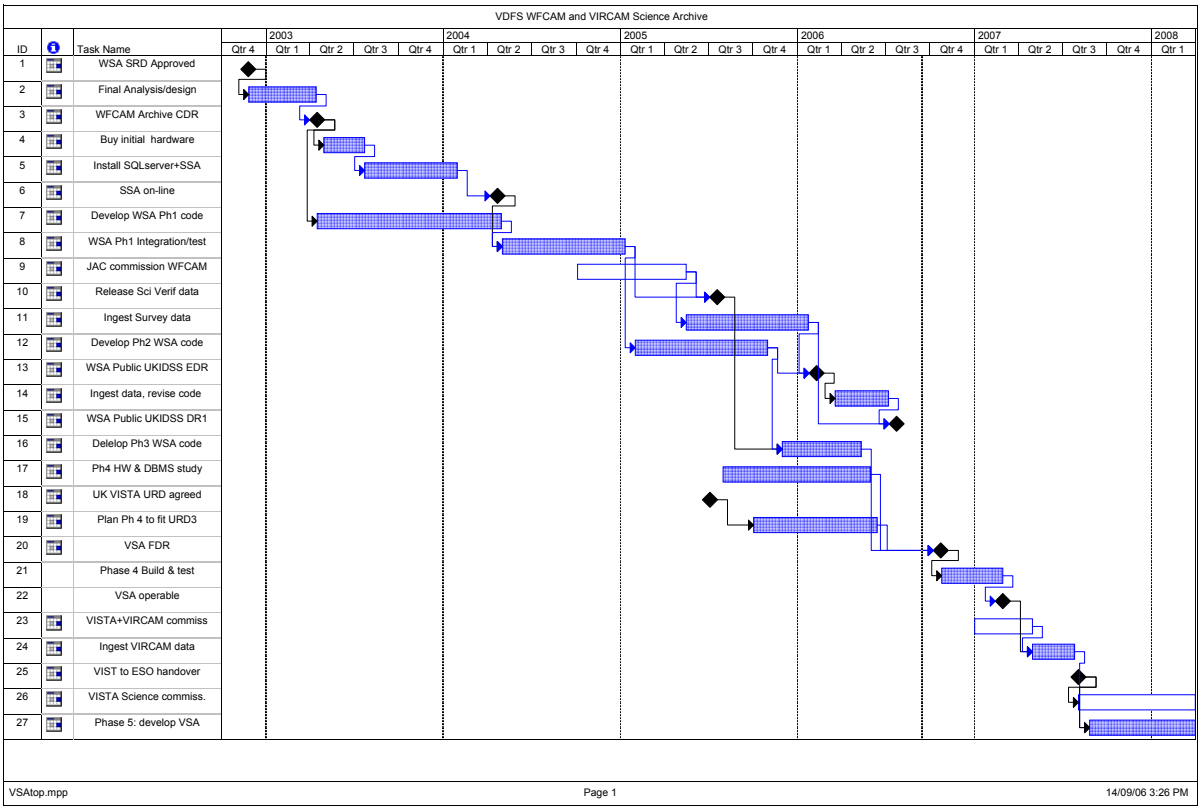

Figure 1: *Top level Project Gantt chart. The open bars mark external tasks.*

3.3 Management

The VDFS management structure is shown in Figure 2, where the Management Team are Emerson (VDFS PI), Lawrence (VDFS co-I), McMahon (VDFS co-I), Stewart (VISTA Software Manager), Irwin (CASU Manager) and Williams (WFAU Manager). Stewart's particular rôle is in ensuring coordination of the VDFS ESO deliverables with the VISTA camera work.

There are four time-scales (funding, quarterly, monthly, weekly) on which the progress of the Archive work is assessed and managed.

The longest is that defining the overall long-term aims. The VDFS Project is currently funded from 2004 October – 2007 September by PPARC through grants to WFAU and CASU.

On a calendar Quarterly cycle, detailed work plans and deliverables are defined and the work allocated amongst the available staff. This is regularly monitored throughout the quarter (see below) and at the end of the Quarter the progress against the plan and deliverables is assessed, and a new Quarterly plan is made which takes account of the successes, and any failures, in the previous quarter. The quarterly past progress and future plans are closely scrutinised by the PI+Management Team, and used as the basis for reporting to the Grid Steering Committee (at intervals of approximately six months).

On a monthly cycle, the WFAU Manager reports progress against plans to the Management Team who discuss them. This allows corrective actions to be taken within the quarterly time-scale. At the same time, the increases in Earned Value and expenditure on the VDFS Grant are reported to the VDMT.

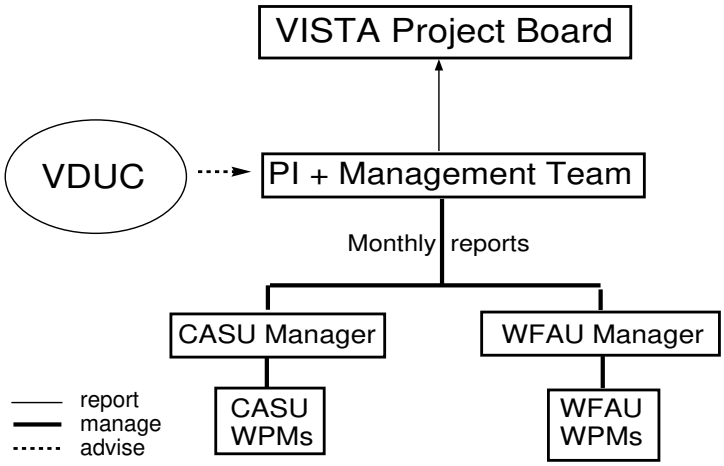

Figure 2: *Management structure of the VDFS Project*

On a weekly cycle, the staff working on the Archive project meet to review progress, allocate tasks and actions. These meetings are minuted and copies of the minutes are provided to the VDFS PI+Management team. They are also available to the JAC.

The Project maintains a Twiki collaborative web site on which all documents, formal and informal, are shared. These include notes from meetings and external visits, equipment set-ups and experiments with different hardware and software.

4 RESOURCES

4.1 Staff

Staff Effort for the Science Archive component of the VDFS Project is provided by members of the Wide Field Astronomy Unit. Those assigned to the Archive development, together with fractions of their effort as funded through the 2004–07 WFAU VDFS Grant, are:

- Dr N. Hambly (Project Scientist) (70%)
- Dr R. Collins (Scientist/developer)
- Dr N. Cross (Scientist/developer) (70%)
- Mr M. Read (Scientist/developer)
- Dr E. Sutorius (Developer/systems)
- Dr P. Williams (Management) (30%)

The following effort is funded by the 2005–10 WFAU Rolling Grant:

- Mr J. Bryant (Archive Operation)
- Dr E. Sutorius (Archive Operation from 1/10/2007)
- Mr M. Holliman (30%, Systems Management)

Ref	SRD1	CU	Ph	VDFS-WP-6 sub-tasks and work programme / 2006 Q3 (July-Sept)	who	sw	Jly	Aug	Sep
				Management and Contingency					
			3	Prepare MNRAS paper on the WSA	NCH, all	4	10	10	
				Prepare reports as required &c.	NCH, PMW	1	10	50	
				Design of Archive VDFS Ph4 following URD 3.0, Survey designs & benchmarking					
40a			**4**	Review URD for Phase 4 and revise Science Requirements Analysis Document (...-VSA-002)	NCH	2	10	50	
40b			**4**	Review VISTA survey designs for impact on Requirements document (...-VSA-002, -VSA-007)	NCH	1	10	50	
42b			**4**	Document VISTA OS/DBMS and ingest software design (...-VSA-006),	ETWS, JB	2	10	50	
42c			**4**	Revise database design document (...-VSA-007)	NJC, NCH	3	0	10	
41h			**4**	Investigate network attached storage options	JB, ETWS	1	10	50	
42c			**4**	Investigate and revise, if necessary, data flow document (...-VSA-005)	RSC	2	10	50	
42d			**4**	Document hardware design for VISTA data volumes (-VSA-006)	JB, ETWS	2	10	50	
41g			**4**	Specify, agree and document ICD with CASU to reflect ESO FITS headers (-VSA-004)	NCH	1	10	50	
42e			**4**	Document network transfer solution for data transfer (-VSA-006)	ETWS, JB	1	10	50	
42f			**4**	Document revised Software Architecture Design (-VSA-009)	RSC	2	10	50	
42g			**4**	Revise management document (-VSA-003)	PMW	1	0	50	
42h			**4**	Produce revised user interface documentation (-VSA-008)	MAR	3	10	50	
				Database and Pixel Data Hardware					
44a			**4**	Procure, install third Catalogue server (Thutmose)	ETWS, PMW	2	50	50	
44b			**4**	Specify, procure, install further pixel data servers ready for first VISTA data	PMW, ETWS	2	0	0	
44c			**4**	Specify, procure, install networking hardware for UKLight connection to CASU	PMW, JB	1	10	10	
				Data transfer and ingestion tools					
32c		CU_all	**3**	Re-factor data transfer and ingest tools and operator GUI as required, add DQC information	ETWS, RSC	2	10	10	
38a			3	Re-factor tools to include Sextractor UDS data from Nottingham	NJC, ETWS	2	100	100	
32d			3	Design and implement Quality bit flagging	RSC, MAR	3	10	10	
				Catalogue maintenance tools					
33a			**3**	Re-factor Phase 2 tools in the light of experience	RSC	3	50	50	
38b				Re-factor to include Sextractor UDS data	NJC	2	100	100	
33				Design and implement tools for recalibration procedures	NJC	3	0	10	
5f		CU9	**3**	Build and test facility to produce list-driven measurements between WFCAM detections using CASU-provided tool.	RSC, NJC	4	0	0	
				Data Delivery and Image processing tools					
37a			**3**	Code maintenance: re-factor Phase 2 tools in the light of experience/feedback	MAR	4	10	50	
				Provide User Interfaces to PTS and calibration databases	*MAR*	*2*	50	100	
8d			**3**	Write and maintain on-line documentation for use of WSA interface & tools; revise for UDS	MAR, JB	2	50	50	
				Provide fast flat file access to ingest DB Survey data	MAR	*1*	*100*	*100*	
						55	12.3	23.2	0.0

Figure 3: *Example of the quarterly chart on which progress is reported to the VDMT at its monthly meetings. The numbers under the months in the last three columns are pencentages complete, reported to high granularity (10% = started, 50% = nearly complete and 100% = complete).*

The Project also receives specialist assistance and advice, particularly in the field of large databases from Dr R. Mann (WFAU, NeSC and AstroGrid). We also receive help and advice from NeSC and the University of Edinburgh, IBM and Microsoft.

As noted above, the VDFS Grant funding the development runs until 30 September 2007. After that date, one member of the development team (Dr Sutorius) is funded to move to Science Archive operations and the remainder of the team is expected to be dispersed to other projects. This is scheduled to occur about one month after the scheduled date for the beginning of VISTA+VIRCAM Science Commissioning, leaving practically no time for the Phase 5 work shaking down the Archive with experience of science data described above.

Funds are therefore being sought from PPARC to extend the VDFS Grant by six months to 31 March 2008, the review date of the WFAU Rolling grant. Proposals for the Rolling Grant will be prepared in March–April 2007 and will include requests for funding for:

- continued operation of the Science Archives at an appropriate level

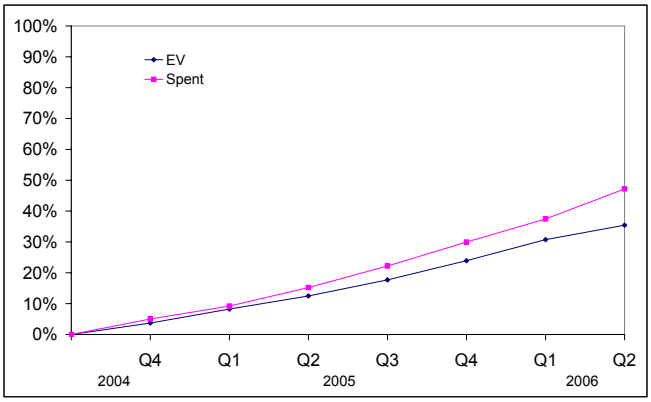

Figure 4: *Plot of percentage Earned Value (EV) and VDFS grant expenditure for the first seven quarters of the Project.*

- possible continued shake-down of the Archive in the light of up-to-date knowledge of the commissioning schedule
- possible continued development of data delivery tools as our experience from the WSA has shown that user requirements evolve during the initial phase of survey operations
- interactive quality control of the data: our experience with WFCAM data and the WSA has shown that significant effort from the team (the Project Scientist in this case) has been required to assist the Survey Scientist in quality control of the data.

4.2 Equipment

Preliminary work on the Project made use of equipment provided by the University of Edinburgh or to the University for the Wide Field Astronomy programme through a major JREI grant and the WFAU PPARC Rolling Grant. This included a 6-TB RAID system used for the SSA. The University also provided file server systems for the WSA Catalogue Servers. The VDFS grant has funded purchase of another Catalogue Server and funding for the pixel data servers is provided by PPARC though the WFAU Rolling Grant.

The equipment identified as being required for the Project is described in the Hardware/OS/DBMS design document (AD04). The equipment will be purchased in accord with standard University of Edinburgh procedures, including tendering where required, with funding provided for Archive Operations in the WFAU Rolling Grant.

__oOo__

5 APPENDICES

5.1 Standards

- DBMS

  The Science Archive uses SQL Server 2000, running on Windows 2003 Advanced Server.

- Languages

  Code and scripts will be written in one of the languages listed below. Whilst recognising the advantages (e.g. maintenance of code) of using as few languages as possible, we settled on a slightly longer list to make best use of the extensive but inevitably varied experience of the staff working on the project.

  – C/C++
  – Python
  – SQL
  – Java

- External subroutine libraries

  The baseline set of external subroutine libraries to be used follows:

  – CFITSIO
  – SLALIB
  – WCSTOOLS
  – ODBC/JDBC
  – Python middleware: e.g. MxODBC, wxPython, PyFITS
  – HTM (as implemented in SkyServer)

- The adoption of coding standards has been discussed within the group. These are in addition to the commenting/documenting procedure. After examining standards used by Starlink, ESO and ALMA, it was decided to follow ALMA but substituting use of CVS for RCS. (The ALMA standard refers to C/C++ only).

- Archiving and versioning code (CVS)

  We use CVS (Concurrent Versions System) for these purposes, following usage at the ATC, JAC and CASU. CVS provides a central, controlled repository for source files (both coding and scripting) and records the history of their modification. The ATC CVS system is being used and all project source code and scripts reside in the VDFS project module.

- Documenting code (doxygen, epydoc, JavaDoc)

  We use a documentation format (i.e. source commenting) standard within source code to allow use of doxgen, a documentation generation utility for production of easily read information (in html, LaTex and postscript) concerning comments and functionality without recourse to the source itself. The standard JavaDoc facility is used for Java code.

5.2 Risk Assessment

We have considered the risks, both external and internal, which might jeopardise completion of the Project on time. We estimate the likelihood (L), effect (E) and impact (I = L × E) of each event, scored on a scale of 1–3, and give for each our mitigation strategy. This table will be kept in a separate Risk Register and the risks re-assessed periodically.

Event	L	E	I	Mitigation
Serious illness or loss of staff	1	3	3	de-scope: delay lower priority tasks
Network too slow for data transfer	1	3	3	use same method as for JAC and ESO to CASU
CASU/external tools delivered late	1	2	2	use preliminary versions supplied
Failure to make interfaces work	1	3	3	seek specialist advice
Development funds cease too early for shake-down with real data	3	3	9	seek extended funding

6 ACRONYMS & ABBREVIATIONS

ADnn : Applicable Document No nn
CASU : Cambridge Astronomical Survey Unit
DBMS : Database Management System
ESO : European Southern Observatory
JAC : Joint Astronomy Centre (Hawaii)
JREI : Joint Research Equipment Initiative
NeSC : National e-Science Centre
SRAD : Science Requirements Analysis Document (AD01)
UKIDSS : UKIRT Infrared Deep Sky Survey
UKIRT : United Kingdom Infrared Telescope
VDUC : VISTA Data Users' Committee
VIRCAM : VISTA Infrared Camera
VISTA: Visible and Infrared Survey Telescope for Astronomy
VO : Virtual Observatory
VSA : VISTA Science Archive
WFAU : Wide Field Astronomy Unit (Edinburgh)
WFCAM : (UKIRT) Wide Field Infrared Camera
WSA : WFCAM Science Archive

7 APPLICABLE DOCUMENTS

AD01	Science Requirements Analysis Document	VDF-WFA-VSA-002 Issue: 1.0 20/09/2006
AD02	UK VISTA User Requirements	VDF-SPE-IOA-00009-0001 Issue: 3.0, 5/07/2005
AD03	VISTA DATA-FLOW SYSTEM SCIENCE ARCHIVE (VSA): REVISED PLAN	VDUC(05)09 06/09/2005
AD04	VSA Hardware/OS/DBMS design	VDF-WFA-WSA-007 Issue 1.0 02/09/06

8 CHANGE RECORD

Issue	**Date**	**Section(s) Affected**	**Description of Change/Change Request Reference/Remarks**
Issue 1.0	13/09/06	All	New document

9 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: A. Lawrence, S. Warren

__oOo__