

VISTA DATA FLOW SYSTEM (VDFS)

for VISTA & WFCAM data

MANAGEMENT AND PLANNING

author

P.M.Williams (WFAU Edinburgh)

WFAU Manager

number

VDF-WFA-WSA-003

issue

Issue 1.0

date

2 April 2003

co-authors

A.Lawrence, N.Hambly

Contents

1 SCOPE	3
2 INTRODUCTION	3
2.1 Context and Stakeholders	3
3 PROJECT DEFINITION AND MANAGEMENT	3
3.1 Rôle of the Science Requirements Analysis Document (SRAD)	3
3.2 Project definition	3
3.3 Management	6
4 RESOURCES	7
4.1 Staff	7
4.2 Equipment	8
5 APPENDICES	9
5.1 STANDARDS	9
5.2 Risk Assessment	10
6 ACRONYMS & ABBREVIATIONS	10
7 APPLICABLE DOCUMENTS	10
8 CHANGE RECORD	10
9 NOTIFICATION LIST	10

1 SCOPE

This document considers and addresses the issues posed by management of the development of the Science Archive that is simultaneously the archive for the UKIRT WFCAM (WSA) and the test-bed for the VISTA Science Archive. It identifies the resources available for the work. It shows how the agreed functional requirements in the Science Requirements Analysis Document (SRAD) have been translated into the work and 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 and Stakeholders

This document is part of the reference documentation for the WFCAM Science Archive part of the VDFS project. There are various groups of interested parties.

The WFCAM Science Archive (WSA) is part of the VISTA Data Flow System (VDFS), funded through the Grid Steering Committee (GSC). The WFCAM work, in addition to its intrinsic importance, is seen as a step towards (and test-bed for) the equivalent system for VISTA. For this reason, it has been agreed that the two be planned and designed together so that we can 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 are taken which militate against scaling for use for VISTA science. It is also necessary that the archive be VO-compatible as soon as possible.

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 has planned and is undertaking a set of public infrared surveys using WFCAM for most of its time on the telescope. These groups will want to be confident that the Archive will be ready to deliver scientific data to members of the Consortium, other WFCAM users and the astronomical community, by the time the WFCAM is on the telescope (January 2004). They were instrumental in generation of the Science Requirements leading to the Science Requirements Analysis Document (SRAD, AD01) and will measure deliverables against the agreed requirements. Their formal input is via the VDFS Advisory committee (see below)

The Science Archive 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. This programme was last reviewed in July–December 2002. The ‘new’ grant is expected to start in April 2003.

3 PROJECT DEFINITION AND MANAGEMENT

3.1 Rôle of the Science Requirements Analysis Document (SRAD)

The Science Requirements Analysis Document (SRAD, AD01), developed closely with and agreed by the stakeholders including UKIDSS, is the primary determinant of the work to be undertaken. A corollary, and secondary advantage, of following the agreed SRAD, is the avoidance of ‘specification creep’.

3.2 Project definition

The top-level project plan is set out in a Gantt chart (Fig. 1), which marks the significant milestones.

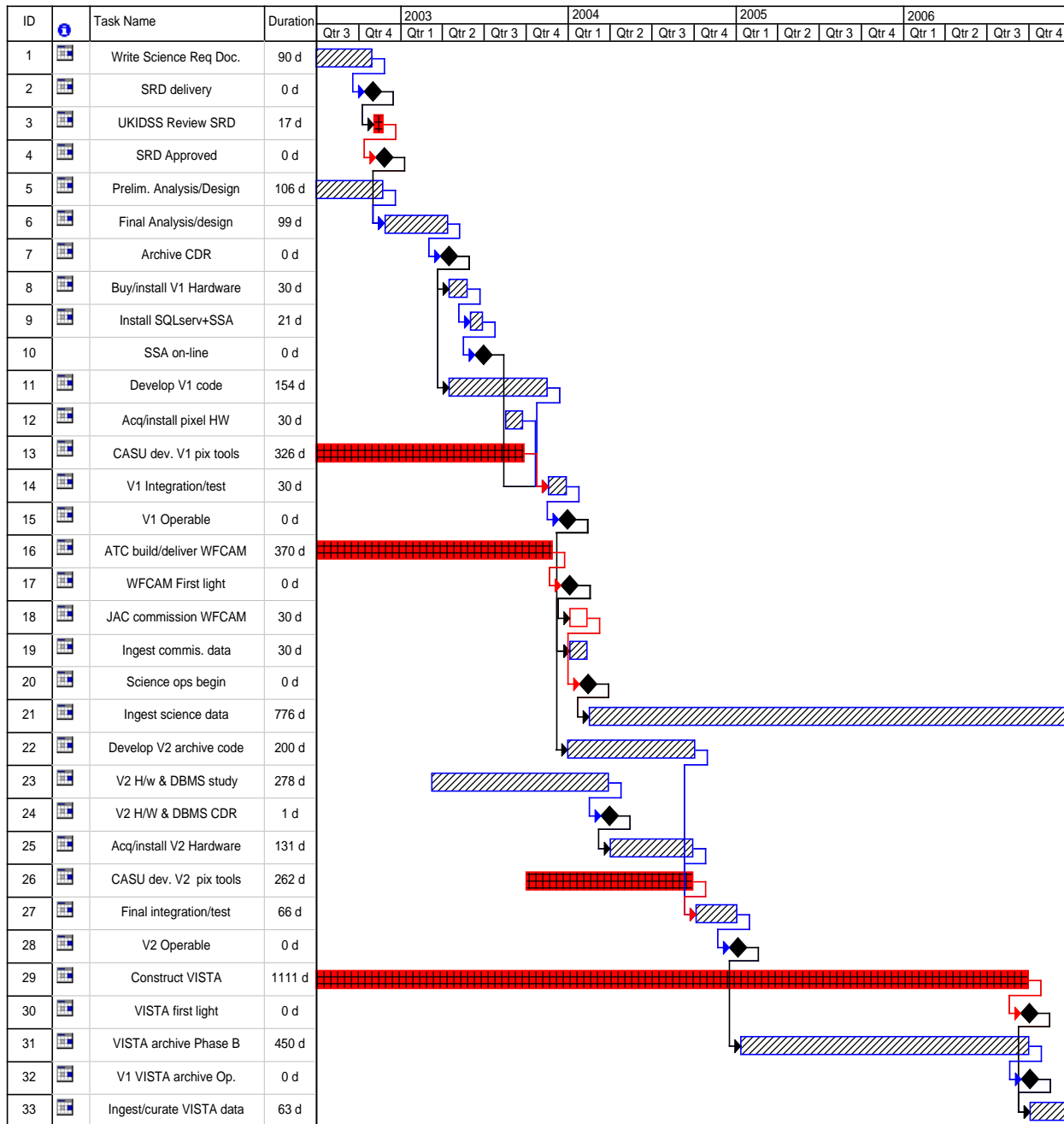


Figure 1: *Top level Project Gantt chart. The darker bars mark external tasks.*

As described in the Project Overview (AD03), in order to allow delivery of a working Science Archive in time for the first use of WFCAM and at the same time allow a thorough analysis of the suitability of available DBMS for large scientific archives, the work was divided into two phases, producing two versions of the Archive. The DBMS in Version 1 will be SQL Server, already used for 6dF Data Archive developed by the WFAU in 2002. To provide experience with scaling SQL Server to databases larger than the 6dF database, the Version 1 programme includes the building of a copy of the SuperCosmos object catalogue on SQL Server (the 'SSA') running on the Version 1 catalogue server hardware. This will give time for tuning the DBMS and provide a platform for testing the access tools as they are developed as well as forming part of the WSA as one of the agreed 'external catalogues'. Studies of the suitability of alternative DBMS, Oracle and DB2, will continue in parallel with the development of the Version 1 Science Archive, leading to a final decision in time for a Review of Version 2 DBMS

and Hardware in April 2004.

The Version 1 Archive development work has been divided into three major phases:

- Analysis and Design, leading up to the CDR in April 15-16,
- Version 1 Build phase, running from this CDR to November 18, and
- Integration and Test, running from November until delivery of the Archive at the end of the year.

The principal milestones in the Analysis & Design Phase were the development and agreement with the UKIDSS Consortium of the SRAD in December 2002 and the preparation of documents for the CDR in April 2003. As reported in the CDR documents, data products and data flow were analysed and DB schemas were developed. Different computer hardware configurations, including some borrowed equipment, were set up and loaded with sky survey data to test data trawl and access speeds. There were also investigations of data transfer rates over the network between the WFAU and CASU.

The requirements of the SRAD, amplified by the 20 ‘Usages of the WFCAM Science Archive’ (AD02), were used to identify the modules required for each version of the Archive. The modules for transfer and ingestion of data from the CASU pipeline and curation of the data were identified from a set of 20 ‘Curation Use cases’ (Appendix to the Database design Document, AD04) and priorities were set to inform scheduling.

The effort, in man-weeks, required for each module was estimated and the work then allocated to different staff members, or groups of staff members where a range of skills and/or experience was believed to be required. From this, the plan for the Version 1 Build phase was developed. This is shown in the second Gantt chart (Fig. 2). The plan was also split up by calendar quarter for detailed planning, reporting and monitoring purposes (see next section).

Formally, the Integration and Test phase of Version 1 begins on November 18 but it is expected that testing of many aspects of the user interface have been undertaken earlier using the SSA database.

The current plan for Version 2 includes migration of the object catalogues to the new DBMS if it is decided to move from SQL Server. This will build on the Version 1 work, with re-use of components, so the effort required will be less than if we were starting afresh. As for Version 1, the tasks and their effort requirements have been identified. The total effort is within the available staff effort to allow completion on time but, given the uncertainty of whether or not a DBMS migration will be required, a detailed plan has not yet been drawn up.

Key dates for the whole programme are summarized below:

WFCAM CDR	Sept 2002
SRAD agreed	Dec 2002
WSA CDR	Apr 2003
SSA prototype	June 2003
V1 operational	Jan 2004
V2 CDR	Apr 2004
V2 operational	Jan 2005
VSA Phase B start	Feb 2005
VSA operational	Sep 2006

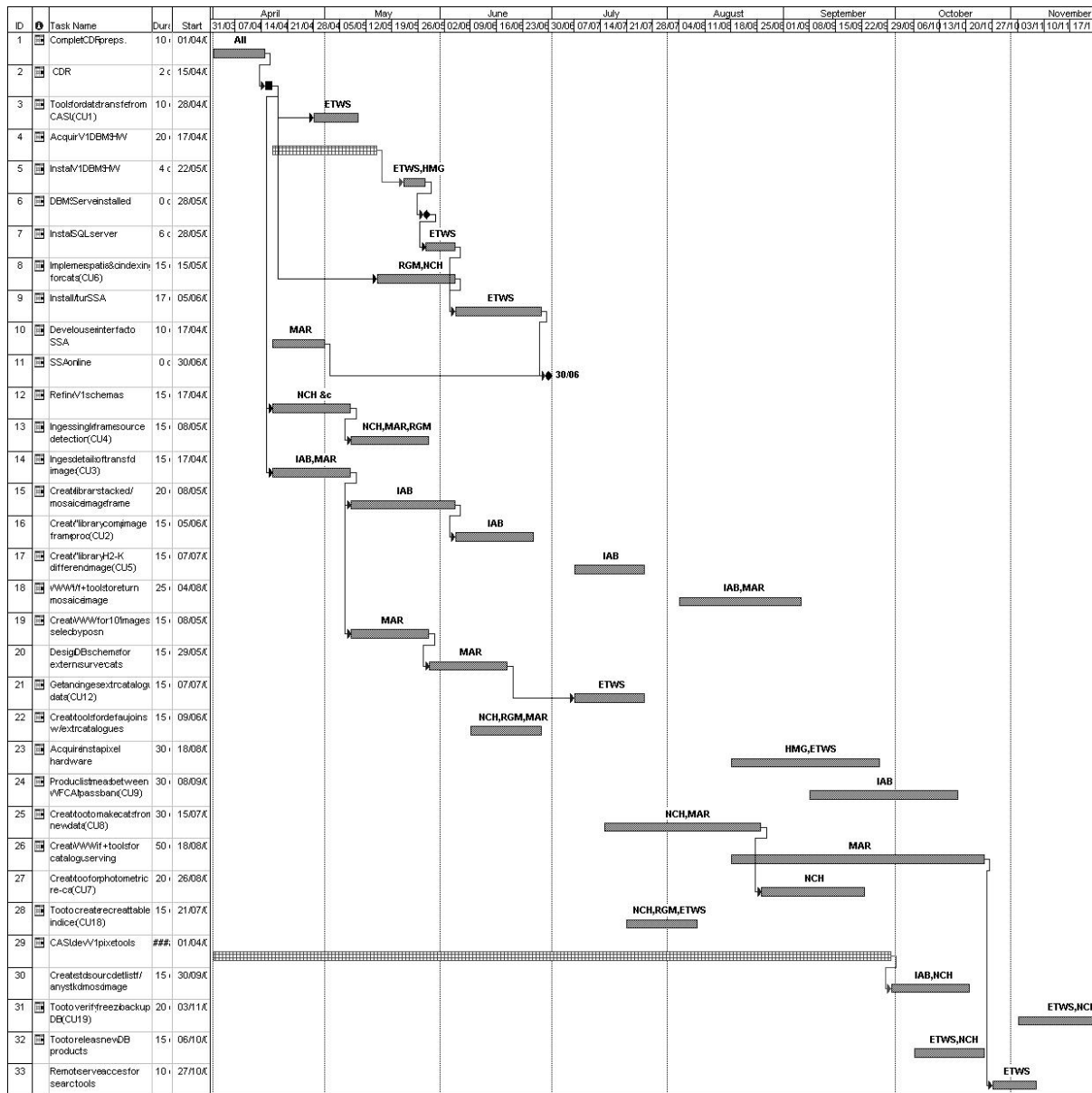


Figure 2: V1 Build phase Gantt chart.

3.3 Management

The management structure is shown in Figure 3, where the Management Team are Emerson (VDFS PI), McMahon (VDFS co-I), Lawrence (VDFS co-I), Stewart (VISTA Project Office), 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. The Advisory Committee include the UKIDSS Project Scientist (Warren) and the UKIRT Director (or his representative).

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. Currently these are proposed to the Grid Steering Committee for funding as part of the UK e-science programme. The current program was funded in response to paper GSC(02)03 - dated March 01 2002, presented to the GSC and seeking

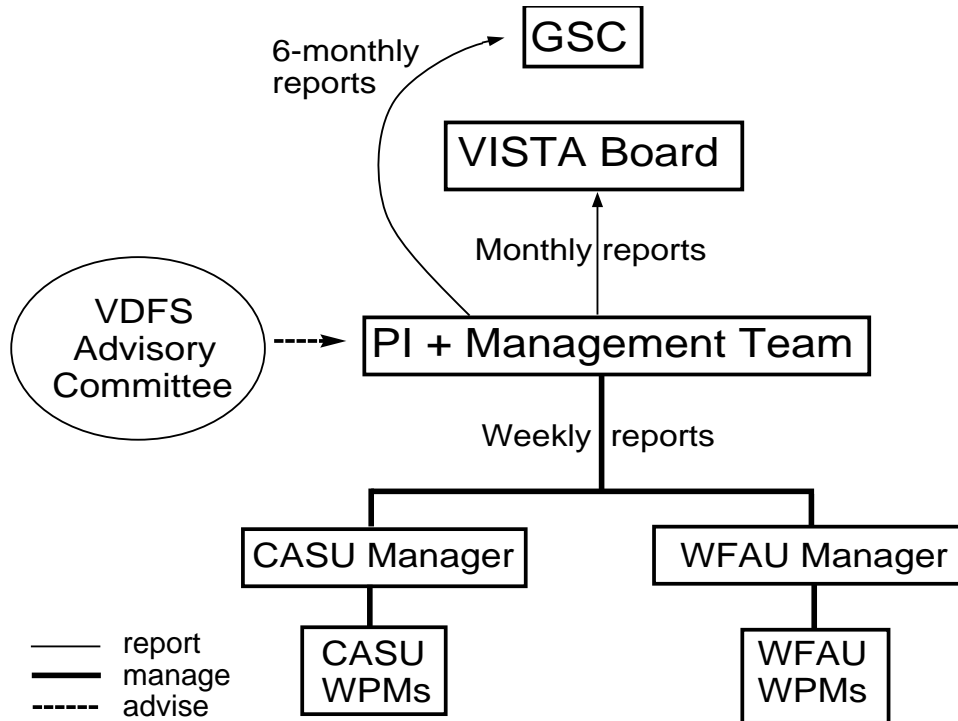


Figure 3: Management structure of the VDFS Project

funding of the VISTA Data Flow System & Test Bed which was submitted by principal applicant JP Emerson with co-applicants RG McMahon & A Lawrence. The funding itself does not come in a separate grant but as part of Edinburgh's Rolling Grant for Wide Field Astronomy.

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, and forms the basis (along with similar reports from the CASU pipeline) of monthly reports to the VISTA Board, who were asked by PPARC's Director of e-science to be an oversight body for VDFS.

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 Project is provided by members of the Wide Field Astronomy Unit. Those assigned to the Project, together with fractions of their effort as funded through the Wide Field Astronomy

Grant, are:

- Dr N. Hambly (proj sci) (70%)
- Dr I. Bond (sci/dev) (70%)
- Mr M. Read (dev) (50% in 2003-04 and 100% in 2004-5)
- Dr E. Sutorius (dev/sys) (50%)
- Dr H. MacGillivray (sys) (10%)

Secondly, the Wide Field Astronomy Unit has the following effort to cover its whole programme, from which the WSA Project derives benefit:

- Dr H. MacGillivray (20%, Systems Management)
- Dr E. Sutorius (50%, Systems Management)
- Dr P. Williams (50%, Unit Management)

Thirdly, the Project 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, Oracle, IBM and Microsoft.

4.2 Equipment

Preliminary work on the Project has made use of equipment provided by the University of Edinburgh or to the University for the Wide Field Astronomy Programme either through two major JREI grants or the WFAU PPARC Rolling Grant. This includes a 6-TB RAID system with tape backup and two Beowulf clusters. The WFAU has received significant help from Eclipse Computing, its External Sponsor and partner in the JREI bids, in the form of the loan equipment for testing disk I/O bandwidth, in particular. This equipment has allowed staff to gain experience with serving of astronomical databases, practical use of the databases Objectivity/DB and SQL Server, and experimentation with different data-storage configurations. This experience has been invaluable in developing the design for the equipment required for the Archive.

The equipment identified as being required for the Project is described in the Hardware/OS/DBMS design document (AD05). The equipment will be purchased in accord with standard University of Edinburgh procedures, including tendering where required. It will be located within a secure environment, room 'C1' at the Royal Observatory, provided with an automatic fire extinguisher system and hardened to Home Office Class 3 standards for security.

--oOo--

5 APPENDICES

5.1 STANDARDS

- DBMS

Version 1 of the Science Archive will use SQL Server 2000, running on Windows 2000 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++
- Java
- Perl
- Python
- SQL
- Fortran 77

- External subroutine libraries

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

- CFITSIO
- SLALIB (C and Fortran versions)
- WCSTOOLS
- PGPLOT
- 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 will be used and all project source code and scripts will reside in the WSA project module.

- Documenting code (doxygen)

We use a documentation format (i.e. source commenting) standard within source code to allow use of doxygen, 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 can also be 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 ($I = L \times E$) 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
Delay in delivery of V1 Server HW	2	2	4	not on critical path for 1 month
SSA Implementation underestimated	1	3	3	not on critical path for 1 month
SQL Server insufficiently scalable	1	3	3	test by producing SSA prototype
DB Relational model failure	1	3	3	Submit plan to CDR; seek advice from SDSS contacts and DBMS experts
Network too slow for data transfer	1	3	3	use same method as for JAC to CASU
CASU/external tools delivered late	1	2	2	use preliminary versions supplied
Catalogue data trawl rate too slow	2	2	4	revise design for Version 2
Failure to agree JAC-CASU ICD meta-data	1	3	3	Sort out through Management Team
Failure to make interfaces work	1	3	3	seek specialist advice
Pixel processing tools fail	1	3	3	seek local/external advice
Delay in delivery of pixel store HW	1	2	2	not on critical path for 1 month

6 ACRONYMS & ABBREVIATIONS

ADnn : Applicable Document No nn
CASU : Cambridge Astronomical Survey Unit
DBMS : Database Management System
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
VISTA: Visible and Infrared Survey Telescope for Astronomy
VO : Virtual Observatory
WFAU : Wide Field Astronomy Unit (Edinburgh)

7 APPLICABLE DOCUMENTS

AD01	Science Requirements Analysis Document	VDF-WFA-WSA-002 Issue: 1.3 20/03/03
AD02	Usages of the WFCAM Science Archive	http://www.roe.ac.uk/~nch/wfcam/misc/wsausage.html
AD03	WSA Project Overview	VDF-WFA-WSA-001 Issue 1.0 02/04/03
AD04	Database design document	VDF-WFA-WSA-007 Issue 1.0 02/04/03
AD05	WSA Hardware/OS/DBMS design	VDF-WFA-WSA-006 Issue 1.0 02/04/03

8 CHANGE RECORD

Issue	Date	Section(s) Affected	Description of Change/Change Request Reference/Remarks
Issue 1.0	02/04/03	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--