VDFS-WSA Quarterly Report: 2003 April – June (Quarter 2)

1. General progress and the archive CDR

The Science Archive CDR 2003 took place on April 15-16. The Panel comprised Luc Simard (CADC, Chair), Frossie Economou (JAC) and John Lightfoot (ATC). Also represented were VISTA, UKIDSS, Subaru, JNO, CASU, ATC and the JAC. Copies of the documents submitted by WFAU, and of the PowerPoint presentations to the Review, are available on the WFCAM documentation site (http://www.roe.ac.uk/~nch/wfcam/).

In its Report, also on the WFCAM documentation site, the Panel reviewed the individual documents and presentations separately and then considered the design against its terms of reference. The Panel considered the end product to be acceptable and external interfaces to be well understood. They raised concerns mainly in two other areas: (a) software architecture and design and (b) feasibility of the project schedule and staffing levels.

The Panel was concerned at the absence of an overall software architecture and design from the documentation and that the code written for V1.0 would be written in an ad hoc ("quick and dirty") way, especially given the tight timescale of the project. Such code would be hard to maintain, with a deleterious impact on the migration from V1.0 to V2.0 of the WSA.

Accordingly, and notwithstanding the tight schedule, the development of a proper top-down design was inserted in the project plan. This builds on the analyses of user requirements and curation use cases, and is documented in VDF-WFA-WSA-009, WSA Software Architecture Design (SAD). Although this interrupted work on individual software modules in the programme, we believe that this work will save time in the long run.

The Panel was concerned that the project schedule was too optimistic. This has been concerning Project for some time. Analysis of the work plan earlier this year showed that with existing staff resources, the functionality defined in the agreed Science Requirements Analysis document could not be completed to fit the timescale set by the scheduled delivery date for the WFCAM. Accordingly, a bid for an additional 0.5 FTE for 18 months from March 2003 was made to the Grid Steering Committee in February 2003, but was not successful. Given fixed resources and a fixed time-scale, we are prioritizing requirements so that basic curation and user access functionality at least is in place when required. As a de-scope option suggested by the CDR Panel, we have given lower priority to installing external catalogues and will begin with only the 2MASS.

In May, we contributed input to the VDFS section of the VEGA bid for esci2 funding, identifying resources required to continue the project and requesting additional staffing from April 2004 to speed deliverables for V2.0.

During Quarter 2, effort was devoted to two experimental studies of system performance: network transfer of data from CASU to WFAU and catalogue trawl rates for realistic SQL queries. In each case, suitable choice of configurations and equipment has led to significant increases in speed. The SQL benchmarking experiments have informed the choice of hardware for the Catalogue Server (SCSI disks with Ultra320 controller) and the first tranche of this equipment was ordered in June. This equipment is slightly

(20%) more expensive than that specified earlier this year but the data trawl rates are over six times faster.

We have since bid for and received SRIF funding to buy additional SCSI disks and the catalogue Load server. With this welcome contribution, we will be able to set up the V1.0 immediate hardware (i.e. including the master node of the pixel server) within the requested budget.

2. Progress on identified tasks and deliverables

Tasks are referred to by Ref numbers (derived from tables prepared for GSC paper and retained for continuity) in the table of tasks, deliverables and completion levels.

(0) CDR documentation and presentations were prepared and are available as described above.

(1) The SQL Server schemas for WFCAM data, UKIDSS surveys and curation log have been written and are stored in the CVS. Parsing scripts have been written to provide HTML documentation from the schema files.

(2a) This task has been given low priority in line with CDR comments and postponed to future quarters, beginning with 2MASS in Q3.

(3) The HTM indexing for the SSA has been done but work on indexing WSA has been put on hold pending availability of new HTM routines from SkyServer.

(4) Testing of the DB Server hardware configurations as reported above have gone on longer than expected so the first tranche has only just been ordered. SQL Server has been installed on another machine (grendel12) and a mini version (6deg square) of the SSA has been built for verification. It has been used for development of positional searching (see Item 8 below).

(5a) Tests on network data transfer between CASU and WFAU have show the way to get adequate bandwidth; writing and implementation of the scripts has been put back until the pixel server hardware is installed in Q3.

(5b) Based on the Software Architecture Design developed (see above) the Python script for this task has been completed and tested.

(5c,d) The scripts for these, again based on the SAD, are under development

(5i) Work on this is in progress but given lower priority until the SSA is in place.

(5g) This has been designed and the script is under development
(8a) The original task has been split into two components: the WWW interface and tools for searching by position to return object tables from the Catalogue and the extension of this to return finder-chart sized (10') images from the pixel data. The latter component has been deferred until installation of the pixel data hardware in Q3. The WWW interface for searching by position (equatorial and galactic) or by position provided by SIMBAD

from a resolvable name, and returning SQL-selected catalogue data has been completed and tested on the prototype SSA. This will be extended to the WSA in Q3.

3. Programme for Quarter 3

The priorities in Q3 are (a) bringing on-line the SSA, (b) continued development of curation tools following the SAD, (c) acquisition and installation of the pixel data storage hardware and (d) continued development of the user interface. We have included a task to exercise the SSA and will solicit feedback from potential users of the WSA. This exercise and shakedown, intended to test scalability of the design, will be all the more important if the delivery of WFCAM data to the VDFS is delayed into the period when we expect to be designing V2.0. Compared with the original plan, we have brought forward slightly the acquisition of the pixel data hardware.

The tasks are set out in the draft programme in the accompanying spreadsheet.

P. M. Williams
2003 July 3