WFCAM Science Archive Critical Design Review

Panel Report

Luc Simard (CADC, Chair)

Frossie Economou (JAC)

John Lightfoot (UKATC)

5 May 2003

1. Summary

The Critical Design Review of the WFCAM Science Archive (WSA) took place at the UK Astronomy Technology Center (UKATC) in Edinburgh on April 15-16, 2003. Review panel members were Frossie Economou (Joint Astronomy Centre), John Lightfoot (UK ATC), and Luc Simard (CADC, Chair). In addition to panel members and Wide-Field Astronomy Unit (WFAU) project staff, representatives from the VISTA project, the UKIDSS consortium, CASU and the Subaru Telescope were in attendance.

The Panel was very pleased by the efforts made by WFAU staff over the past months to produce a vastly improved set of documents for the WSA. Presentations by WFAU staff were very professional and instructive. We were also pleased by the work spent to determine hardware performance and design elaborate database relational models.

Section 2 gives recommendations from the Panel regarding the documents presented at the Review, and Section 3 goes over the terms of reference to determine if they have been met.

2. Section Reviews

A set of documents was submitted to the Panel for critical design review, and each document was reviewed following presentations by WFAU staff. This section presents the concerns and recommendations of the Panel.

2.1 Overview

The main issue in this section was the question of the UKIDSS Survey Manager. It was clear to the Panel and the other attendees that data quality information from both the summit pipeline and the Science Archive products needed to be fed back to the observing plan through the Survey Manager. Once a Survey Manager has been designated by the UKIDSS consortium, this person will have to provide detailed requirements to WSA project members on the type of survey progress information needed to conduct survey operations in a way that will meet survey goals. These requirements should be communicated to WFAU very soon especially if some survey progress information cannot be extracted from the Science Archive itself, and a separate database (and associated tools) must be designed and implemented to supply the missing information.

2.2 Science Requirement Analysis

In order to meet science requirement provided by UKIDSS, a list of archive content and functionalities was defined and presented, but this list was not prioritized. Priorities must be assigned to items in that list to help in identifying de-scoping options and their impact should a need for them arise. The Panel strongly felt that the hardware and software

needed to produce satisfactory version 1.0 pixel data and source catalogs and curate them should be the top priority.

The document presented a forward mapping from science requirement to archive requirements, but there is a need for backward mapping (i.e., archive to science) as well. Backward mapping will be important to evaluate the impact on science of any de-scoping of the archive.

The version 2.0 archive requirements will need additional inputs from V1.0 archive users, but the Panel was concerned to see that the schedule did not allow time for this process to occur.

2.3 Management and Planning

The project schedule, staffing and risk assessment were a major source of concern for Panel members.

The schedule fails to include time and resources for the following essential activities: (1) contingency plans, (2) recasting of the V1.0 Science Archive should it fails to meet some of its goals, (3) acceptance tests for the V1.0 Science Archive (this activity has an important external dependency given that UKIDSS scientists will have to participate.), and (4) completion of the design/analysis for V2.0.

The timeline for the project schedule is overall far too optimistic. As an example, V1.0 Science Archive integration and testing will certainly take longer (factor of 2 at least) than one month especially if this activity takes place over Christmas. The schedule should place reasonable demands on staff to maintain a positive and effective working environment.

The SSA prototype will be a very important test for the V1.0 Science Archive.

Staffing is not adequate. A 2.5 FTE effort for software development/implementation and hardware acquisition/deployment is uncomfortably low. Hardware alone could easily consume 1.0 FTE. Allocating 0.5 FTE for V1.0 operations will not be enough. A plan must be developed to juggle development and operations work. A “HelpDesk” person will be required following deployment, and this person will need to be a scientist.

Project risk assessment needs to be greatly expanded to be more realistic. The statement “Not on critical path for 1 month” is not a mitigation plan. In addition to likelihood and effect, each risk should also be assigned a “time penalty” factor to estimate its effect on the overall project schedule.

2.4 Interface Control

The Panel felt that interfaces were properly defined.

2.5 Data Flow

The current Science Archive design calls for PATT open-time programs to be processed and archived in V2.0. This is not acceptable given that many programs carried out soon after WFCAM's first light will be shared-risk, open-time programs. We recommend that open-time program processing and archiving be included in V1.0.

Data compression algorithms such as RICE in the CFITSIO library should be investigated. A compression factor of 2 would have a very significant effect on network transfers between CASU and WFAU and between WFAU and archive users as well as on pixel data storage requirements.

2.6 Hardware

There are no plans to acquire spares (systems *and* disks) for any of the proposed WSA subsystems (web server, SQL server, ingest server, pixel server, etc.). This was worrisome to the Panel, and we recommend that failure scenarios be studied so that a proper plan can be developed. IDE disks have a non-negligible failure rate especially for large systems such as the WSA.

The WSA is intended to be a *Science* archive. This requires a higher level of uptime/availability than a typical data archive. We recommend that metrics be developed to measure uptime/availability of current WFAU services to determine whether current uptime levels will be acceptable for WSA operations.

The discussion of trawl rates in the current documentation as well as tests performed by WFAU did not take into account concurrent use. Concurrent use has been a major source of problems for other big archives such as SDSS. It is therefore very important that tests be conducted as soon as possible to measure the impact of concurrent use on the proposed hardware solutions and develop more powerful solutions should they be necessary.

The Panel strongly felt that a software layer should be put in place to protect the database from "unreasonable" queries that could easily tie up all available resources and disrupt user access. The proposed use of direct SQL was not acceptable to the Panel. Direct SQL should not be used to make up for deficiencies in user interfaces. We recommend that form-based queries be used. Different forms could be designed and implemented for different UKIDSS surveys. We also recommend that a query queue mechanism be put in place to manage query priorities.

WSA should be connected to the JANET backbone to handle expected data transfer volumes.

We suggest that proxy data delivery be studied. Proxy servers eliminate the need for staging areas.

Given that WSA will be performing significant amount of processing for re-stacking, re-mosaicking, and astrometric/photometric re-calibrations, we were concerned that the current hardware plans did not include a CPU processing farm. Processing demands will rapidly grow as more and more survey data accumulate with time.

2.7 Database Design

We were pleased with this section. Considerable work was spent making sure proper relational models were in place. The Panel was told that iterations on the design would probably be necessary, but we note that no time has been allocated to this activity in the current project schedule.

We emphasize that portable code with clean interfaces must be written for the required software.

A number of curation tasks were described in this section, but it was unclear to the Panel how these tasks would actually be triggered, i.e. we did not understand how the list of items to curate would be tallied up for the curator (human or machine). Scheduling daily, weekly or monthly curation tasks does not take into account data associations within the observing programs. We recommend that a "discovery" mechanism be designed and implemented to help the curator.

Curation tasks will update important information such as astrometric and photometric calibrations, and the headers of images delivered to users should always contain the most up-to-date values. A mechanism to generate headers "on-the-fly" (i.e., when the image is requested) from values stored in the database will thus have to be put in place.

2.8 User Interface

The big issue in this section was data coverage. WSA will contain many different (WFCAM and non-WFCAM) datasets, and there must be a way for users to visualize the overlap between all those datasets so that users are not put through countless, unproductive queries that could quickly frustrate them. The WSA user interface must therefore include some means to measure and display survey coverage.

2.9 Other

An important component is completely missing from the current set of documents namely software architecture and design (libraries, classes, API's). The Panel therefore had no way to determine whether the proposed WSA design would end up with a large amount of un-maintainable code upon its completion. This has a direct impact on the path from V1.0 to V2.0. The temptation to write "quick and dirty" code will be very strong in a

project as tightly scheduled as the WSA, and this situation should be avoided. Choosing good, coherent software architecture and design practices from the start would make that path considerably easier and shorter.

3. Terms of Reference

The main task of the Panel was to determine whether the WFCAM Science Archive design put forward by WFAU members met eight terms of reference.

3.1 The end product is acceptable

Yes.

3.2 The project as conceived is feasible

The Panel had major concerns regarding the project schedule and risk assessment. Both were far too optimistic, and timescales could easily be scaled up by a factor of two.

3.3 Detailed plans for building the complete system are in place

A plan is in place, but important areas need to be better defined and planned. One important example is software architecture and software design procedures.

3.4 The budget and staffing resources are sufficient for the project and commensurate with the estimated time-to completion

Staffing is a concern especially after deployment. Too little staff was allocated for version 1.0 operations.

3.5 A reasonable estimate of the risk has been made, and whether there are useful de-scope options

Risk assessment needs to be planned in greater detail. No de-scoping options were presented although the Panel strongly felt that the hardware and software needed to produce satisfactory pixel data and source catalogs and curate them should be in place before any resources were allocated to the ingestion of external catalogs for example.

3.6 External interfaces are well controlled

Yes.

3.7 The path from version 1 to version 2 is understood and feasible

The path from version 1 to version 2 will be easier if sufficient attention is paid to proper software architecture and design procedures that will maintain portable code and clean interfaces between subsystems.