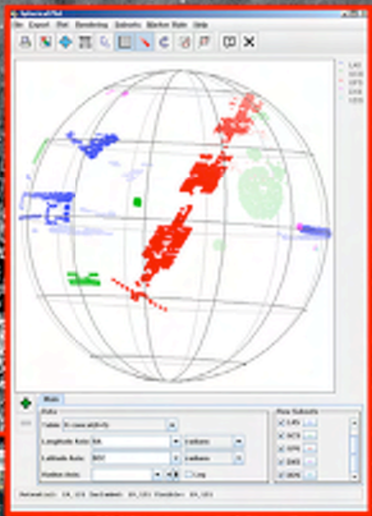
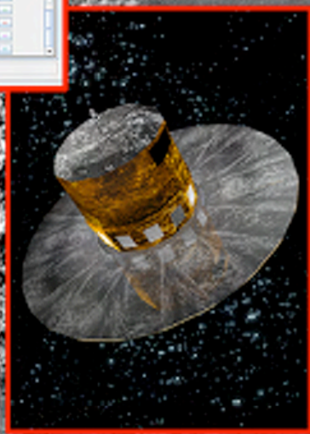
# Wide-Field Astronomy at Edinburgh





R.G. Mann
A. Lawrence
J.S. Dunlop
A.M.N. Ferguson
J.A. Peacock
W.K.M. Rice

## Cover Image

The background image shows a WFCAM mosaic of the Helix nebula taken in the H2 (2.12$\mu$m) filter. The inset panels show: top - a TOPCAT plot of the sky coverage of the UKIDSS surveys in DR2; middle row - (left) the Pan-STARRS 1 telescope, (middle) the VISTA telescope and (right) the Gaia satellite; and bottom - accessing UKIDSS DR2 data via AstroGrid.

# Contents

# Section 2 Summary Science Case

## 1 INTRODUCTION

We propose to undertake a survey astronomy programme of international calibre in support of the research priorities of the UK astronomical community. The programme of the Wide Field Astronomy Unit (WFAU) for the next five years centres on finishing development, and continuing operation, of the state-of-the-art Science Archives we have designed for WFCAM and VISTA, plus preparation for full UK exploitation of the next generation of world-leading sky surveys, namely Pan-STARRS, Gaia and LSST. This is a golden era of survey astronomy, and WFAU's experience in curating and publishing sky survey data, together with our developing expertise in the technologies needed to exploit them scientifically within the context of the Virtual Observatory, will help ensure that the UK astronomical community gains full benefit from the riches it has on offer.

## 2 WFAU ACHIEVEMENTS

The photographic programme that WFAU inherited from the UK Schmidt Telescope Unit has been successfully completed, followed by the decommissioning of SuperCOSMOS and the creation of permanent storage for the long-term preservation of the plate collection; both achieved at no cost to the WFAU rolling grant. We built the SuperCOSMOS Science Archive (SSA) as a prototype for the WFCAM and VISTA Science Archives, and the SSA is now an important scientific resource itself, with $10^9$ rows of data extracted during its first three years of operation. Our other SuperCOSMOS–based archives remain actively used (see Figure 1) and a total of 192 refereed and 77 non-refereed papers based on SuperCOSMOS data have now been published.[1] We will soon release the final SSA, which will be a unique homogeneous all-sky multi-colour, multi-epoch digital optical sky survey, with high precision astrometry and enhanced photometric calibration via 2MASS and SDSS.

WFAU led the development of the science cases for UKIDSS and AstroGrid, and we are seeing these ideas through to fruition, via robust, well-engineered implementation. This has required the development within the Unit of a new skill set, obtained through a combination of hiring new staff with the relevant IT expertise and on-the-job training of the existing team. WFAU now boasts the expertise in astrophysics, computer science, hardware technology, software engineering, database design and curation, and professional project management it requires as a world-class data centre.

The key to WFAU's success is belief in its public service remit, coupled with strong local scientific interest in all of its projects. The success of the WFCAM Science Archive (WSA) has shown that we deliver interfaces which are both user-friendly and provide powerful access to data. WSA availability (>96%) exceeds the requirement[2] for the VISTA Science Archive (VSA) and the WSA helpdesk system has been well used, and provides quick responses to user queries. We have worked closely with the UKIDSS Survey Heads to develop thorough Quality Control procedures, and are currently liaising with the five (out of six) VISTA Public Survey PIs who have requested that we curate their data, to ensure that we shall deliver a service meeting their requirements and those of ESO. Local scientific interest, on the part of the Investigators, and fostered through modest research allocations for a few key staff, ensures that WFAU's work is driven by the requirements of scientific excellence, while our leading role in the developing Virtual Observatory, together with our focus on providing data analysis services as part of our Science Archives, will ensure that the UK community can continue to exploit fully the sky survey datasets we shall curate in the future.

[1] See http://www.roe.ac.uk/ifa/wfau/cosmos/newpubs/newpubs.html.

[2] The WSA itself has no numerical availability requirement.

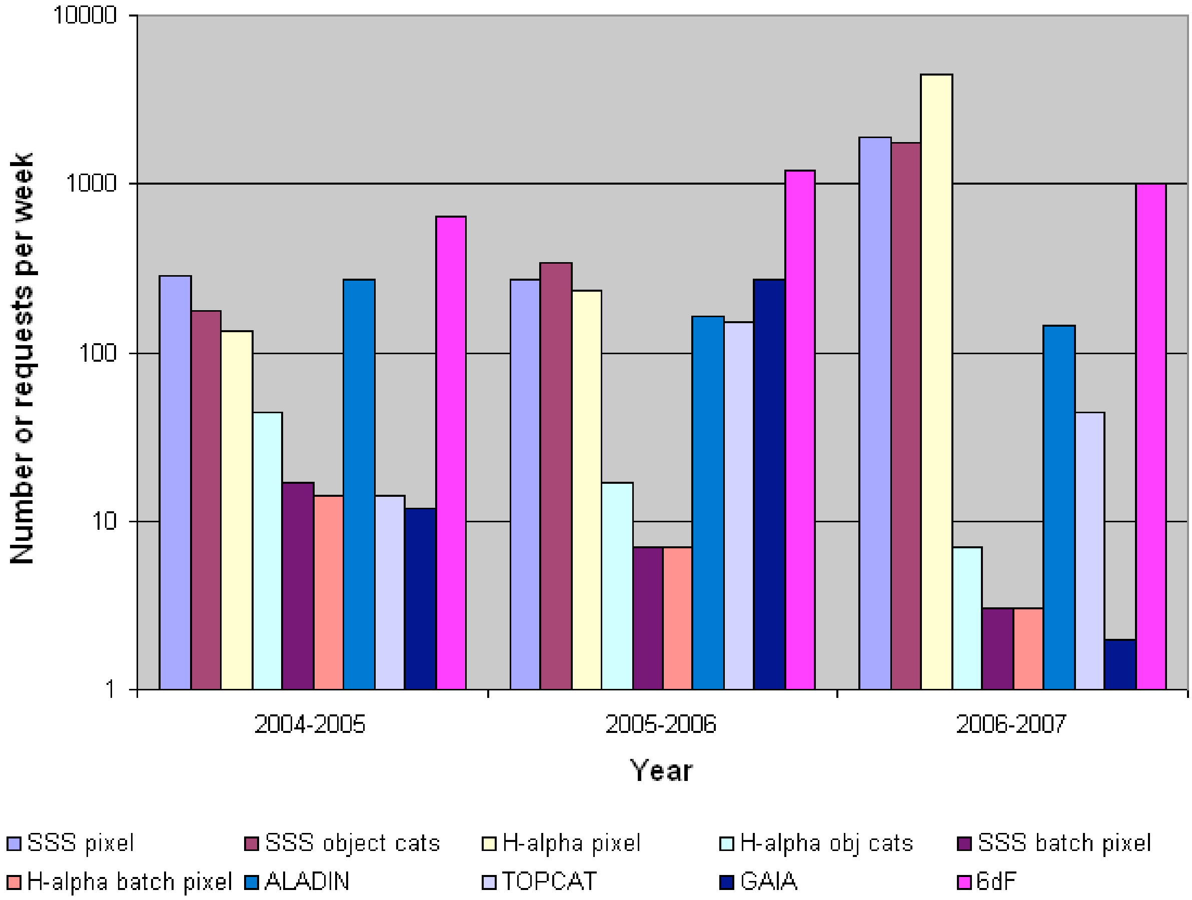



Figure 1: *Usage statistics for WFAU-curated archives derived from SuperCOSMOS data, classified by data type and access method. For each year there are ten categories, reading from left to right as follows: (i) requests made through the online webform for SuperCOSMOS Sky Surveys (SSS) pixel data; (ii) SSS object catalogues retrieved the same way; (iii) SuperCOSMOS Hα Survey (SHS) pixel data from the online form; (iv) SHS object catalogue data from the online form; (v) batch script requests for SSS pixel data; (vi) batch script requests for SHS pixel data; (vii) SSS and SHS data queries from the Aladin image and catalogue analysis tool; (viii) SSS catalogue data queries from the TOPCAT catalogue analysis tool; (ix) SSS data requests from the Starlink GAIA tool; and (x) SSS pixel data retrieval using a batch script supplied to the 6dFGRS and RAVE teams to generate finding charts in preparation for spectroscopic observations.*

# 3 THE WFAU PROGRAMME

We briefly describe the full WFAU programme below, indicating those components for which we request funding here. We expect that WFAU will continue to hold a portfolio of project-specific grants in the future – as for AstroGrid, VOTECH, and the VISTA and Gaia Data Flow Systems, at present – but that the rolling grant will support our core data curation and technology R&D activities, so we retain the expertise necessary for us to continue taking on new projects in a cost-effective manner.

**WFCAM/VISTA Science Archive development: final funding requested.**
Science Archive development for WFCAM and VISTA has been undertaken by the VISTA Data Flow System (VDFS, a component of the VEGA (VISTA + GAIA) proposal) project, with funding from PPARC's e-Science programme. The WSA is now in its operations phase, and our plans to scale it up for VISTA have been endorsed by the VDFS Final Design Review. Experience with

WFCAM suggests that development effort will continue to be required for some time after survey operations begin, as the whole system – from telescope to archive – gets shaken down using survey data. We therefore request a final 18 months of effort to complete development of the VISTA Science Archive (VSA), whose detailed scientific requirements are only now being finalised through the ESO Public Surveys process: this case is presented in Section 4(a).

**WFCAM/VISTA Science Archive operations and analysis services: funding requested.** WSA/VSA operations were judged the highest priority portion of the WFAU programme by the Panel assessing our last rolling grant renewal and this will certainly remain one of our core activities for many years to come. We have now been operating the WSA for two years, and that experience, and our analysis of the requirements from ESO's VISTA Public Surveys programme in the light of WSA experience, enable us to make an informed assessment of the resource requirements for archive operations over the grant period: this is detailed in Section 4(b). One of the main lessons from the WSA is the importance of our supporting survey PIs in conducting Quality Control procedures, and we have now defined a Survey Quality Assurance Scientist position within the Unit, in recognition of the significance of this activity. The value of an archive is determined ultimately by its effectiveness in facilitating the scientific exploitation of the data it holds, and it is clear that, with large-scale statistical analyses driving the development of the coming generation of sky surveys, the role of the data centre must expand, to encompass the provision of the hardware and software to support core data analysis operations *in situ*. This will prevent duplication of effort in the development of software to perform basic operations underpinning many analyses, and becomes necessary once data volumes are too large for users to download all the data they need for analysis in their own institutions. Section 4(d) presents our case for continued development of core data analysis services.

**Virtual Observatory development: already funded**
WFAU continues to play a leading role in the development of the Virtual Observatory (VO): Lawrence leads both AstroGrid, the UK's VO project, and VOTECH, an EU-funded design study for the future European Virtual Observatory; Mann is leader of one of VOTECH's constituent design studies, assessing how Data Exploration (=Data Mining + Visualization) should be undertaken in the VO; and there are currently four AstroGrid/VOTECH-funded staff in WFAU. The interaction between the science archive and VO teams within WFAU benefits both significantly: for example, WFAU archives have been amongst the first to be integrated into AstroGrid and, conversely, the WFAU science archive team have helped influence the design of AstroGrid components and elucidate practical issues relating to their installation in data centres. WFAU's work on VO development is all funded from outwith the WFAU rolling grant, although the cases made in Sections 4(a) and 4(b) for support of Science Archive development and operations do include development of science archive software adhering to VO standards and installation of AstroGrid components, respectively.

**Gaia Data Flow System development: already funded**
As part of the PPARC/STFC-funded Gaia Data Flow System (GDFS) project, we are developing the data management facility needed for Gaia photometric data processing. This requires design of a system which can support the high data volumes and iterative data processing which characterise Gaia photometric analysis and involves interaction with the Gaia central database at ESAC, as well as the other institutions across Europe contributing to CU5, the photometric processing portion of the Gaia Data Processing and Analysis Consortium. As shown in Section 3, this activity starts in Q3 2007, and is currently funded to the end of Q2 2011.

**Science Archive development for synoptic sky surveys: funding requested**
Technological advances are producing survey systems with very high étendue, capable of covering wide areas of sky to interesting depths each night. Such systems can be used not only to build deep, large area surveys, but also to provide a time-resolved view of the sky, by building up integration time on each field through multiple exposures, with the possibility of making those observations

with a cadence optimised for detection of particular time-varying phenomena. These synoptic sky surveys clearly represent the next development in wide-field astronomy. They make possible a wide range of analyses requiring deep and/or time-resolved data, but pose challenges to data centre staff, due to the very high data rates and the new, temporal analysis modes which must be supported by Science Archives. Section 4(c) presents our case for support of WFAU's initial involvement in projects providing public access to data from three of the most important future synoptic sky surveys: Pan-STARRS, Gaia and LSST. (Pan-STARRS is initially a private survey, but the data will become public well within the lifetime of this grant, as discussed in Section 4(c).) These are all world-leading projects, and we intend that provision of full and ready access to their data to the UK community will become a core activity for WFAU after its next rolling grant renewal.

**Project-related research: funding requested**
As noted above, the range of activities undertaken by WFAU requires staff with a range of skills. A crucial part of this mix is a small number of research-active astronomers, who are fully cognizant of the detailed scientific requirements of our projects and able to interact effectively with the external astronomers leading them. Section 4(e) details our request for 0.3 FTE allocations for three such key staff to conduct research programmes using the data we curate. Experience shows that these allocations are necessary for recruiting and retaining high calibre staff, as well as providing them with the strongest motivation for ensuring the highest quality in their project duties.

**Project Management: funding requested**
Section 4(f) presents our case for Project Management support. Unlike in previous WFAU rolling grant proposals, we have written a separate case for this activity this time, to provide justification for the increased management support we request for the later stages of the grant period. This is required for two reasons: firstly, as WFAU becomes involved in more international projects, the project management overhead from planning, reporting and financial procedures becomes greater, and, secondly, the computational requirements we shall face will increase dramatically with the start of our synoptic sky surveys activities. We propose, therefore, to combine a continued 0.5 FTE Project Manager role with a new 0.5 FTE Technical Lead position, to provide a full-time post (from Q4 2009) which we would anticipate filling by recruitment from the commercial IT sector.

# 4 PERSONNEL

We supply, in Section 3, a chart which summarises the personnel deployment across projects during the lifetime of the grant. Its apparent complexity is a result of our policy of phasing staff effort, so that new projects are started by experienced staff, whose effort is gradually replaced by less experienced staff over time, thereby ensuring that the right expertise is available at the right time.

**WFAU Personnel Changes**
*(i) Departures*
Several WFAU staff have left, following the completion of the photographic programme: Allan left once the photolabs ceased operation; and MacGillivrary and Thomson left upon completion of plate-scanning by SuperCOSMOS. Tritton has left the WFAU grant, and is currently supported by a temporary contract to oversee the relocation of the plate collection. Hill, an AstroGrid/AVO developer left in December 2005, at the end of his contract, while two remotely-located members of WFAU's VOTECH team Smith (Portsmouth) and Holbrey (Leeds) left during 2006. Williams will retire during the period of the new grant.
*(ii) Arrivals*
Bryant took up a Science Archive operations post in September 2005, and Holliman started a sysadmin position jointly funded by the WFAU rolling grant, VOTECH and AstroGrid in March 2006. Andrews moved from Cambridge to Edinburgh in August 2006, continuing her work on AstroGrid.

Walshe joined the VOTECH team in October 2006, and Rimoldini took up an appointment on the WFAU rolling grant in November 2006.

*(iii) Proposed New Appointments*

We intend to appoint a new developer for the Gaia Data Flow System (GDFS) project in Q4 2007, who will initially work on the VDFS to free up experienced effort from Hambly and Collins to get the GDFS work going. This proposal seeks funding for three new appointments during the course of the new grant: (a) a half-time Technical Lead, to be appointed in Q4 2009, who will also take over Williams' half-time project manager post; (b) a synoptic sky surveys scientist, with a 30% research allocation, also to be appointed in Q4 2009; and (c) a developer to work on synoptic sky surveys, from Q2 2010.

# 5 SUMMARY OF RESOURCES REQUESTED

**Funded Staff Effort:** We request support for a total of 39.3 d.s.y. of staff effort across six themes and at varying levels across the grant period, plus sysadmin and secretarial support at average levels of 0.7 and 0.5 FTE, respectively. A detailed profile of project staff effort across the grant period is shown in Section 3. The Investigators contribute an additional 2 d.s.y. of effort over the lifetime of the grant.

**Equipment:** We request archive hardware to the tune of £250K, plus 10% for spares (disks, power units), as detailed in Section 4(a). Personal computer equipment is counted as Consumables, as no individual item exceeds £2K.

**Other costs:** Distributed projects involve significant amounts of travel, so we request a total of £89.6K for Travel & Subsistence. This covers attendance of conferences, oversight committee and project consortium meetings, as well as travel associated with planning new projects. We request a sum of £101.5K for Consumables: of this, £47K is for hardware maintenance, and the remainder for computing supplies, renewal of personal computers, and other sundries.

# Section 3 Summary Staff Profile by Theme

We provide overleaf a chart summarising the evolution of the WFAU staff profile, distributed according to theme, to illustrate the phasing of effort between the themes described in the following Sections. The profiling of staff appointments is shown in the Gantt charts with the supplememtary material, in Section 5.

We include the separately funded Gaia data-flow system (GDFS) work, to show staff effort moving on to, and off, this work during the period.

We include the effort of Holliman (Sysadmin) and Dupin (Secretary), based on 0.1 FTE and 0.08 FTE per FTE in the program averaged over the five-year period as discussed in Section 2. Not shown are the AstroGrid and VOTECH staff, whose effort does not overlap with the work funded by this grant, apart from small amounts of Holliman and Williams in the early part of the period.

Funding requested

## Staffing (FTE) by Theme

| | Calendar year | 2007 | | | 2008 | | | | 2009 | | | | 2010 | | | | 2011 | | | | 2012 | | | | 2013 | Total staff-years by Theme | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Quarter | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | (a) | (b) | (c) | (d) | (e) | (f) |
| | **Grant Year** | **Current (2007-08)** | | | | **Year 1 (2008-9)** | | | | **Year 2 (2009-10)** | | | | **Year 3 (2010-11)** | | | | **Year 4 (2011-12)** | | | | **Year 5 (2012-13)** | | | | **VDFS** | **Ops** | **SynS** | **Data** | **Res** | **Mgt** |
| **Hambly** | (a) VDFS development | 0.7 | | 0.2 | | | | **0.2** | | | | | | | | | | | | | | | | | | 0.3 | | | | | |
| | Gaia DFS | | | 0.5 | | | | | | | | 0.5 | | | | | | | | | | | | | | | | | | | |
| | (e) Research | | 0.3 | | | | | | | | | | | | **0.3** | | | | | | | | | | | | | | | 1.5 | |
| | (c) Synoptic Surveys | | | | | | | | | | | | | | **0.2** | | | | | | | **0.7** | | | | | | 1.62 | | | |
| **Cross** | (a) VDFS development | | 0.7 | | | **0.7** | | | **0.2** | | | | | | | | | | | | | | | | | 0.55 | | | | | |
| | (e) Research | | 0.3 | | | | | | | | | | | | **0.3** | | | | | | | | | | | | | | | 1.5 | |
| | Gaia DFS | | | | | | | | | | | | 0.5 | | | | | | | | | | | | | | | | | | |
| | (c) Synoptic Survs | | | | | | | | | | | | | | **0.2** | | | | | | | **0.7** | | | | | | 1.62 | | | |
| **Collins** | (a) VDFS development | 1.0 | | 0.5 | | **0.5** | | | | | | | | | | | | | | | | | | | | 0.25 | | | | | |
| | GAIA DFS | | | 0.5 | | 0.5 | | | | | | | 1.0 | | | | | | | | | | | | | | | | | | |
| | (c) Synoptic Surveys | | | | | | | | | | | | | | | | | | | | | **1.0** | | | | | | | | | |
| **VEGA developer** | (a) VDFS development | | | 1.0 | | **1.0** | | | **0.5** | | | | | | | | | | | | | | | | | 1.0 | | | | | |
| | Gaia DFS | | | | | | | | **0.5** | | | | | | **1.0** | | | | | | | | | | | | | | | | |
| **Read** | (a) VDFS development | | 1.0 | | | | | **0.5** | | | | | | | | | | | | | | | | | | 0.75 | | | | | |
| | (b) Survey Quality Ass | | | | | | | **0.5** | | | | | | | | | | **1.0** | | | | | | | | | 4.25 | | | | |
| **Sutorius** | (a) VDFS development | 1.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | (b) Archive Operations | | | 1.0 | | | | | | | | | | | **1.0** | | | | | | | | | | | | 5.0 | | | | |
| **Bryant** | (b) Archive Operations | | 1.0 | | | | | | | | | | | | **1.0** | | | | | | | | | | | | 5.0 | | | | |
| **Rimoldini** | Data analysis services | | 1.0 | | | | | | | | | | | | **1.0** | | | | | | | | | | | | | | 5.0 | | |
| **New1** | (c) Synoptic Surveys Sci | | | | | | | | | | | | | | | | | **0.7** | | | | | | | | | | 2.45 | | | |
| | (e) Research | | | | | | | | | | | | | | | | | **0.3** | | | | | | | | | | | | 1.05 | |
| **New2** | (c) Synoptic Surveys Dev | | | | | | | | | | | | | | | | | | **1.0** | | | | | | | | | 3.0 | | | |
| **Williams** | VDFS Manager | | 0.3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | WFAU Manager | | 0.2 | | | | | | **0.5** | | | | **0.2** | | | | | | | | | | | | | | | | | | 0.92 |
| **New3** | (f) Project Manager | | | | | | | | | | | | | | | | | **0.5** | | | | | | | | | | | | | 1.75 |
| | (f) Technical Lead | | | | | | | | | | | | | | | | | **0.5** | | | | | | | | | | | | | 1.75 |
| **Holliman** | WFA System Manager | | 0.3 | | | | | | | | | | | | **0.7** | | | | | | | | | | | | | | | | |
| | GAIA DFS System Mgr | | | | | | | | | | | 0.2 | | | | | | | | | | | | | | | | | | | |
| **Dupin** | GAIA DFS Secretarial | | | | | | | | | | | 0.2 | | | | | | | | | | | | | | | | | | | |
| | Secretarial / Clerical | | | | | | | | | | | | | | **0.5** | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | 2.9 | 14.3 | 8.7 | 5.0 | 4.1 | 4.4 |

Gaia DFS funded separately but included to show funding evolution

# Section 4 Cases for Support by Theme

4(a) VDFS Science Archive Development

4(b) WFCAM/VISTA Science Archive Operations

4(c) Science Archiving for Synoptic Sky Surveys

4(d) Data Analysis Services

4(e) Project-Related Research

4(f) Project Management

# Section 4(a) : VDFS Science Archive Development

**Staff involved**

N. Hambly
N. Cross
R. Collins
M. Read
VEGA developer (from Q2 2008)

## 1 BACKGROUND

The VISTA Data Flow System (VDFS[1]; see Emerson et al. 2004) programme consists of providing systems engineered end–to–end data flow for transfer, pipeline processing and science archiving of WFCAM and VISTA data from the respective observatories through UK data centres to the end users. Pipeline processing is handled at the Cambridge Astronomy Survey Unit (CASU), while WFAU concentrates on the provision of science archives. VDFS development was a self–contained project funded at e–science round 1, and the allocated resources for science archiving were distributed via an award supporting a case incorporated within the WFAU grant renewal of 2002. Support for VDFS development was reviewed and further funds awarded at e–science round 2 as the major component (with GAIA data processing) of the VEGA proposal. Combined with a small (6 month) extension to the second e–science award (primarily for scaling the existing WFCAM system to VISTA data volumes and in recognition of the slip in VISTA first–light) VDFS science archive development has been funded via e–science monies from Q4 2002 through to the end of Q1 2008. At an average level of around 3 FTE per year, the resulting science archive system (see below) represents the culmination of $\sim 15$ staff years of effort.

Throughout the bidding process for VDFS development funds, we have consistently argued that resources are necessary beyond first–light since a period of deployment and shake–down of software systems using real data is vital. Our approach of using WFCAM as a test–bed for VISTA has, as far as possible, minimised this period of phasing from development to operations (described in Section 4(b) of this proposal). Nonetheless, here we present a final case for development funds for the VISTA Science Archive to cover an 18 month period after survey operations begin (assumed to be during Q1 2008) in order to deploy a fully scaled and shaken–down archive system to service the requirements for VISTA.

The justification for this request comes from two major reasons. Firstly, our experience in developing and deploying the science archive for WFCAM has shown that user requirements only become finalised after a first iteration with *real* survey data; similarly, quirks of the data only become apparent in real survey data. Secondly, for VISTA, the ESO Public Survey (EPS) designs are only crystalising now, and user requirements are still under review (see later). This is leading to new requirements over and above those already covered by the WFCAM science archive system defined in 2002.

## 2 REPORT ON RECENT WORK: 2004 – 2007

We reiterate that VDFS science archive development resources have been provided to date from the e–science line. Nonetheless, in this Section we report briefly on work done with those resources, despite their being outwith previous rolling grant awards.

[1] http://www.maths.qmul.ac.uk/~jpe/vdfs/

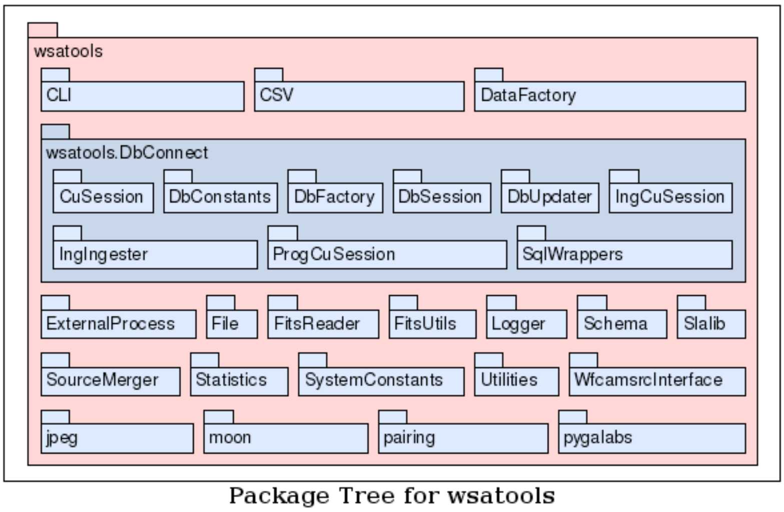



Figure 1: *Package organisation for the WSA Python module "wsatools" (see text).*

The concept of the science archive (e.g. Gray et al. 2002) implies much more than a simple data repository – it is a system whereby science exploitation of very large datasets is facilitated by well–designed curation procedures and provision of flexible user interfaces. The increasing importance of science archives to the user community is demonstrated by the usage being made of such facilities (e.g. for the SDSS and indeed our own VDFS archive for WFCAM – see Section 4b). VDFS science archive development work consisted of detailed data flow, data modelling and software design studies in the context of a phased delivery of prototype, basic and full–blown archive systems: our legacy SuperCOSMOS Science Archive (SSA[2]), WFCAM Science Archive (WSA[3]) and the VISTA Science Archive (VSA) respectively. The final system underwent a formal Final Design Review (FDR) in October 2006 (following on from a successful Critical Design Review of the prototype system in April 2003) for which we delivered a set of design documents (requirements analysis, management and planning, interface control, hardware design, database design, user interface design, software architecture and virtual observatory integration) for review by an external panel. All these documents are available for download from the development web site[4], and their contents are summarised in conference papers presented by Hambly et al. (2004a; 2004b), Collins et al. (2006) and Cross et al. (2007). The review panel reports are also available from the development web site (the FDR panel report in particular praised the VDFS design).

In addition to the documentation and archive system deliverables achieved within the reporting period, there is, of course, the associated software. A large amount of curation, and web interface, applications code has been written in Python, Java, C/C++ and SQL, where we chose the most appropriate language for a given application set (i.e. to take advantage of the respective strengths of the different languages). The code is contained in a centralised repository under formal revision control using the industry–standard Concurrent Versions System (CVS). The code is self–documented using in–line comments and 'ePyDoc', 'JavaDoc' and 'Doxygen' tools. The CVS repository server is not externally browsable for security reasons, but the current set of source code documentation is available online[5]. For example, Figure 1 shows the package organisation for the Python module "wsatools" available online[6] (see the associated description for further details).

[2] `http://surveys.roe.ac.uk/ssa`
[3] `http://surveys.roe.ac.uk/wsa`
[4] `http://www.roe.ac.uk/~nch/wfcam/index.html`
[5] `http://www.roe.ac.uk/~rsc/wsa`
[6] `http://www.roe.ac.uk/~rsc/wsa/Epy_Wsa/wsatools-module.html`

# 3 PROPOSED PROGRAMME OF WORK Q2 2008 – Q3 2009

The following programme of work is based on the VISTA UK User Requirements document (Sutherland 2005; hereafter the URD), the science requirements analysis document presented for the VDFS FDR (Hambly et al. 2006; hereafter the SRAD), the resulting review panel report (Simard et al. 2006; hereafter the Panel report), and the most recent revision of VISTA EPS requirements from the Survey Management Plans being prepared (in consultation with VDFS) at the time of writing. With reference to Table 2 in the SRAD, we note that the PIs of the VISTA Magellanic Clouds (VMC), the VISTA Variables in the Via Lactea (VVV), the VISTA Hemisphere Survey (VHS), the VISTA Kilo–degree Infrared Galaxy (VIKING) and VISTA Deep Extragalactic Observations (VIDEO) surveys have all expressed the wish to use the VDFS VISTA Science Archive (VSA) as the primary point of dissemination of their survey data.

Tasks to be carried out, roughly prioritised in order of decreasing difficulty are as follows:

**1: Curation of underexposed strips in VISTA tiles**
The URD gives details of the focal plane layout for the VISTA infrared camera, and in particular gives an exposure map for a typical 6 paw–print filled tile. Because the detectors are spaced by 50% of a detector width in one direction, the tile exposure map contains two strips, at a pair of opposite edges, that are underexposed by a factor 2× with respect to the rest of the tile area. For maximum survey speed, some EPS PIs are considering a tiling strategy that overlaps these underexposed areas between adjacent tiles to yield combined full exposure in the strips. This results in the requirement for a generalised stacking and catalogue extraction procedure running at the archive end (as opposed to during nightly pipeline processing) for wide, shallow surveys that was not required for the WSA. An archive–end, database–driven solution to this problem is needed since, at the observing block level, during nightly processing, there is, of course, no guarantee that the required adjacent tile observation is available to be stacked.

**2: Interface with the ESO Science Archive Facility**
To date, VDFS science archive design has been based on the premise that the WFAU archives (both WSA and VSA) would be the end–point in the data flow, with individual end–users accessing processed data from servers located at Edinburgh (raw data being available from ESO/Garching in common with all other ESO telescopes). Recently, however, a requirement has emerged for VISTA EPS primary data products (i.e. photometrically and astrometrically calibrated FITS science images and associated single passband and band–merged FITS binary table catalogues) to be delivered to a centralised ESO Science Archive Facility (SAF[7]). Note that despite the name, the ESO–SAF is not a *survey* science archive as we have designed for the VDFS – rather it is a fairly static repository of basic calibrated data products from all ESO telescopes/instruments. Nonetheless, it is the responsibility of every individual EPS PI to deliver such survey data products to the ESO–SAF central repository; however we would argue that is more cost–effective and simpler for a single systems–engineered solution to this requirement – exactly what VDFS can provide. We propose to develop, in collaboration with our counterparts at the ESO–SAF, a controlled interface and semi–automated bulk ‘outgest’ applications for database metadata and science image/catalogue products to service this requirement for VISTA EPS PIs wishing to use VDFS.

**3: Enhanced user completeness tests**
The SRAD specifies (e.g. Sections 5.6.2 and 5.7) that VDFS should provide both simple completeness estimates and detection probabilities and/or completeness functions. Currently, provision of the simplest possible empirical completeness is made, for example, via number–magnitude counts in Structured Query Language (SQL). However, in the most general case it would be advantageous, for example, if synthetic images could be run through the processing and archiving systems (see

[7] `http://archive.eso.org`

also URD science example 11.3). This requires significant enhancements to the science archive user interface, and provision of new curation applications that, once again, were not required of the WSA.

**4: Image stacking options**
The existing science archive system includes provision of stacking during normal curation of deep surveys using a VDFS (CASU–supplied) toolkit application. Section 5.4.2 of the SRAD, however, specifies further user–selectable stacking options; moreover certain end–user requirements (e.g. Jarvis 9 in the Panel report) would even go so far as to suggest on–the–fly stacking in the User Interface (e.g. for the best possible image thumbnail in an arbitrary, user–specified region). Once again, neither requirement was specified for the WSA design, but both imply significant coding enhancements which we propose to make.

**5: Pixel analysis code**
Section 5.7 of the SRAD (along with URD 11.17) specify the ultimate aim of upload of user code to run server–side for wholesale pixel reanalysis. This is an ambitious undertaking for VDFS; furthermore it remains to be seen what is the likely demand for such a facility. As a significant step towards this ultimate goal, we propose to facilitate user–specified wholesale pixel analysis via upload of parameter configuration files for standard source extraction software (e.g. S–Extractor, Bertin & Arnouts 1996). Example applications that have been requested by users are a galaxy photometry package to provide enhanced extraction parameter attributes (e.g. half light radii, model profile fit parameters) and model point spread function fitted photometric attributes. We propose to support user–led developments by collaborating in their development and implementing the final packages at the archive end.

**6: List–driven photometry**
The concept of list–driven photometry (also known as 'co–located' photometry in VDFS parlance) is described in, for example, Irwin et al. (2007). Briefly, optimal aperture photometry is performed at positions, and with apertures, externally defined by a master list across a series of images. This is essential, for example, for constructing consistent photometry across a set of different passband images (e.g. for object extended source colours) or in constructing precise relative photometry across a run of distinct epochs in the same filter/field (e.g. for light–curve generation). Irwin et al. (2007) demonstrate the power of this technique, which is capable of producing accurate results even in relatively crowded fields. In line with FDR review recommendations (e.g. Simard 17 in the Panel report) we propose to implement this CASU–supplied analysis tool at the archive end in lieu of much more intensive (from the pixel processing standpoint) difference image analysis techniques for the synoptic components of several of the VISTA EPSs, e.g. light–curve generation for variables in the VVV and VMC surveys. List–driven photometry may also be implemented across filter passbands for consistent colours in other surveys, depending on user requirements.

**7: Enhanced object catalogues**
Section 5.6 of the SRAD specifies requirements to flag known fast–moving solar system objects (minor planets) in VISTA object catalogues, and furthermore to automatically detect, and flag, detections of any asteroids, since these can confuse usages requiring to find high proper motion stars and/or sources having extreme colours. Moreover, our experience of operating the WSA has shown that generic quality error–bit flagging[8] can be implemented only based on experience with real data. It is impossible to predict beforehand certain data handling issues until real survey data begins to flow through the VDFS, and developments are inevitably required in archiving (in addition to pipeline processing) to react to unpredictable changes and features in the data. We propose to implement moving object flags and enhanced quality error–bit flags at the archive end as necessary during the first 18 months of real survey data flow for the VISTA EPSs.

[8] `http://surveys.roe.ac.uk/wsa/ppErrBits.html`

**8: Complementary optical surveys**

Many usages of the VISTA infrared EPS data sets require complementary optical imaging (conversely, many usages of optical surveys require complementary infrared imaging). The VDFS science archive team have been approached by the PI of the VLT Survey Telescope (VST) EPS 'ATLAS' (Shanks) to curate and disseminate the survey data products from this optical survey, not least because of our experience in operating the WSA. The VDFS science archive design includes a high–performance and flexible cross–matching scheme between VDFS–curated surveys and/or externally–produced, static catalogue products[9]; other VST EPS PIs (e.g. Kuijken for the KIlo–square Degree Survey, KIDS) have approached VDFS to discuss inclusion of their catalogues in the science archive. The resource requirements are dominated by ATLAS, however, since we are required to curate the survey (e.g. ingest standard nightly pipeline–processed products; produce a seamless merged source catalogue etc). We propose to adapt and deploy survey data models and database–driven curation applications (e.g. source merging) in order to curate the ATLAS complementary optical imaging VST EPS.

**9: AstroGrid deployment**

The VDFS science archive development team maintain close links with the VO development team within WFAU (both UK AstroGrid and European VOTECH). For example, members of both development teams routinely attend each others' weekly project team meetings. The VDFS science archives have been early adopters of VO techniques, protocols and infrastructure: user–selectable output format options include VOTable[10]; archive metadata schemas includes VO–style Unified Content Descriptors[11]; WSA user groups are organised into VO–style 'communities'[12] ready for mapping onto AstroGrid authentication/authorisation infrastructure; user query results pages offer a facility to save results to the AstroGrid 'MySpace'; and most recently we have deployed secure DataSet Access (DSA) infrastructure alongside the WSA servers. All this enables, for example, inclusion of UKIDSS data in workflows generated within the AstroGrid workbench. We propose to continue this deployment over the first 18 months of VISTA survey operations as AstroGrid/VOTECH infrastructure developments mature to fully integrate the VDFS science archives into the Virtual Observatory.

[9] `http://surveys.roe.ac.uk/wsa/sqlcookbook.html#crossmatch`
[10] `http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOTable`
[11] `http://www.ivoa.net/Documents/latest/UCD.html`
[12] `http://surveys.roe.ac.uk/wsa/login.html`

# 4 KEY DELIVERABLES and MILESTONES

The key deliverables (D) and milestones (M) of the VDFS Science Archive development theme up to the end of the period covered by this grant renewal are as follows:

| Date | Type | Details |
|---|---|---|
| Q4 2007: | M | VISTA first light |
| Q1 2008: | M | VISTA EPS data flow begins |
| | | VISTA Science Archive operations begin |
| Q1 2008 (end): | M | e–science funded VDFS development ends |
| Q2 2008: | D | Code and documentation for underexposed strip stacking/cataloguing |
| | D | Code and documentation for user completeness tests |
| Q3 2008: | M | First release of EPS prepared databases |
| | D | Code and documentation for image stacking options |
| | D | Code and documentation for pixel analysis (catalogue extraction) options |
| | D | Code and documentation for ESO–SAF interface software |
| Q4 2008: | D | Code and documentation for list–driven photometry |
| Q1 2009: | D | Code and documentation for enhanced object catalogue flagging |
| Q2 2009: | D | Release of documented VST ATLAS curation database |
| | M | Completion of VO infrastructure deployment |
| | D | Release of fully VO-integrated VSA |
| Q3 2009 (end): | M | Phase–over from VDFS development to operations complete following shake–down of deployed system |

# 5 RESOURCES REQUESTED

We request staff resources, travel (for attendance of management, survey consortium, and ADASS meetings, plus trips to ESO) and consumables, as follows:

(a) *Staff.* 2.9 d.s.y. over 18 months from Q2 2008
(b) *Travel and subsistence.* £2K/FTE/year = £5.8K
(c) *Consumables:* £1K/FTE/year = £2.9K
(d) *Maintenance:* none required
(e) *Equipment:* none required

The total staff resource in (a) is spread as follows: Hambly 0.3 staff years; Cross 0.6; Collins 0.25; Read 0.75; and finally 1.0 continuing funding for the VEGA developer being appointed in Q3 2007. We reiterate that the deployment of staff effort on this and the GAIA Data Flow System project (funds provided outwith this grant; details can be found elsewhere in the current proposal) features a phased transition to ensure continuity and transfer of expertise.

The figure of 2.9 staff years is the minimum required to do the VSA work outlined in Section 3. As stated above, our experience of doing similar work for the WSA over the first 18 months of WFCAM survey operations is that just under 3 FTE was necessary over that period (i.e. 4.5 staff years). The reduced staff level is commensurate with the gains in having already shaken down generic curation and interface software that is applicable to all VDFS archive systems. During the work period, we anticipate that Cross will be promoted from grade 7 to 8. It is vital that some experienced staff (as opposed to all new, junior staff) are retained for this work to benefit from the work already done for the WSA. Staff costings for Cross reflect this – see the JeS–1 for details.

# 6 ADDITIONAL REFERENCES

Bertin, E. & Arnouts, S., 1996, A&AS, 117, 393

Collins, R.S., Cross, N.J.G., Hambly, N.C., Mann, R.G., Read, M.A., Sutorius, E.T.W., Williams, P.M., Bond I.A., 2006, In: Proceedings of the 15th ADASS meeting, ASP. Conf. Ser., 351, 743

Cross, N.J.G. et al., 2007, In: Proceedings of the 16th ADASS meeting, ASP Conf. Ser., in press

Emerson, J.P. et al., 2004, In: Optimizing Scientific Return for Astronomy through Information Technologies, Proceedings of the SPIE, 5493, 401

Gray, J., Szalay, A.S., Thakar, A.R., Stoughton, C., Vandenberg, J., 2002, "Online scientific data curation, publication and archiving", In: Virtual Observatories, ed. A.S. Szalay, Proc. SPIE, 4846, 103

Hambly, N.C., Williams, P.M., Read, M.A., 2006, "VDFS Science Archive Science Requirements Analysis Document", (SRAD) `http://www.roe.ac.uk/~nch/wfcam/VDF-WFA-VSA-002-I1/VDF-WFA-VSA-002-I1.html`

Hambly, N.C., Read, M.A., Mann, R.G., Sutorius, E.T.W., Bond, I., MacGillivray, H.T., Williams, P.M., Lawrence, A., 2004a, In: Proceedings of the 13th ADASS meeting, ASP Conf. Ser., 314, 137

Hambly, N.C., Mann, R.G., Bond, I.A., Sutorius, E., Read, M.A., Williams, P.M., Lawrence, A, Emerson, J.P., 2004b, In: Optimizing Scientific Return for Astronomy through Information Technologies, Proceedings of the SPIE, 5493, 423

Irwin, J. et al., 2007, MNRAS, in press (astro–ph/0612395)

Simard, L. et al., 2006, "UK VISTA Data Flow System Review Panel Report", available from `http://www.roe.ac.uk/~nch/wfcam`

Sutherland, W.J., 2005, "UK VISTA User Requirements", VDFS Document VDF–SPE–IOA–00009–001 (Issue 3.0, 5/07/05), available from `http://www.maths.qmul.ac.uk/~jpe/vdfs` (URD)

# Section 4(b) : WFCAM/VISTA Science Archive Operations

**Staff involved**

J. Bryant
E. Sutorius
M. Read

## 1 BACKGROUND

This case for support for WFCAM/VISTA science archive *operations* (as opposed to *development* – see Section 4a) continues from our last grant renewal proposal. At that review point the panel assessed this activity as "*the highest priority element of the WFAU programme*". Operations were funded at the level of 1.0 FTE from Q2 2005 to Q3 2007 rising to 2.0 FTE from Q4 2007 to Q1 2010. The panel stated that "*the need to exploit the scientific potential of the WFCAM* [and VISTA] *instrument*[s] *meant that both posts . . . were critical*". The panel anticipated that further reviews of support should be made in the light of operations experience, particularly after commencement of VISTA survey science operations. At the time of writing, VISTA data have yet to start flowing, owing to a delay in delivery of the primary mirror to Chile; moreover we do not anticipate any VISTA data flow until after review of the current proposal (early 2008). However, experience of WFCAM Science Archive (WSA) operations for two years and crystalising VISTA ESO Public Survey designs (see Section 4a and references therein) give us a firm idea of the staff and hardware resources necessary to support this work.

The present bid consists of a case for support for operations consisting of the day–to–day running of the science archives including transfer, ingest, curation and serving of science quality data, plus maintaining the hardware and software systems required for those tasks. In 2 we report on the first two years of successful WSA operations. In 3 we go on to describe in some detail the tasks necessary, and hardware resource requirements needed, to fulfil routine operations as a service to the UK and ESO communities. This justifies the present request for funding which is summarised in 5. Note that we make an explicit case for hardware resources here since the funds required are greater than those available on a simple *pro rata* based on the number of operations staff.

## 2 REPORT ON RECENT WORK: 2005 – 2007

### 2.1 Achievements: general operations

During early 2005 operations effort began to phase over from development, and shake–down of the WSA system took place. Starting with engineering commissioning data in Q1 2005, science verification data in Q2 2005, and then full–blown survey science data from Q3 2005 onwards, a total of over 35 TB of image and source catalogue FITS files has been transfered from the pipeline processing centre at Cambridge and ingested – this is a greater volume of data than that produced by the SuperCOSMOS machine during more than 10 years of plate scanning operations. Curation of all ingested data included production of quick–look JPEG images for browsing, quality error bit flagging[1] and production of multi–colour merged source lists as science–ready catalogue products. For the UKIDSS surveys, a rigorous quality control (QC) procedure was developed in collaboration with consortium survey scientists (e.g. Dye et al. 2006; see Figure 4). Following QC for survey data, more advanced database–driven curation is applied to the data. Deep stacks/mosaics and

[1]`http://surveys.roe.ac.uk/wsa/ppErrBits.html`

associated extracted source lists are produced (e.g. for the Deep eXtragalactic and Ultra–Deep Surveys). Broadband–narrow band PSF–matched difference images are also produced (K–H2) images for the Galactic Plane Survey. Following source merging a crossmatching procedure automatically creates neighbourhood pointers between IR sources and those in a host of locally stored but externally produced catalogue datasets (e.g. SDSS, 2MASS, radio and x–ray survey catalogues). These external datasets themselves have a considerable data volume: the WSA operations staff transfered databases or flat files for more then 5 TB of data (mainly the SDSS catalogue access databases from John Hopkins University) for local storage, crossmatching and user querying.

During the reporting period, in addition to commissioning and science verification data releases, the WSA operations staff have prepared three science–quality UKIDSS data releases[2] for the community totalling nearly 2 TB: the Early Data Release (EDR) in Q1 2006 (0.18 TB; see Dye et al. 2006); DR1 in Q3 2006 (0.43 TB; see Warren et al. 2007a); and DR2 in Q1 2007 (1.22 TB; see Warren et al. 2007b). DR3 is being prepared during the review process of this proposal and release is anticipated later in 2007. Operations staff have also made survey–like prepared database releases for 47 private PI programmes[3].

Administration of the IT infrastructure on which the WSA is built is a critical aspect of operations. During the reporting period, operations staff have maintained and expanded the large RAID arrays necessary for online storage; expanded and maintained the network infrastructure, including deployment of an experimental UKLight[4] network link with the processing centre; maintained and upgraded the archive software including patches and updates to third–party software; and finally have implemented and operated the system backup policy. Maintenance of the overall hardware systems has also included upgrading PC servers and LTO tape devices. Archive downtime is recorded[5] for the WSA – over the period Q3 2005 to Q2 2007 (731 days) archive downtime is less than 30 days, i.e. archive availability has been $> 96\%$.

## 2.2 Achievements: archive usage

At the time of writing, 626 individual users are registered for proprietary (18 month) UKIDSS database access in the WSA. These users are distributed over 72 distinct institutions, 42 of which are outwith the UK. Furthermore, 58 private PI programme registrations allow access by small PI–led teams to their proprietary (12 month) private datasets. The UKIRT Board–sanctioned access policy is that those proprietary datasets will become world–readable (i.e. accessible to unregistered users) after the proprietary period lapses; the first major world release will be the UKIDSS EDR in Q3 2007.

Figure 1 illustrates the level of archive usage over the reporting period (NB: these statistics exclude *all* Edinburgh access to avoid testing activities skewing the results). The WSA provides two distinct access modes: 'traditional' flat file access to standard data products (i.e. pipeline–processed image/catalogue FITS files and JPEG compressed images) for casual browsing and external QC; and flexible Structured Query Language (SQL) access to tabular datasets (mainly seamless, merged source lists) in prepared, static database products. Our baseline assumption during design was that SQL access would facilitate the easiest science exploitation; for the majority of our users, this appears to be the case, based on usage so far. The top panels in Figure 1 show flat file numbers and volumes (note the logarithmic ordinate scale). A peak in activity is seen in JPEG downloads just prior to the last two major UKIDSS releases (DR1 in July 2006 and DR2 in March 2007) due to the Consortium survey heads eyeballing all science images as part of the QC procedure in the

[2] `http://surveys.roe.ac.uk/wsa/releasehistory.html`
[3] `http://surveys.roe.ac.uk/wsa/nonSurveyStatus.html`
[4] `http://www.uklight.ac.uk/index.html`
[5] `http://surveys.roe.ac.uk/wsa/downtime.html`

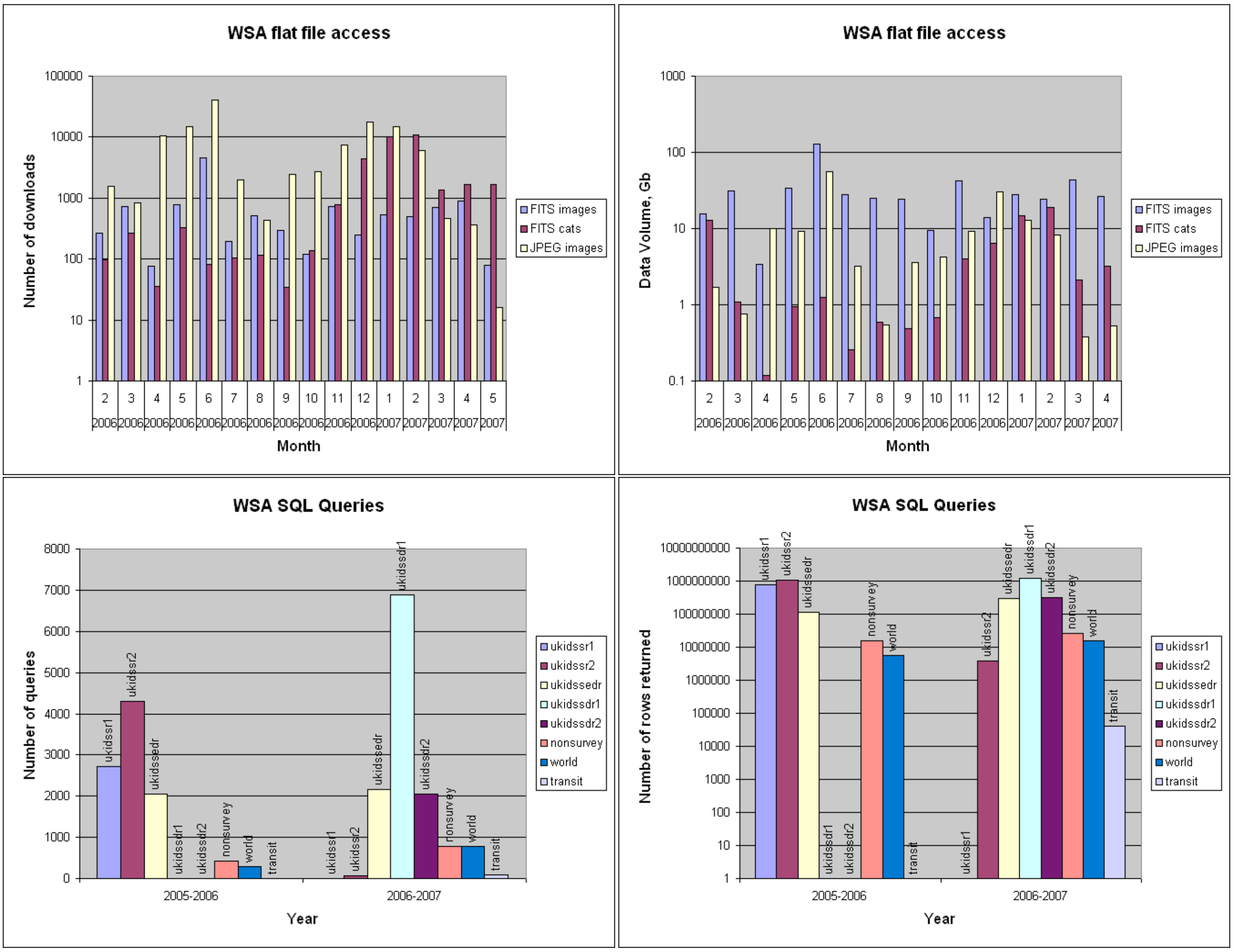



Figure 1: *WSA usage activity over the reporting period. Top panels: flat–file access (image/catalogue files and compressed JPEG images); bottom panels: SQL queries. Databases in the latter are: UKIDSSR1/2 – science verification data; UKIDSS EDR/DR1/DR2 – respectively the science–quality Early Data Release and Data Releases 1 & 2; World – a small 'taster' data subset from the EDR released for unregistered (world) access; and finally nonsurvey and transit are private PI programmes.*

run–up to the release. In terms of data volume, a steady download rate of > 10 GB/month is observed. The lower panels of Figure 1 show SQL activity. On the left, the number of queries from 'freeform' SQL, cross–ID and menu–driven queries input into the WSA web forms, broken down by survey database, indicates a significant increase from the first to the second year with $> 10^4$ queries in the past year. Most impressively, on the right those same queries but plotted in terms of rows returned shows that *billions* of rows of tabular data are being extracted by our users, with the most popular dataset over the past year being the UKIDSS DR1.

Finally, a 'helpdesk' system has been set up, and is operated, to support users of the archive. Over the reporting period, $\sim 100$ requests for support have been received. A response is normally made within a few hours of receipt during normal working hours. Any responses that are felt to be more widely useful are put on the static Q&A web pages[6] while general SQL query solutions

[6] `http://surveys.roe.ac.uk/wsa/qa.html`

are included in the SQL cookbook[7] maintained online. Sometimes users make requests for small changes and/or enhancements to archive user interface functionality. Maintaining the website and enhancing interface functionality have been further operational activities.

# 3 PROPOSED PROGRAMME OF WORK Q2 2008 – Q1 2013

## 3.1 Staff effort for science archive operations

Science archive operations entails managing data ingress/egress and curation, provision of database system administration, hardware maintenance, software maintenance, support for survey PI–led QC procedures, and provision of general user support in use of the more sophisticated aspects of the science archive user interface (e.g. SQL). An intimate knowledge of the underlying applications software is not essential but can help; astronomy knowledge and understanding of databases and procedures is essential, however, so we envisage these posts as being different to technician–type work. Below we describe the operations tasks (with the highest priority tasks listed first) in order to justify the staff resource requested later.

**General archive curation**
For the purposes of managing transfer, ingest and curation of pipeline processed data a set of curation 'use cases' has been designed[8] for the VDFS science archives. These consist of daily/weekly tasks, e.g. instigating data transfer, verification and ingestion; weekly/monthly tasks, e.g. source merging, image stacking and other database–driven activities like astrometric/photometric recalibration; and longer-timescale tasks, e.g. preparing release database products for publishing online and managing VISTA data products egress to the central ESO repository (see Section 4a and references therein). At their core, the curation applications are automated for bulk data handling operations. However higher level instigation of these (often complex) procedures requires a great deal of care and attention by skilled operators in order to maintain the high quality of the end product.

**Supporting survey PI–led quality control**
Public survey data quality control (QC) is an open–ended process in terms of evolving requirements and is intensively interactive. In the science archives, a deprecation attribute[9] is set using a manually instigated data modification script for a host of survey data quality issues. Example QC results for the UKIDSS EDR are shown in Dye et al. (2006). Our experience of working with the UKIDSS consortium survey scientists and survey heads has shown that a significant amount of effort is required to support survey QC. While PIs take the responsibility for leading QC procedural design, it is vital that the science archive operations team includes a specialised and experienced staff astronomer to work on local QC implementation in order to maintain survey data quality in released database products.

**Hardware expansion and maintenance**
Hardware expansion consists of specifying, procuring and installing new hardware (primarily disk storage) for the pixel/catalogue file store and catalogue database servers; upgrades are occasionally necessary (e.g. LTO tape upgrades, retiring/replacing the oldest PCs). Maintenance includes monitoring the health and performance of the large RAID5 arrays attached to the archive servers, e.g. checking for failing disks and sub–optimal arrays; swapping in replacements for failed redundant units (disks/PSUs); liaising with local system support, network administrative support upstream in the data flow system and hardware contract suppliers in maintaining the physical health of

[7] `http://surveys.roe.ac.uk/wsa/sqlcookbook.html`
[8] `http://www.roe.ac.uk/~nch/wfcam/VDF-WFA-VSA-007-I1/VDF-WFA-VSA-007-I1.html`
[9] `http://surveys.roe.ac.uk/wsa/gloss_d.html#multiframe_deprecated`

all the archive hardware and network infrastructure; and finally managing downtime, e.g. liaising with users to minimise the impact on their archive usage by scheduled maintenance of the archive system.

**User support**
Archive operations requires provision of user support. Access to WFAU archives is via the internet – this in turn implies a comprehensive and well–maintained set of web pages. Documentation and cookbooks become increasingly important as the user interfaces become increasingly sophisticated, so the archive operators are required to maintain a helpdesk system to provide a rapid response to requests for help and advice from users in addition to the production of documentation. Queries to the helpdesk system need to be fed back into online documented 'FAQs'. Hence, the operators are required to have a good working knowledge of relational databases, their primary applications interface (i.e. SQL) and their application in astronomy. Finally, the User Interface web applications themselves need to be maintained and enhanced in line with evolving user requirements.

**Database administration**
Database administration is a specialised IT support role. Experience of use of relational (SQL) database managements systems is vital for effective administration. Tasks include monitoring the health and performance of the server database instances; performing general housekeeping duties, e.g. making small schema modifications as requirements evolve; performing system and applications database backups as part of the overall disaster recovery policy; and finally monitoring usage and creating/dropping indexes as usages change. This final task is particularly important since we expose the full power and flexibility of the SQL interface to our users. Complex, long–running queries often can be sped up dramatically by judicious choice of indexed attributes.

**Support for PI programmes**
Private PI programmes ('non–survey' programmes in WFAU parlance) are an important component to the observing missions of WFCAM and VISTA. VDFS has always aspired to processing and archiving *all* data produced by the cameras (provided standard observing protocols are followed) not least because IR pipeline processing (e.g. sky estimation and calibration) benefits from use of as much data as is available from the complete night of observations from flexibly–scheduled programmes. As well as making basic FITS image/catalogue files available to PI teams, the science archive system allows for the curation of 'survey–like' databases for non–survey programmes. For WFCAM we already support UK PATT and service users (see Section 2.2 previously); the PIs of the new UKIRT Board sanctioned campaigns GABARIT (PI Kerrins), WFCAM Transit Survey (PI Pinfield), Coma cluster survey (PI Mouhcine) and HiZELS (PI Smail) have all been in touch with the WSA team concerning curation and/or publication of their campaign data.

**Support for complementary surveys and 'added value' datasets**
The VDFS science archives contain local copies of survey datasets created elsewhere for joint querying in the archive. For UKIDSS, the SDSS and 2MASS surveys have been particularly popular – see, for example, Kendall et al. (2007) and Lodieu et al. (2007). Survey datasets covering all wavelength ranges were either originally specified at the initial requirements capture phase or have been requested since (e.g. GLIMPSE, NVSS); moreover, we anticipate inclusion of the VST optical survey datasets ATLAS (PI Shanks) and KIDS (PI Kuijken) after commencement of those surveys. Finally, survey teams occasionally require publication of their own derived data through the science archive. An example of such an 'added value' dataset is the UKIDSS UDS deep image, which requires such specialised processing as to be beyond the scope of VDFS currently. For the UDS, a team led by PI Almaini at Nottingham has created the deep stack/mosaic at each release point so far and this is transfered, ingested, curated and served from the WSA as part of each UKIDSS data release. Much of the science from the SDSS, for example, has come from such value–added products. Supporting complementary surveys and user–supplied added value datasets is obviously important for facilitating full science exploitation of the main infrared surveys.

**Software maintenance**
Once VDFS science archive development is complete, the software package is handed over to the operations team for any residual software maintenance (e.g. bug fixes and minor enhancements). Equally important, moreover, is maintenance of third–party modules included as part of the overall system. Our experience is that updates and bug fixes that become available in third–party software (e.g. Python modules, the Python language itself, database connectivity middleware, software libraries like CFITSIO and AST) must be applied reasonably regularly ($\sim$ annually) to keep each component in step with the others to maintain a healthy working system. Even at the lowest (operating system) level, kernel patches must be applied (for example, to maintain a secure system) and those patches sometimes have knock–on effects that require attention when recompiling/reinstalling the higher level software. Hence, software maintenance is an occasional, but nonetheless essential and ongoing part of general archive operations.

**VO datacentre activities**
As stated in Section 4(a), WFAU is committed to implementing VO protocols and access infrastructure as soon as is reasonably possible given concurrent development work in e–science projects like AstroGrid and VOTECH. We plan to continue this work, which implies an operational overhead in, for example, deploying the software infrastructure developed by AstroGrid on hardware specially set up for this purpose. In particular, our request for hardware funding includes additional disk space on database servers for the personal database storage made available through AstroGrid's implementation of the planned VOSpace standard (discussed in greater detail in Section 4(d)). It is not clear how much disk space will be required for this, or whether this will place a significant additional load on our database servers, but, by the time of the next grant renewal, we shall have experience of running such a service for a couple of years, and will make a further request for hardware funding in the light of that. By continuing to deploy AstroGrid components, we will achieve full integration of the infrared science archives into the VO. Note that this activity benefits greatly from local VO expertise in the AstroGrid and VOTECH development teams that exist within WFAU.

## 3.2 Storage hardware expansion for science archive operations

Data rates, volumes and archive hardware design are all discussed extensively in our VDFS science archive hardware design document[10]. We originally estimated growth (e.g. previous case for hardware resources) in WFCAM pixel data at 1.7 TB per quarter assuming a yearly processed data volume of 20 TB which can be losslessly 'Rice' compressed (Sabbey 1999) by a factor $\sim 3$. Experience with two years of WFCAM operations has shown that this is an underestimate, chiefly because the requirement to store the ancestor frames (individual pointings and unstacked interleaved microstep frames) of final science stacks was neglected. A total of 16 TB of *compressed* image data and catalogues is stored for the first year of WFCAM operations; the figure for the second year is similar. Hence, for WFCAM, we require 4 TB of flat file storage space per quarter. For the database servers, our storage requirement is for an incrementally growing ingest database, while online public catalogue storage grows as an arithmetic progression with time on an approximately six month timescale. This is because we are required to keep online all previous catalogue database releases (see, for example, the VDFS Final Design Review panel report, item Simard 13 – Section 4a and references therein) and each release contains data (sometimes reprocessed and recalibrated) from all previous releases as well as new additions. Our experience is that database server storage is $\sim 5\%$ of the flat file requirement, i.e. 0.2 TB per quarter for WFCAM. For example, the UKIDSS EDR ($\sim 1$ quarter) was $\sim 0.2$ TB; DR1 (2.5 quarters) was $\sim 0.5$ TB; DR2 (6 quarters) was $\sim 1.2$ TB. For VISTA, it is anticipated that the storage requirements will be a factor 2 larger,

[10] http://www.roe.ac.uk/~nch/wfcam/VDF-WFA-VSA-006-I1/VDF-WFA-VSA-006-I1.html

| Year | Q | Moore factor | WFCAM Files | WFCAM DBs | VISTA Files | VISTA DBs | Total per quarter | LTO–III tapes |
|---|---|---|---|---|---|---|---|---|
| 2008 | 2 | 1.000 | 2.359 | 0.737 | 4.718 | 1.474 | 9.289 | 0.106 |
| 2008 | 3 | 0.890 | 2.100 | 6.368 | 4.199 | 5.537 | 18.204 | 0.569 |
| 2008 | 4 | 0.792 | 1.869 | 0.584 | 3.737 | 1.168 | 7.358 | 0.084 |
| 2009 | 1 | 0.705 | 1.663 | 7.271 | 3.326 | 7.343 | 19.604 | 0.699 |
| 2009 | 2 | 0.627 | 1.480 | 0.463 | 2.960 | 0.925 | 5.828 | 0.066 |
| 2009 | 3 | 0.558 | 1.317 | 7.987 | 2.635 | 8.773 | 20.712 | 0.801 |
| 2009 | 4 | 0.497 | 1.172 | 0.366 | 2.345 | 0.733 | 4.616 | 0.053 |
| 2010 | 1 | 0.442 | 1.043 | 8.553 | 2.087 | 9.907 | 21.590 | 0.883 |
| 2010 | 2 | 0.394 | 0.929 | 0.290 | 1.857 | 0.580 | 3.657 | 0.042 |
| 2010 | 3 | 0.350 | 0.827 | 9.002 | 1.653 | 10.804 | 22.286 | 0.947 |
| 2010 | 4 | 0.312 | 0.736 | 0.230 | 1.471 | 0.460 | 2.896 | 0.033 |
| 2011 | 1 | 0.278 | 0.655 | 9.358 | 1.309 | 11.515 | 22.837 | 0.998 |
| 2011 | 2 | 0.247 | 0.583 | 0.182 | 1.165 | 0.364 | 2.294 | 0.026 |
| 2011 | 3 | 0.220 | 0.519 | 9.639 | 1.037 | 12.078 | 23.273 | 1.038 |
| 2011 | 4 | 0.196 | 0.462 | 0.144 | 0.923 | 0.288 | 1.817 | 0.021 |
| 2012 | 1 | 0.174 | 0.411 | 9.862 | 0.822 | 12.524 | 23.619 | 1.070 |
| 2012 | 2 | 0.155 | 0.366 | 0.114 | 0.731 | 0.228 | 1.439 | 0.016 |
| 2012 | 3 | 0.138 | 0.325 | 10.039 | 0.651 | 12.878 | 23.893 | 1.096 |
| 2012 | 4 | 0.123 | 0.000 | 0.000 | 0.579 | 0.181 | 0.760 | 0.009 |
| 2013 | 1 | 0.109 | 0.000 | 0.000 | 0.515 | 13.158 | 13.673 | 0.629 |
| | | | | | Totals: | | 249.646 | 9.186 |

Table 1: *Projected hardware costs (£K including VAT) per quarter for curating and serving the large infrared public surveys from WFCAM and VISTA. Assumptions are: todays prices (excluding VAT) at £0.8K/TB for mass file storage (SATA RAID), £2.5K/TB for database server storage (SCSI RAID) and LTO–III tapes at £75/TB; 11% reduction per quarter in prices due to Moore's law (50% reduction for the same capacity over 18 months); arithmetically accumulating public survey database releases every 6 months in addition to constant accumulation in the load server ingest database. Note that the WFCAM surveys (UKIDSS) are expected to run for 7 years and started in Q2 2005, hence we assume a final WFCAM survey release in Q3 2012. The totals (including VAT) are £250K for online storage and £9.2K for consumables (i.e. tapes).*

i.e. flat file storage at 8 TB per quarter and database storage of 0.4 TB per quarter starting in Q2 2008.

For the purposes of projecting storage costs into the future we apply a Moore's law adjustment. This is also based on our own experience: in 2000, 50 GB SCSI disks cost around £1K whereas in mid–2003, 180 GB SCSI disks cost around the same price – this corresponds to a classic Moore–like scenario of a factor $\sim 2$ gain in 18 months. This in turn equates to a 11% reduction in storage costs per quarter *for the same capacity.* Table 1 shows the cost breakdown of the required science archive storage, scaling by Moore's law from projected price at the start of the grant period.

Backup media, currently LTO Ultrium–III tapes[11], 0.4 TB per tape, £30 per tape (excluding VAT), are required to secure online database data for quick recovery in the event of major disk array failure; we do not require to backup FITS flat file data since a copy is held online at CASU. Hence, we request resources to cover tape consumables as indicated in Table 1 where we have once again scaled by Moore's law (LTO has a clear upgrade path over the next few years).

[11] `http://www.ultriumlto.com/`

### 3.3 Maintenance of archive hardware

Operational requirements for the WFCAM and VISTA archives come from their respective user requirements documents (Adamson 2002; Sutherland 2005) and the associated analysis document[12] For example, Requirement 16 in Section 5 of the VISTA User Requirements specifies archive availability figures that in turn imply the need for a quick response in the event of hardware failure. Our archive design includes a large degree of redundancy for robust operation – for example, all file systems are RAID5 arrays which are fault–tolerant against single disk failure in any one RAID set, and the disk housing units themselves have redundant power supplies. There is no need to have extended maintenance for such redundant systems since all our disk controllers support 'hot swapping' of failed disks, and redundant power supplies automatically switch on in the event of a unit failure. All that is required for these systems is a small surplus of spare disks and other modular components; then we anticipate that the archive operators will be able to cope with low–level system maintenance without impacting the archive up–time. Continued on–site maintenance contracts are required only for critical single modules within the archive system (e.g. web server, tape backup unit) after the normal 3 year contract included in the original purchase price has expired.

Hence, for the purposes of archive hardware maintenance we request a 10% overhead on the hardware cost for spares (i.e. one hot–swappable spare disk for every 16 purchased, and spare power supply modules etc.); we request also resources for maintenance for the mission–critical servers and other components to the tune of £8K per year, excluding VAT. Maintaining our own redundant system with a small pool of spares represents a considerable saving over simply requesting resources to cover maintenance contracts for all the archive hardware purchases (for which we have been quoted in the past £28K per year), and we emphasise that this is the minimum maintenance resource necessary to service the user–required archive availability quoted in, for example, Sutherland (2005).

## 4 KEY DELIVERABLES and MILESTONES

The key deliverables (D) and milestones (M) of the WFCAM/VISTA Science Archive Operations theme up to the end of the grant period are as follows:

| Date | Type | Details |
|---|---|---|
| Q1 2008: | M | VISTA Science Archive (VSA) operations begin |
| Q2 2008: | M | VISTA survey QC starts |
| Q2 2008: | D | UKIDSS DR4 |
| Q3 2008: | D | First VSA survey data releases |
| Q3 2012: | D | Final UKIDSS Data Release |
| Q3 2012: | M | WFCAM science archive operations complete |

In addition, data releases of the UKIDSS and VISTA ESO Public Surveys and private PI 'non–survey' datasets will take place throughout the grant period, at dates TBC.

[12] `http://www.roe.ac.uk/ nch/wfcam/VDF-WFA-VSA-002-I1/VDF-WFA-VSA-002-I1.html`

# 5 RESOURCES REQUESTED

We request continuing staff resources (2.0 FTE) for WSA/VSA operations over the full period. Additionally we request 0.5 FTE from Q2 2008, rising to 1.0 FTE in Q4 2009 for M. Read to phase over from VDFS user interface development (Section 4a) to become Survey Quality Assurance Scientist with overall responsibility for implementing survey QC and to supplement general operations. Read is a highly experienced survey astronomer whose presence will be essential to provide QC support and also to provide continuing user support for the science archives.

The requested travel funds will support operations staff presenting reports to UKIRT Board meetings, and other external oversight bodies; interacting with survey consortia and liaising with ESO over VISTA Public Survey data and attending conferences such as NAM, JENAM and ADASS[13] for the purposes of providing 'roadshow' demonstration of archive functionality and for face–to–face interaction with the user community.

In summary:

(a) *Staff.* 2.5 FTE from Q2 2008 rising to 3.0 FTE from Q4 2009, totalling 14.25 d.s.y.
(b) *Travel and subsistence.* £2K/FTE/year = £28.5K
(c) *Consumables:* £9.2K for LTO tapes, plus £1K/FTE/year for PC upgrades and other consumables = £23.45K
(d) *Maintenance:* £8K+VAT per year for the whole period (£47K in total)
(e) *Equipment:* £250K + 10% (spares) = £275K for science archive hardware

We reiterate that astronomy knowledge and understanding of databases and procedures will be required on the part of the operators so these are not technician–type posts.

# 6 ADDITIONAL REFERENCES

Adamson, A., 2002, "WFCAM Science Archive Science Requirements", `http://www.jach.hawaii.edu/~adamson/wfarcrq.html`

Dye, S. et al., 2006, MNRAS, 372, 1227

Kendall, T.R. et al., 2007, A&A, 466, 1059

Lodieu, N. et al., 2007, MNRAS, 374, 372

Sabbey, C.N., 1999, In: Proceedings of the 8th ADASS meeting, ASP Conf. Ser., 172, 129

Sutherland, W.J., 2005, "UK VISTA User Requirements", VDFS Document VDF–SPE–IOA–00009–001 (Issue 3.0, 5/07/05), available from `http://www.maths.qmul.ac.uk/~jpe/vdfs`

Warren, S.J. et al., 2007a, MNRAS, 375, 213

Warren, S.J. et al., 2007b, astro–ph/0703037

[13] `http://www.adass.org/`

# Section 4(c) : Science Archiving for Synoptic Sky Surveys

**Staff involved**

R. Collins (from Q3 2011)
N. Cross (from Q4 2009)
N. Hambly (from Q4 2009)
New Synoptic Surveys Scientist (from Q4 2009)
New Synoptic Surveys Developer (from Q2 2010)

## 1 BACKGROUND

### 1.1 Introduction

This case describes the early stages of WFAU's proposed work on the three new projects which, together with continued operation of the WFCAM and VISTA Science Archives, are intended to form the core of WFAU's programme after our next rolling grant renewal. These projects are (i) development of a public Science Archive for data from the Pan-STARRS 1 (PS1[1]) project; (ii) development of a Science Archive for Gaia[2]; and (iii) design and development of data management and data analysis capabilities for LSST[3], the Large Synoptic Survey Telescope. These are three of the most powerful sky survey projects planned for the next 10–15 years and what unites them is their repeated observations of each field – $\sim$10 times for PS1, $\sim$100 times for Gaia, and $>$1000 times for LSST – across a wide area (at least half the sky in each case). These surveys represent optical astronomy's first large–scale forays into the time domain, opening up a wide range of scientific topics, as well as presenting a number of significant new challenges for those curating, and publishing, the data from them. By following a phased approach to the three projects we can exploit the synergies between them in the same manner as we did with WFCAM and VISTA, as well as integrating the new optical data with these existing near–IR datasets, and thereby securing for the UK community the best chance to exploit these major new surveys to the full.

### 1.2 PS1

PS1 is a survey system built centred on a dedicated 1.8m telescope on Haleakala, Maui, with a 7 sq. deg. FOV and the world's largest CCD camera ($1.4 \times 10^{12}$ pixels). It is a prototype for a larger system – PS4, with four such telescopes, is under development, and a further extension under discussion – but it is a very powerful survey facility in its own right. Construction of PS1 has been funded primarily by the US military, and it will be operated by the PS1 Science Consortium (PS1SC[4]). The PS1SC will conduct a 3.5 year survey programme, starting in late 2007. The main survey, taking $\sim$55% of the time, will cover $3\pi$ steradians to $r = 24$ (plus corresponding depths is $g$,$i$,$z$ and $y$), while a Medium Deep Survey will cover 84 sq. deg. in the same filters, to a depth approximately three magnitudes fainter, and a series of smaller surveys are targetted at specific areas of sky, such as M31 and the fields of a Stellar Transit Survey. The PS1 survey programme

[1] http://pan-starrs.ifa.hawaii.edu/public/
[2] http://www.esa.int/science/gaia
[3] http://www.lsst.org
[4] The PS1SC comprises the following institutions: the University of Hawaii, with the National Central University of Taiwan; Johns Hopkins University; the Max Planck Society (MPA Garching and MPIA Heidelberg); the Harvard/Smithsonian CfA with Las Cumbres Observatory; and a UK consortium composed of Durham, Edinburgh and Queen's Belfast.

centres on a set of twelve Key Science Projects, covering a wide range of topics, namely: Populations of objects in the Inner Solar System; Populations of objects in the Outer Solar System; Low-Mass Stars, Brown Dwarfs, and Young Stellar Objects; Search for Exo-Planets by dedicated Stellar Transit Surveys; Structure of the Milky Way and the Local Group; A Dedicated Deep Survey of M31; Massive Stars and Supernovae Progenitors; Cosmology Investigations with Variables and Explosive Transients; Galaxy Properties; Active Galactic Nuclei and High Redshift Quasars; Cosmological Lensing; and Large Scale Structure. While PS1 is not currently on the STFC Roadmap, it addresses a number of the key Science Questions that define it, notably *How do galaxies, stars and planets form and evolve?*, and, perhaps most powerfully, through a combination of weak gravitational lensing, baryon acoustic oscillations and Type Ia supernovae, *What is the Universe made of and how does it evolve?* Moreover, PS1's deep optical data can significantly enhance a number of projects already on the Roadmap: for example, as the Planck Core Programme takes shape, it is becoming clear that the PS1 $3\pi$ survey will be crucial for the detailed study of the Integrated Sachs-Wolfe effect and attempts to discover the *"missing baryons"* in the local Universe through cross-correlating Planck's Sunyaev-Zel'dovich Effect maps with PS1's 3D map of the local galaxy distribution to reveal the predicted *"Warm/Hot Intergalactic Medium"*.

## 1.3 Gaia

Gaia is an ESA Cornerstone mission, due for launch in Q4 2011, and with a lifetime of five to six years. It will conduct whole–sky photometric and spectroscopic surveys, visiting each field an average of $\sim 70$ times, and generating a photometric dataset comprising $\sim 10^{12}$ observations. Gaia's principal goal is the generation of a high precision 3D map of $> 10^8$ stars in the Milky Way, enabling a detailed investigation of the formation and evolution of the Galaxy, as well as fundamental studies of stellar physics. It is also expected to detect thousands of new extrasolar planets, determine accurate orbits for hundreds of thousands of minor planets in the Solar System, and test General Relativity by following the bending of star light by the Sun and major planets across the whole celestial sphere. Gaia's benefits to extragalactic astronomy will include improving the determination of the cosmological distance scale, through geometric distances to Cepheids and RR Lyraes, and fixing the cosmological reference frome, through precise astrometry of quasars.

## 1.4 LSST

While PS1 has yet to start survey operations and Gaia is still four years from launch, the Large Synoptic Survey Telescope (LSST) represents the generation of sky survey facilities to follow them. Current planning sees LSST commissioned in 2015, but the scale of the technological challenges it will pose – most notably a data flow of $\sim$2 TB *per hour*, which must be mined in real time to meet many of LSST's key science goals – means that detailed design work is already underway in the US, where LSST has received high–priority ranking from recent national reviews in astronomy, solar system science and particle physics. The power of LSST is most vividly demonstrated by Figure 1, which compares the grasp of a range of current and planned sky survey facilities. The grasp of a telescope (which is defined as the product of area of its primary mirror and the solid angle of its field of view) is its fundamental figure of merit as a survey system, as it dictates how quickly it can cover the sky to a given depth. LSST achieves more than an order of magnitude increase in grasp over current planned facilities through a design which featues an annular primary mirror with an outer diameter of 8.4m and a 9.6 sq. deg. FOV, but which also retains the image quality needed for weak lensing studies. Current LSST plans envisage a survey of 20,000 sq. deg. conducted over a ten year period, reaching final depths of $u = 26.2$, $g = 27.4$, $r = 27.6$, $i = 26.9$, $z = 26.1$ and $Y = 24.8$, and visiting each patch of sky $\sim$2000 times, with a range of cadences, as required to

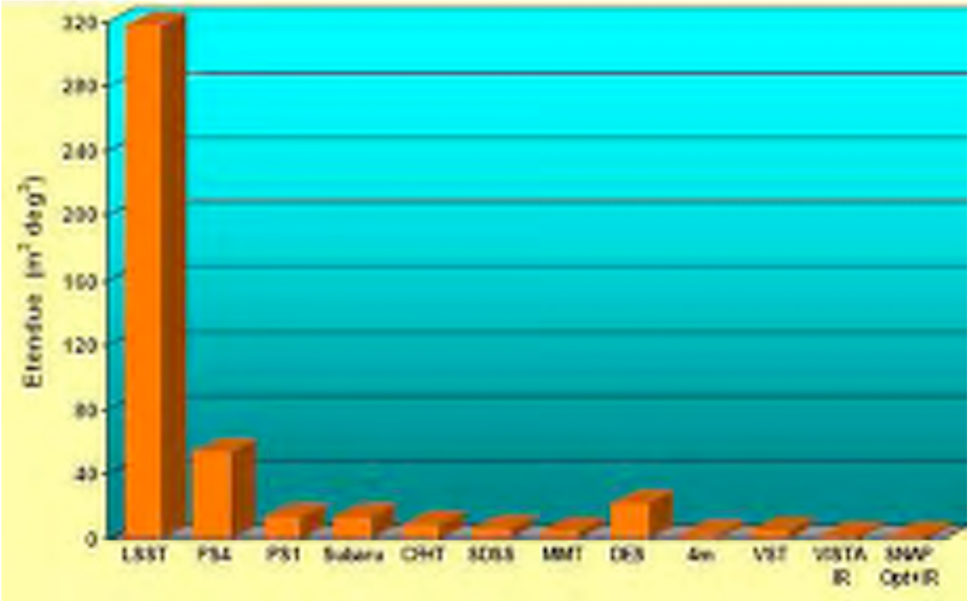



Figure 1: *The grasp, or étendue, of a range of current and planned sky survey facilities. (LSST Corp.)*

detect and study a range of time–varying phenomena, from meeting the US Congressional Mandate for identifying Potentially Hazardous Asteroids to finding $> 10^6$ Type Ia SNe per year. The focus of LSST science is, however, complementary techniques – weak lensing, baryon acoustic oscillations, galaxy cluster counts and SNe Ia – for probing the nature of dark matter and dark energy.

# 2 PROPOSED PROGRAMME OF WORK Q4 2009 – Q1 2013

## 2.1 Introduction

The strength of the science case for each of these three surveys is unquestionable, and the UK community will require full and ready access to all three datasets. We argue below that WFAU is uniquely qualified to provide that access, and, while the bulk of the work on all three projects will take place after our next grant renewal, significant work must start before then. The detailed resource requirements for the post-2010 will only become clear in the next couple of years, for the reasons we describe below, so we focus here on the justification of the resources needed before then.

## 2.2 PS1

The PS1 Design Reference Mission[5] notes that the *PS1 Science Consortium has agreed to allow the release of all PS1 Data Products, including image data, signal to noise maps, catalog data, derived data products and metadata to the astronomical community and public at large* but that there *is no commitment by the PS1SC to fund any means of serving PS1 Data Products.* The PS1SC's policy on consortium proprietary rights and public data release is due to be debated at a meeting of its Science Council in August 2007, but the intention of the Pan-STARRS PI[6] is that regular releases of data products will be made (possibly by individual Key Project teams) after short proprietary periods, once the survey enters routine operations, and data quality can be assured.

[5]PS1 DRM: http://www.ifa.hawaii.edu/users/chambers/ps1sc/documents/DesignReferenceMission_01.pdf
[6]N. Kaiser, priv. comm., May 2007

We propose to develop a Science Archive to publish PS1 data products, as produced by the standard PS1 Image Processing Pipeline and, where possible, value–added products from Key Project teams. These would be integrated into the WFCAM Science Archive in the manner of existing auxiliary datasets, like the Sloan Digital Sky Survey (SDSS): i.e. we would generate cross-neighbours tables enabling PS1 catalogues to be jointly queried with our existing sky survey datasets, thereby facilitating a range of analyses, such as the selection of low mass stellar and substellar samples on the basis of UKIDSS colour selection plus proper motion and/or variability information from PS1. We anticipate making a series of releases of coherent collections of PS1 data products — all the data, not just in the UKIDSS overlap regions — as they become public, with consortium members having access to them while they remain proprietary, thereby validating the quality of the data and the functionality of the archive. This work is likely to be conducted in collaboration with the team at Johns Hopkins University responsible for the SDSS archive, with whom we have a very good existing working relationship. As well as providing the UK community with ready access to PS1 data in a manner not available to them otherwise, this work would also form part of our R&D programme leading to LSST (see below), by providing us with practical experience of creating and curating a large, multi-epoch optical dataset, to be analysed for its temporal, as well as spatial and photometric, properties.

The deliverables and timetable are summarised below.

### 2.3 Gaia Catalogue Access

Gaia's data products fall into three categories. Firstly, alerts will be made during operations, to facilitate the timely follow-up of transient events and unexpected indications of variability in known sources. Secondly, the final Gaia catalogue, due for release in 2020, will contain astrometric and spectro-photometric data for $\sim 10^9$ sources, including variability and classification information. These first two categories of data product are relatively well defined in the Gaia Science Management Plan and the Announcement of Opportunity for the Gaia Data Processing and Analysis Consortium (DPAC). What is far less well defined so far is the third category of intermediate data products, to be released – on a schedule to be decided by the Gaia Science Team – before the final catalogue is available in 2020. To a first approximation, Gaia's first few all-sky surveys will define a source catalogue and subsequent surveys will provide further epochs of data relating to the sources in that catalogue, as well as improving the accuracy of the global astrometric solution and of the calibration of Gaia data. It will be possible, therefore, to release all-sky catalogues during the course of Gaia operations, and these are likely to be of great interest to the community.

As part of the PPARC/STFC-funded Gaia Data Flow System (GDFS) project, WFAU is developing the data management system needed to support the Gaia photometric analysis pipeline, within Coordination Unit 5 (CU5) of the Gaia Data Processing and Analysis Consortium (DPAC). Within DPAC, public access to (both intermediate and final) Gaia catalogues will be the responsibility of CU9, which has yet to be activated. The DPAC *Proposal for the Gaia Data Processing*, accepted by the ESA Science Working Group in April 2007, outlines the Gaia catalogue access to be provided by CU9 in terms very reminiscent of WFAU's core Science Archive concept: i.e. that online access to the Gaia catalogues will be provided by a web–based portal, complemented by additional server-side data mining facilities, all of which are implemented in a VO–compliant manner. In the light of that, it is unsurprising that WFAU's initial expression of interest in working on CU9 was welcomed by the Gaia Project Scientist[7], who confirmed that CU9 will be the subject of separate Announcement of Opportunity, to be issued by ESA no earlier than 2008.

We propose to contribute to the development of a Science Archive for Gaia's photometric data

[7] T. Prusti, priv. comm, March 2007

products, both intermediate and final: *N.B.* this is separate from, but builds upon, our work on CU5, funded through the GDFS. This concept was endorsed by the Panel assessing our last Rolling Grant bid, but no funding was awarded, as it was felt that this should be included in the GDFS proposal, rather than funded separately. The decision by ESA to activate CU9 later than the other CUs precluded inclusion of our resource request in the GDFS bid, so we request the initial funding for this work here. Once the CU9 A0 has been released, we will be in a better position to assess the resource requirements for this project as a whole, and we expect to make a revised request as part of our next grant renewal.

## 2.4 LSST

The data rate from LSST – around ∼2 TB *per hour*, and, so, ∼7 PB per year – and its focus on time-dependent phenomena, necessitates a data management and analysis paradigm reminiscent of a particle physics experiment. From the summit in Cerro Pachon, in Chile, data will be transfered to a Base Facility in La Serena, where nightly data products (including transient alerts) will be generated on a ∼25 Tflop–class supercomputer. Data will then pass to Archive Centres, which will perform more advanced processing and may be the location for specialised data analysis tasks, and then on to several Data Centres, which will store data products and serve them to the wider community through the Virtual Observatory (VO). The main Archive Centre will be at the National Center for Supercomputing Applications (NCSA) in Illinois, but there could be one or more others. LSST development is currently being funded by a combination of National Science Foundation (NSF) and Department of Energy (DOE) awards, plus (currently token) contributions from the institutions (mainly US universities, and national labs, plus Google Corporation) which make up the LSST Corporation. This mixed funding model will continue, and the LSST management is very interested in securing a European contribution to the project. An initial meeting to discuss possible European involvement was held in Potsdam in March 2007, at which it was clear that existing commitments to ALMA and ELT development mean that ESO cannot fund a European contribution, so this is likely to have come from a consortium of institutions/agencies: participants from Edinburgh and Durham attended the Potsdam meeting, and representatives of Oxford, Portsmouth and QUB have also indicated interest in joining a UK component of a European LSST consortium.

We propose to take a leading role in the development of a European contribution to LSST and to secure the location of an Archive Centre in Edinburgh. This plan rests not only on WFAU's expertise in sky survey data curation, but also on the University's existing strengths in high-performance computing. The University's Advanced Computing Facility (ACF), on the Edinburgh Technopole site, just outside the city, will host HECToR[8], the UK's new national supercomputing facility, and has been designed to provide flexible accommodation for large-scale computational projects, such as an LSST Archive Centre. The key to successful UK participation in the LSST is to build on the experience in engineering public access systems for earlier projects that are similar in experimental nature (albeit considerably smaller in scope). Much time initially will be spent on discussion meetings and planning with international partners. Staff experienced in such interactions in PS1 and Gaia will be invaluable, so we propose first to phase over our science archive R&D scientists, to be followed later by development effort once the initial planning work is complete and the requirements have been captured and analysed). LSST work will follow the same systems–engineered and formally documented/reviewed approach that has been successful in the past: requirements capture and analysis, data flow analysis and data modelling followed by software architectural design leading to a coding/implementation phase. However, the main activity for LSST work prior to the next rolling grant renewal is development of a plan of work for inclusion in that case for resources as part of a large, international collaborative effort.

[8]http:/www.hector.ac.uk

# 3 KEY DELIVERABLES and MILESTONES

The key deliverables (D) and milestones (M) during the relevant period are as follows:

| Date | Type | Details |
|---|---|---|
| Q4 2007 (end) | M | PS1 proprietary surveys up and running |
| Q4 2008 | M | ESA AO for GAIA Catalogue Access (at the earliest); |
| | D | response to the ESA AO 6 months later |
| Q3 2009 (end) | M | VDFS Science Archive development ends |
| Q4 2009 | M | Synoptic surveys scientist in post to complement effort from VDFS; public PS1 science archive requirements capture and analysis begins; planning for GAIA and LSST work commences. |
| Q4 2009 (end) | D | Requirements and analysis documents for public PS1 SA |
| Q1 2010 (end) | D | Data flow and data modelling documents for public PS1 SA |
| Q2 2010 (end) | D | Requirements and analysis documents for public GAIA catalogue access |
| Q2 2010 | M | Synoptic surveys developer in post for coding public PS1 SA. |
| Q2 2010 (end) | M | Initial planning for LSST work complete |
| Q2 2010 (end) | D | Next RG renewal case (incorporating LSST planning) |
| Q4 2010 | D | PS1 science archive goes live |
| Q2 2011 (end) | M | PS1 survey complete |
| Q4 2011 | M | GAIA launch |
| Q1 2015 | M | LSST first light (at the earliest) |

# 4 RESOURCES REQUESTED

We request staff resources totally 8.7 d.s.y., as summarised in part 3 of the overall case, starting in Q4 2009. At that point, 0.2 FTE of each of Hambly and Cross will transfer to this theme, to kick-start its work, in conjunction with a new Synoptic Survey Scientist. Six months later they will be joined by a new developer. The two new appointees will work full-time (modulo the 0.3 FTE research allocation for the scientist) on this project throughout the lifetime of the grant. Further staff resources will be needed once work on the Gaia catalogue access starts in earnest, and as preparations for LSST ramp up, so Hambly and Cross will increase the fraction of their effort devoted to this work to 0.7 FTE each in Q3 2011, and they will be joined by Collins working full-time on development for Gaia catalogue access (building on the Gaia CU5 work he will have been doing until then).

We request travel and subsistence at the rate of £2K/FTE/year for the project staff, plus a £8K for the Investigators: they (principally Mann and Lawrence) will undertake significant amounts of travel in support of this theme, primarily during the establishment of a European LSST consortium and during the LSST planning phase. This totals £25.4K

We request funding for consumables at a rate of £1K/FTE/year for the project staff, plus 2 × £2K to buy workstations for the two new staff. This totals £12.7K.

In Summary:

(a) *Staff.* 8.7 d.s.y.
(b) *Travel and subsistence.* £25.4K
(c) *Consumables:* £12.7K
(d) *Maintenance:* none requested
(e) *Equipment:* none requested

# Section 4(d) : Data Analysis Services

**Staff involved** L. Rimoldini

## 1 BACKGROUND

WFAU's Science Archives allow users to run arbitrary SQL queries against data stored in a relational database management system (RDBMS), and astronomers are becoming increasingly adept at exploiting the power of SQL. The advantages of building sky survey Science Archives on RDBMS greatly outweigh the disadvantages, but SQL does have its limits: it is a set-based language, so it can only be used to select sets of database entries satisfying particular selection criteria. Sky surveys have primary science drivers which centre on either searching for very rare objects, such as high-redshift quasars or low mass (sub)stellar objects, or generating statistical summaries of the populations of sources, such as measures of their clustering or correlations between particular physical properties. These kinds of operation often cannot be expressed efficiently in SQL, so users cannot perform these analyses using the existing Science Archive SQL query interfaces provided by WFAU. Currently, they must extract the data needed for their analysis from the database using an SQL query, transfer them over the network to their own institution and analyse them using code they write for the purpose. This is wasteful of human resources – many researchers writing code to perform the same basic statistical operations – and, as sky survey datasets grow in size, it becomes increasingly difficult for users to download and store all the data that they require for their analyses. Evidence for this comes from the Sloan Digital Sky Survey (SDSS), most of the large-scale statistical analyses which have been undertaken by researchers at SDSS consortium institutions, with access to a local copy of the SDSS dataset, and thereby exploiting privileged access to the data, which is not acceptable for the public datasets curated by WFAU. A better approach would be to have the basic statistical code running at the data centre directly against the database. Security concerns mean that it is not currently possible to allow archive users to upload and run unchecked analysis code to be run in that manner[1], and, clearly, the adminstrative overhead precludes the checking of code provided by users. It follows, therefore, that the code must be implemented by the data centre, as part of the Science Archive.

In the last WFAU rolling grant renewal we successfully sought funding for a new position – a data mining applications scientist – to start the development of such data analysis services as part of our Science Archives. The case for retaining that post has been strengthened by a number of recent events. Firstly, the science cases for several of the VISTA Public Surveys (e.g. VIDEO, VIKING and VHS) all centre on exactly the kind of analyses – e.g. finding quasars at $z > 5$ and galaxy clusters at $z > 1$ – requiring the provision of data analysis services as part of Science Archives. Secondly, recent R&D work undertaken as part of the Data Exploration Design Study (DS6) within the VOTECH project[2] shows how such services can readily be made callable from within the existing VO infrastructure, and has also helped identify how data centres can make such services more efficient. Thirdly, we are seeing evidence of some WSA users downloading in bulk UKIDSS source catalogues in FITS format, which suggests there may already be some duplication of software development effort in the community, which may be prevented by provision of standard analysis services as part of the WSA. Fourthly, the plans for analysis of data from LSST (described in Section 4(c)) suggest that the close coupling of data management and data analysis that we are

[1] WFAU is collaborating with researchers from the University of Edinburgh's Laboratory for the Foundations of Computer Science on ways to do that securely – see http://groups.lfcs.inf.ed.ac.uk/request – but that research is some way from producing readily deployable software.

[2] See http://wiki.eurovotech.org/twiki/bin/view/VOTech/DS6StudyReport

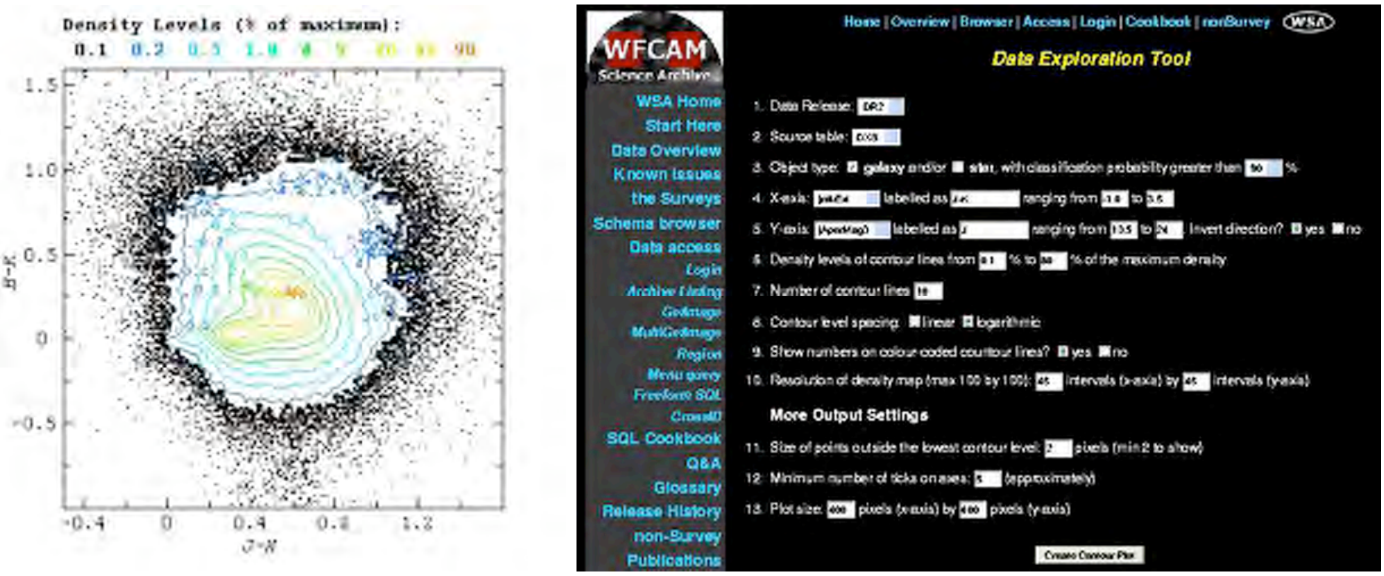



Figure 1: *(Left) A plot from the WSA website of the distribution of UKIDSS LAS DR2 stars in H-K versus J-H colour-colour space. The data are plotted as contours down to a density level of 0.1% of the peak value, and those (highly unusual) objects lying outside that lowest contour are plotted as points. (Right) A screenshot of the dynamic contour plotting tool, which enables users to specify the data and plotting parameters.*

prototyping here will be the solution adopted for future surveys.

## 2 REPORT ON RECENT WORK: 2006-2007

The decision not to fund the research allocation sought for this post made its recruitment much harder than anticipated, requiring a re-advertisement. We did eventually appoint a strong candidate, Lorenzo Rimoldini, but his arrival in November 2006 means that only six months of work have been undertaken on this project to date. Rimoldini had no prior experience of either of the two languages – Java and Python – currently used to access WFAU's databases, so his initial task was designed to train him in Java, as well as to deliver a service to WSA users.

Many sky survey analyses involve selections in colour-colour or colour-magnitude space, so Rimoldini's first task has been to develop code for generating contour plots of these spaces for each of the UKIDSS surveys: firstly, static plots and, secondly, a tool enabling users to customize the selection of the data subsets for which they wish to generate a contour plot (both shown in Figure 1).

## 3 PROPOSED PROGRAMME OF WORK Q2 2008 – Q1 2013

In the last WFAU grant proposal, the proposed programme of work for the holder of this new post for the period up to Q1 2008 comprised three topics: (i) implementing prototype data structures to aid UKIDSS science analyses; (ii) experimenting with hardware architectures, to see which is best suited to the provision of data analysis services; and (iii) investigating possibilities for the secure running of user-supplied analysis code. The late starting date for this theme has slowed progress towards these goals. As noted above, item (iii) is being investigated with collaborators from the University's School of Informatics, but is clearly some way from delivering generally-deployable code, but significant progress on (i) and (ii) should be made by the end of Q1 2008, as originally

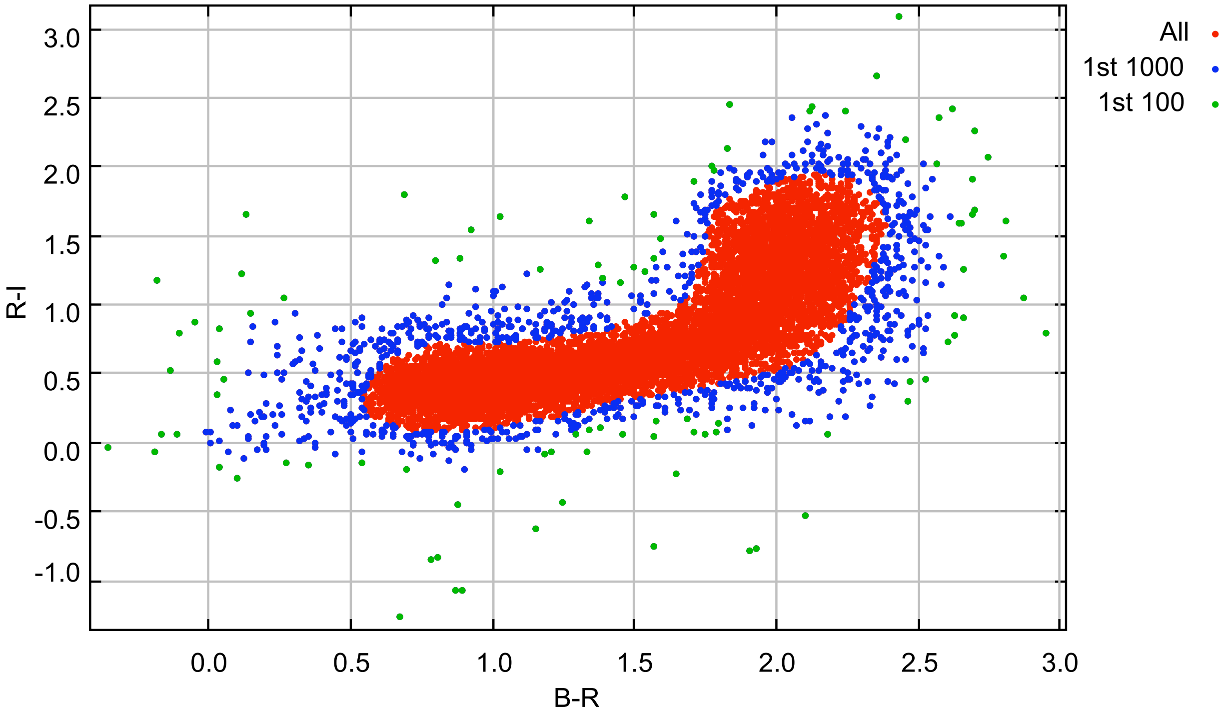



Figure 2: *A $B - R$ vs. $R - I$ colour-colour diagram, generated from the SuperCOSMOS Science Archive. Sources are labelled by how "unusual" they are, quantified using a Gaussian mixture model to describe the distribution of sources in the diagram. The 100 most unusual sources are coloured green, the next 900 blue, with the remaining sources in red.*

planned.

**Data structures**

Progress on (i) has been made already. As part of the VOTECH project's Data Exploration design study (DS6, which is led by Mann) VOTECH-funded staff within WFAU have been investigating how to call data mining algorithms running on sky survey databases from within VO workflows. As a proof-of-concept, a workflow was created to find outliers in a $B - R$ $v$ $R - I$ colour-colour diagram using data from the SuperCOSMOS Science Archive (SSA), through the use of a Gaussian mixture model to describe the distribution of sources in that colour-magnitude space. The results from this are shown in Figure 2, generated by TOPCAT[3]. The outliers in that plot which are blue in both colours are hot stars (e.g. white dwarf stars) and those red in both colours are cool stars (e.g. mid/late M dwarf stars), while those which are blue in $R - I$ and quite red in $B - R$ are candidate high-z quasars. Underlying this analysis is the loading of sky survey data into a k-d tree data structure to speed up the generation of the mixture model, and the success of this proof-of-concept demonstrates that we can readily make available data analysis tools running at the data centre wrapped as web services for VO users. Further work on (i) will continue as originally planned, through prototyping a tool which performs one of the analyses from the UKIDSS science case.

**Analysis of requirements for WSA/VSA data analysis services**

The data analysis services to be developed under the aegis of this theme must meet the requirements of the communities using the WSA and VSA. To ensure that is the case, we shall undertake a systematic analysis of those requirements, resulting in a pair of reports: one for the WSA, to be delivered in October 2008; and one for VISTA, due in October 2009. To help prepare the WSA report, we anticipate that Rimoldini will make a presentation at a forthcoming UKIDSS consortium

[3]http://www.star.bris.ac.uk/~mbt/topcat/

meeting to solicit requirements, as well as seeking more detailed information from the UKIDSS Survey Heads. In the case of the VSA, he will interact with the PIs of the ESO Public Surveys, but this is likely to yield useful information only once users have experience of manipulating significant quantities of VISTA data, hence the later delivery date for this report.

**Light-curve analysis tool for synoptic data**
One specific example of analysis functionality for which there is already a clear requirement is a light-curve tool to support the *VISTA Variables in the Via Lactea* (VVV) survey, which will produce a deep, multi-epoch near-infrared map of the entire Milky Way Bulge and a large portion of the Southern Galactic Plane. This will yield a catalogue of $\sim 5 \times 10^8$ sources, roughly $10^6$ of which will be variables. The VVV catalogue is too large to be readily downloaded by researchers searching for different classes of variable, so it behoves WFAU to develop a light-curve analysis tool to facilitate scientific exploitation of the VVV, and later synoptic surveys, such as PS1, Gaia and LSST (see Section 4(c)). This is better developed in-house, rather than by an external team, because its development is likely to require changes to the database schema used to store the VVV data.

**Data analysis services in VOSpace**
As described in the VDFS Operations case (Section 4(b)), one of the most significant parts of our AstroGrid deployment programme will be the installation of the VOSpace[4] virtual storage component. The initial VOSpace implementation will be a distributed flat-file storage system (like AstroGrid's current MySpace service), but the next stage in its development will include functionality like the "MyDB" system at Johns Hopkins, where users are provided with storage space in an RDBMS instance co-located with the databases from which they wish to extract data. We shall investigate how this facility may be used to allow users to develop stored procedures to run against their VOSpace tables to perform analysis operations efficiently.

**Column-ordered storage**
Another piece of work conducted under the aegis of DS6 has shown[5] how operations involving only a few columns from a wide table (such as the *Source and *Detection tables in the WSA) can be performed much more quickly on data stored in a column-ordered format on disk than through a query on an RDBMS (which stores all data in row order): roughly speaking, the time taken for an analysis requiring only $N$ columns from a table which is $M$ columns wide, can be reduced by a factor of $\sim M/N$, if the data are stored in a column-ordered disk format, and the application code is able to take advantage of appropriate O/S level file mapping techniques. Analysis operations which could benefit from this approach include the generation of summary statistics like histograms and scatter plots, so we shall investigate the development of a web-based statistics package based on column-ordered disk storage of popular attributes. Experience shows that obtaining the full performance benefit from file mapping techniques can require tweaking of the O/S configuration parameters, which is more likely to be possible for dedicated data centre hardware, than for the desktop workstations of science archive users. In the longer term, as disk space is becoming increasingly cheap, we anticipate being able to store on disk, in a column-ordered format, second copies of (at least, subsets of) our Science Archives, and, in due course, to direct queries to the RDBMS-stored copy or the column-ordered copy, based on which will execute each query most efficiently.

**Hardware support for *in situ* analysis of WFAU–curated data**
As noted in our last grant renewal case, we have on offer locally a wide range of hardware on which to test prototype data analysis services – from PC clusters to shared-memory machines and multiprocessor supercomputers – and will produce for our next rolling grant renewal a costed

[4]http://www.ivoa.net/Documents/cover/VOSpace-20070304.html
[5]See http://wiki.eurovotech.org/twiki/bin/view/VOTech/DS6StudyReport

proposal for the hardware needed to support *in situ* analysis of WFAU-hosted datasets, in the light of the experience gained from the remainder of this theme and the reports on WSA and VSA data analysis service requirements.

## 4 KEY DELIVERABLES and MILESTONES

The key deliverables (D) and milestones (M) of the Data Analysis Services theme up to the end of the grant period are as follows:

| Date | Type | Details |
|---|---|---|
| Q1 2008: | M | VISTA Science Archive (VSA) operations begin |
| Q2 2008: | D | Prototype tool implementing one of the principal analyses from the UKIDSS Science Case, using a novel data structure |
| Q3 2008: | D | Report on WSA data analysis service requirements |
| Q3 2008: | M | First VSA survey data releases |
| Q1 2009: | D | Design for light-curve analysis tool |
| Q3 2009: | D | Report on VSA data analysis service requirements |
| Q1 2010: | D | Prototype implementation of light-curve analysis tool |
| Q2 2010: | D | Assessment of hardware requirements for data analysis services, for inclusion in WFAU rolling grant proposal |
| Q2 2010: | M | Next WFAU rolling grant renewal due |
| Q2 2012: | D | *"Production"* system for data analysis services, on dedicated hardware resources |
| Q3 2012: | M | WFCAM Science Archive (WSA) operations end |

## 5 RESOURCES REQUESTED

We request continued support for Rimoldini for the whole lifetime of the new grant. This position will require attendance of various consortium meetings and consultations with survey PIs to elicit requirements, plus attendance of conferences and consortium meetings to present results, so we request £2K per year for travel and subsistence. We request £1K per year for renewal of personal computer hardware, and other consumables.

In Summary:

(a) *Staff.* 1 FTE for five years = 5 d.s.y.
(b) *Travel and subsistence.* £2K/FTE/year = £10K
(c) *Consumables:* £1K/FTE/year = £5K
(d) *Maintenance:* none required
(e) *Equipment:* none required

# Section 4(e) : Project-Related Research

**Staff involved**

N. Cross
N. Hambly
New Synoptic Surveys Scientist

## 1 BACKGROUND

The success of the WFAU programme depends on ensuring that what the Unit delivers is what the community requires. The Investigators can steer the Unit's strategic direction to match developing trends in astronomy, and, within their individual areas of expertise, can help elaborate the broad–brush science requirements often supplied by external sources, but, ultimately, project staff must understand those requirements at the most detailed level, so they can translate them into functional requirements for the systems the Unit will deliver.

This requires a few key project staff to remain research–active in areas directly related to WFAU's project work and, so, in this theme, we request continuation of the 0.3 FTE research allocations for two such existing staff – Cross and Hambly – and inclusion of such an allocation within the work programme of the new Synoptic Surveys Scientist. Together, these three staff will span the science areas covered by WFAU's data curation activities: Cross and Hambly have expertise in extragalactic and Galactic astronomy, respectively, and we anticipate recruiting into the Synoptic Surveys Scientist post someone with a research background in time–domain astronomy.

Cross and Hambly have strong track records in their respective areas, and have both proven to be very productive – together, they have co-authored 43 refereed and 16 non-refereed publications during the period 2005-2007 from a total research allocation of 0.6 FTE over that period – and we expect to hire someone of similar calibre for the Synoptic Surveys Scientist post. The requested project-related research allocations for Cross, Hambly and the new Synoptic Surveys Scientist together represent only 10% of the staff effort for which we seek funding in this proposal, but bring a disproportionate benefit to the WFAU programme as a whole. Not only are they crucial for the recruitment and retention of high–quality staff in key project roles, and provide those staff with the strongest personal motivation for ensuring that WFAU's deliverables are of the highest quality, but they also lead directly to delivery of enhanced functionality: for example, we intend to use the galaxy photometry package being developed by Cross in his research time for future UKIDSS and VISTA data products, and anticipate that the new Synoptic Surveys Scientist may well develop similar new tools for time-domain astronomy as a by-product of the research s/he conducts.

We report below on the research activities of Cross and Hambly during the period 2005-2007 and present their planned project-related programmes for the coming five years. In the case of the new Synoptic Surveys Scientist, we have no named postdoc and so no track record or definite plan to outline, so Section 3.3 below presents several possible project-related programmes that this person might pursue, each in collaboration with one of the Investigators.

## 2 REPORT ON RECENT WORK: 2005 – 2007

### 2.1 Cross:

Currently Cross is working on a first author paper on the galaxy photometry of UKIDSS galaxies, featuring a comparison with 2MASS, and has completed code to accurately measure the half-light semi-major axes of galaxies. The paper discussing this work is in a draft form, but is expected to be submitted in summer 2007 (Cross et al. 2007). This paper uses the extraction of artificial galaxies to understand systematic effects in UKIDSS, multiple real observations of the same targets to understand random errors, and tracks down sources of contamination and incompleteness. This is an important step in understanding both random and systematic errors. This will be a useful reference for all extragalactic astronomers using UKIDSS and 2MASS data, as Cross et al. (2004) was for MGC, SDSS and 2dF data. This work will also prepare the ground for a galaxy photometry package for UKIDSS and VISTA data, which we intend to produce in later releases.

Working with colleagues from the Millennium Galaxy Catalogue (MGC) collaboration and others, Cross has been involved in studies of the space density of galaxies and compact galaxies in the local Universe (Driver et al. 2005; Liske et al. 2006). Cross has been involved in matching the MGC to the UKIDSS-LAS data to study the infrared properties of the local galaxy population. With colleagues at the University of Sussex he is working on measuring the K-band luminosity function of galaxies from UKIDSS and SDSS datasets (Smith et al. 2007). This will use results from Cross et al. (2007) to correct for incompleteness and surface brightness limits.

Finally, Cross has worked with colleagues on HST *Advanced Camera for Surveys* data, in particular on the formation and evolution of galaxies in proto-clusters at high redshift (Overzier et al. 2007; Overzier et al. 2006) and in clusters at $z \sim 1$ (Postman et al. 2005; Goto et al. 2005). Recently Cross has been reducing deep VLA 1.4GHz observations of one of these $z \sim 1$ clusters (RX J0152.7-1357) and matching the radio catalogue to the deep ACS optical imaging and VLT spectroscopic data.

During the reporting period (2005 to 2007), Cross has co–authored 19 peer–reviewed journal papers and 7 non–refereed articles on the galaxy luminosity function, galaxy size distributions, galaxy evolution, surveys and database design. During the last five years (2002 to 2007), refereed/non–refereed papers amount to 42/33.

### 2.2 Hambly: low luminosity degenerate stars

Hambly conducts several lines of research under the broad heading of low luminosity degenerate stars. The greater part of Hambly's research work has been expended as UKIDSS Galactic Clusters Survey[1] working group head. This work consisted of co-ordinating survey observations (preparing the Minimum Schedulable Blocks; taking part in survey observing itself); overseeing science exploitation, e.g. science verification (Lodieu et al. 2007a) and extracting the first science results (Lodieu et al. 2006; 2007b); and preparing the case for renewal (the UKIRT Board recently awarded the UKIDSS GCS a further 35 nights with WFCAM/UKIRT as part of the recent AO for large campaign proposals). Other UKIDSS–related work has included contributions to model predictions of expected brown dwarf numbers in the Large Area Survey (LAS; Deacon & Hambly 2006) as well as to the first discoveries of very cool brown dwarf stars (e.g. Kendall et al. 2007; Warren et al. 2007a). General survey work for UKIDSS has resulted in major contributions to data releases (e.g. Dye et al. 2006; Warren et al. 2007b).

Work on nearby stars from proper motion surveys has continued to be a fruitful area of research. With student Deacon, Hambly has pushed the techniques of proper motion surveys into the infrared (Hambly & Deacon 2005) by exploiting the legacy SuperCOSMOS Sky Survey (SSS; Hambly et al. 2001) in conjunction with 2MASS (Deacon, Hambly & Cooke 2005; Deacon & Hambly 2007). As a member of the Research Consortium On Nearby Stars[2], Hambly has continued to act as proper motion survey specialist, facilitating exploitation of the SSS. This has resulted in further papers whose ultimate aim is the completion of the census of nearby stars (e.g. Finch et al. 2007). Survey

[1] `http://www.roe.ac.uk/~nch/gcs`
[2] `http://www.chara.gsu.edu/RECONS`

discovery papers have resulted in further follow–up activities. RECONS proper motion survey analysis has yielded samples for spectroscopic follow–up (e.g. Subasavage et al. 2007). Finally, in Ducourant et al. (2007) we report the results of a long–term trigonometric parallax campaign which aims to provide further information concerning the evolutionary and kinematic status of the oldest and coolest high–velocity WDs.

During the reporting period (2005 to 2007), Hambly has co–authored 24 peer–reviewed journal papers and 9 non–refereed articles on low luminosity stars, surveys and associated data management. During the last five years (2002 to 2007), refereed/non–refereed papers amount to 40/19. A complete and up–to–date bibliography with ADS links through to online journals is available at `http://www.roe.ac.uk/~nch/publications/`.

# 3 PROPOSED PROGRAMME OF WORK Q2 2008 – Q1 2013

## 3.1 Cross

The galaxy photometry and half-light measurement paper which Cross is currently working on forms the basis of Cross' work over the next few years. Cross intends to measure accurately the mass-size relationship for disk galaxies to understand the role of angular momentum in the formation of spiral disks. Global properties of spiral disks are dependent on the total mass and angular momentum of the dark matter halo (DM), see Fall & Efstathiou (1980). Models of disk formation have consistently shown that the angular momentum is distributed in a log-normal distribution. The mass and angular momentum of the baryons that make up the stars, gas and dust that is visible should be proportional to the mass and specific angular momentum of the DM halo. While angular momentum is difficult to measure, it is related to the mass and size of the stellar disk. De Jong & Lacey (2000), Cross et al. (2001) and Driver et al. (2005) have all looked at the global optical properties of disk galaxies and compared them to the Fall & Efstathiou model, finding some agreement, but also several disparities.

Some of these disparities may be caused by the difficulty of relating luminosity to mass, and measuring disk scale lengths in galaxies where dust and spiral arms are dominating features. Both of these problems are significantly reduced in the near infrared. The WFCAM Science Archive includes data from the UKIDSS-LAS that is cross-matched to optical imaging and spectroscopic data from the SDSS and MGC, and accurate disk-bulge decompostion data from the MGC, is ideal for selecting and understanding the evolution of spiral disks. Recent work by Driver et al. (2007) has empirically measured the optical extinction in spiral disks to be $0.2 - 1.1$ mag depending on inclination. Their models predict a much lower variation $0.0 - 0.5$ mag attenuation in the K-band, with most variation at the highest inclinations. Working with Driver, Cross will test this prediction as part of a programme to understand the systematic errors in the measurement of mass and size.

Other disparities between the model and data may come from dynamical interactions and friction between baryons can change the specific angular momentum of baryons. By removing worries about measuring mass and size, we can start to understand the physics of interactions.

## 3.2 Hambly

Continuing the topics described above, Hambly plans continued involvement in UKIDSS. For the GCS, preparation of the observations, taking part in survey observing, quality control and co-ordination of science exploitation are all important tasks to be done. New research collaborations are already being set up (e.g. with new ESO members in Spain) to ensure the widest possible use of the data. UKIDSS will be approaching completion towards the end of the current grant period, and we anticipate a major effort in co-ordinating pan–European teams involved in getting the most out of the large GCS dataset via robust membership analyses (e.g. Lodieu et al. 2007b) and modelling of such effects as binarity and dynamical evolution (cluster evaporation) on the derived mass functions amongst the diverse set of GCS targets. The UKIDSS wide, shallow survey datasets, in particular the LAS, hold much promise for the study of nearby stars via infrared proper motion surveys (e.g. Hambly & Deacon 2005). Extending this into the southern hemisphere will be possible from late–2008 when it is anticipated that VISTA survey data will first become available; Hambly is a co–I in several of the wide, shallow ESO Public Surveys (e.g. the VISTA Hemisphere Survey) currently at the preparation stages. Even legacy all–sky datasets such as the SSS in conjunction with 2MASS have yet to be fully exploited for the purposes of completing the nearby substellar census. Work in this area is planned with Henry, the lead investigator of RECONS, who is on sabbatical later this year and plans an extended visit to the University of Edinburgh.

Finally, Hambly's research allocation will be used in working on a study of the Galactic white dwarf (WD) populations using the legacy SSS data (e.g. Hambly, Digby & Oppenheimer 2005). With new student Rowell, we plan to select the largest sample of WDs yet defined (e.g. several times larger than that described in Kilic et al. 2006). This sample will be large enough to define more accurately kinematic sub-samples of old disk, thick disk and spheroid WDs, enabling a more accurate determination of their relative numbers, a WD cooling age for the progenitor population (e.g. Harris et al. 2006) and their contribution to the total mass of their respective Galactic components. This is important because even the most recent microlensing experiments (Calchi Novati et al. 2005) stubbornly point to the presence of a significant population of high mass–to–light ratio lenses ($m \sim 0.5$ $M_\odot$) that could be in the form of WDs. Moreover, a large, all–sky WD sample allows the detection of significant numbers of rare, exotic types. The 'ultracool' WDs are an important area to study (e.g. Gates et al. 2004). These objects show remarkable departures from black–body SEDs and their evolutionary status is difficult to elucidate without detailed follow–up of a sample of reasonable size. Recent awards of telescope time on the CTIO 4m (spectroscopy), WHT 4m (spectroscopy) and Liverpool Telescope 2m (imaging for trigonometric parallaxes) are producing an unrivalled dataset for the study of ultracool WDs (Rowell, Hambly & Kilic, in preparation). With these data, we plan a comprehensive study of these fascinating but rare WDs and for the first time, we intend to quantify their affect on ages derived from WD luminosity function modelling (e.g. Harris et al. 2006).

### 3.3 Synoptic Surveys Scientist

We anticipate recruiting into the Synoptic Surveys Scientist position someone with a strong research background in one of the areas of time-domain astronomy directly related to WFAU's future project work on PS1, Gaia and LSST. We expect that s/he will quickly establish an independent programme of research related to WFAU project work, in the way that Cross and Hambly have, and we outline below several possible areas within which s/he could collaborate with the one of the Investigators.

**Identifying RR Lyraes in the outer Milky Way**. The time sampling of PS1 ($4\times$ per year for the entire sky visible from Hawaii in the $3\pi$ survey, and daily for $\sim$6 months a year for the M31 Microlensing survey) will allow the identification and analysis of the variable star content in the Milky Way and throughout the Local Group. Edinburgh scientists plan to exploit the higher temporal sampling, and photometric depth, of the M31 Microlensing dataset to search for, and study, distant Milky Way RR Lyrae stars along the M31 sightline. RR Lyraes are excellent standard candles, so this work will address such issues as the radial profile, and extent, of the stellar halo in this direction and search for localised stellar overdensities in the outer halo. Working with Ferguson and her group, the Synoptic Surveys Scientist would play a major role in developing techniques to identify variable stars within this dataset, and extract and characterise their lightcurves. This work would also feed back directly into the work programmes outlined in Section 4(c) and 4(d), by sharpening scientific requirements for, and prototyping approaches to implementing, the lightcurve analysis and variability-detection services needed for future synoptic sky survey science archives.

**Finding new structures in the local Universe**. PS1 is expected to lead to the discovery of a variety of previously-unknown structures in and beyond the Local Group such as dwarf satellites, faint globular clusters and stellar streams. Indeed, the SDSS has already considerably improved our census of such structures (e.g. Belokurov et al. 2007) and the increased depth and areal coverage of the $3\pi$ survey promises to continue this trend. Ferguson and her group are particularly interested in identifying new structures in external galaxies and the Synoptic Surveys Scientist would play a major role in developing techniques to identify stellar overdensities [in (colour,magnitude,location) space] and extract, and characterise, candidate structures: this could also lead to a generic data analysis tool whose development could be advanced through the work programme outlined in Section 4(d).

**Applying machine learning techniques to transient detection**. One of the major problems in the automated discovery of transient events is the expense of false positive events: e.g. wide-field imaging surveys with the correct cadence can identify large numbers of candidate supernovae, but current methods tend to require visual inspection to weed out the high fraction of false positives. The data volumes generated by synoptic sky surveys like PS1, Gaia and LSST are so large that human intervention for false positive removal will not be possible, especially for those classes of transients for which rapid triggering of follow-up observations is essential. The first steps have recently been taken (Bailey et al. 2007) to apply machine learning techniques (e.g. boosted decision trees, support vector machines) to transient detection in sky survey datasets. These results are promising, but there is a long way to go before they can readily meet the requirements of future surveys. Mann has a long-standing collaboration with researchers interested in the application of machine learning techniques to astronomy (see, e.g. Storkey et al 2004, Taylor 2006, Taylor et al. 2006) and the Synoptic Surveys Scientist would further develop such approaches with the goal of facilitating transient detection in synoptic sky surveys with low false positive rates.

**Finding "Hot Jupiters"**. The presence of "Hot Jupiters" – Jupiter-like planets orbiting very close to their parent stars – can in a general sense be explained by migration of the planet through the circumstellar disc early in the star formation process. What is unclear is the exact details of the process. Do the planets simply spiral inwards, stopping when the circumstellar disc has dissipated? Do the planets become highly eccentric and are they then tidally circularised at twice the Roche radius (Ford & Rasio 2006)? Is the disc truncated at co-rotation by the stellar magnetosphere, producing a cavity in which the migration may stall? Rice is actively studying these question, mainly from a theoretical perspective, but also wants to initiate an observational programme, using PS1. The Synoptic Surveys Scientist would collaborate with Rice in the creation of a sample of "Hot Jupiters" from PS1 data, through accurate transit timing (e.g. Agol et al 2005): this is likely to require quite challenging combined photometric and temporal analysis, and this project would provide an invaluable assessment of whether WFAU's schema designs for synoptic sky survey databases are capable of supporting such analyses, which are found amongst the key science drivers for future synoptic sky surveys.

# 4 KEY DELIVERABLES

Papers, articles and other written reports in peer–reviewed journals, conference proceedings and WWW publications.

# 5 RESOURCES REQUESTED

Our requested research allocation for the period Q2 2008 to Q1 2013 is 4.05 d.s.y. (0.3 FTE for each of Cross, Hambly and the new Synoptic Surveys Scientist) along with conference travel and consumables as follows:

(a) *Staff.* 0.3 FTE Cross, 0.3 FTE Hambly, 0.3 FTE Synoptic Surveys Scientist = 4.05 d.s.y. in total)
(b) *Travel and subsistence.* £2K/FTE/year totalling £8.1K for conference attendance
(c) *Consumables:* £1K/FTE/year = £4.05K in total
(d) *Maintenance:* none required
(e) *Equipment:* none required

– see the JeS–1 form for details.

# 6 ADDITIONAL REFERENCES


Agol, E., et al, 2005, MNRAS, 359, 576

Belokurov et al. 2007, ApJL, 654, 897

Calchi Novati, S. et al., 2005, A&A, 443, 911

**Cross, N.J.G.**, et al. 2001, MNRAS, 324, 825

**Cross, N.J.G.** et al. 2004, MNRAS, 349, 576

**Cross, N.J.G.** et al. 2007, in preparation

De Jong R.S. & Lacey C. 2000, ApJ, 545, 781

Deacon, N.R., **Hambly, N.C.**, Cooke, J.A., 2005, A&A, 435, 262

Deacon, N.R., **Hambly, N.C.**, 2006, MNRAS, 371, 1722

Deacon, N.R., **Hambly**, N.C., 2007, A&A, in press (astro–ph/0702692)

Driver S.P., Liske J., **Cross N.J.G.**, De Propris R., Allen P.D. 2005, MNRAS, 360, 81

Driver S.P. et al. 2007, MNRAS, accepted

Ducourant, C., Teixeira, R., **Hambly, N.C.**, Oppenheimer, B.R., Hawkins, M.R.S., Rapaport, M., Modolo, J., Lecampion, J.F., 2007, A&A, in press (astro–ph/07040355)

Dye, S. et al. incl. **Hambly, N.C.**, 2006, MNRAS, 372, 1227

Fall S.M., Efstathiou G., 1980, MNRAS, 193, 189

Finch, C.T., Henry, T.J., Subasavage, J.P., Jao, W.-C., **Hambly, N.C.**, AJ, in press (astro–ph/0703133)

Ford, E.B., Rasio, F.A., 2005, ApJ, 638, L45

Gates, E. et al., 2004, ApJ, 612, L129

Goto T., Postman M., **Cross N.J.G.** et al. 2005, ApJ, 621, 188

Liske J., Driver S.P., Allen P.D., **Cross N.J.G.**, De Propris R. 2006, MNRAS, 369, 1547

Kendall, T.R. et al. incl. **Hambly, N.C.**, 2007, A&A, 466, 1059

**Hambly, N.C.**, et al. 2001, MNRAS, 326, 1279

**Hambly, N.C.**, Digby, A.P., Oppenheimer, B.R., 2005, In: Proceedings of the 14th European Workshop on White Dwarfs, eds. D. Koester & S. Moehler, ASP Conf. Ser., 334, 113

**Hambly, N.C.**, Deacon, N.R., 2005, AN, 326, 1011

Harris, H.C. et al., 2006, AJ, 131, 571

Kilic, M. et al., 2006, AJ, 131, 582

Lodieu, N., **Hambly, N.C.**, Jameson, R.F., 2006, MNRAS, 373, 95

Lodieu, N., **Hambly, N.C.**, Jameson, R.F., Hodgkin, S.T., Carraro, G., Kendall, T.R., 2007a, MNRAS, 374, 372

Lodieu, N., Dobbie, P.D., Deacon, N.R., Hodgkin, S.T., **Hambly, N.C.**, Jameson, R.F., 2007b, MNRAS, submitted

Overzier R.A., Bouwens R.J. **Cross N.J.G.**, et al. 2007, ApJ, submitted

Overzier R.A., Miley G.K., Bouwens R.J., **Cross N.J.G.** et al. 2006, ApJ, 637, 580

Postman M., Franx M., **Cross N.J.G.** et al. 2005, ApJ, 623, 721

Smith A. et al. 2007, in preparation.

Storkey A.J, Hambly N.C., Williams C.K.I, Mann R.G., 2004, MNRAS, 347, 36

Subasavage, J.P., Henry, T.J., Bergeron, P., Dufour, P., **Hambly, N.C.**, Beaulieu, T.D., 2007, AJ, in press (astro–ph/07040894)

Taylor, E.L., 2006, PhD thesis, University of Edinburgh

Taylor, J.D., Taylor, E., Mann, R.G., Prina Ricotti D., Bose R., 2006, in Astronomical Data Analysis Software and Systems XV ASP Conference Series, Vol. 351, Proceedings of the Conference Held 2-5 October 2005 in San Lorenzo de El Escorial, Spain. Edited by Carlos Gabriel, Christophe Arviset, Daniel Ponz, and Enrique Solano. San Francisco: Astronomical Society of the Pacific, 2006., p.467

Warren, S.J. et al. incl. **Hambly, N.C.**, 2007a, in preparation

Warren, S.J. et al. incl. **Hambly, N.C.**, 2007b, MNRAS, 375, 213

# Section 4(f) : Project Management

**Staff involved** P. Williams (until Q1 2010)
New Technical Lead/Project Manager (from Q4 2009)

# 1 INTRODUCTION

We are presenting *Project Management* as a separate theme in this proposal to give us the opportunity to justify the increased level of project management effort in the later portion of the grant for which we are requesting support. This additional effort is required for two reasons: (i) WFAU's programme is increasingly centred on large, multi-institutional – and, often, multi-national – projects, which results in an increased overhead, in terms of formalised procedures for project planning and for progress, and financial, reporting; and (ii) the scale and complexity of the computational side of WFAU's work is set to increase markedly with the start of the Synoptic Sky Surveys theme.

Our solution to this latter problem is to create a new, half-time WFAU Technical Lead position, starting (along with the Synoptic Sky Surveys theme) in Q4 2009 and, for practical reasons, to merge this with the continuation (after Williams' retirement) of the existing half-time WFAU Project Manager post, to create a full-time position. We anticipate recruiting for this position someone from the commercial IT sector with experience managing software development projects, who would meet the requirements of both halves of the post. We view this position as being of significant strategic importance for the future of WFAU, both facilitating successful execution of the currently proposed programme and aiding WFAU involvement in the kind of large-scale, computationally-intensive projects which will come to dominate wide-field astronomy in the future. We expect that fractions of the cost of this post will be recovered from other project-specific grants (e.g. a post-2009 AstroGrid operations budget) at various times, so that what we are requesting here is, essentially, underwriting for a key WFAU post.

# 2 CURRENT PROJECT MANAGEMENT ARRANGEMENTS

## 2.1 The role of the Investigators

Lawrence has been the PI of WFAU since its creation in 1999, but, in recent years, his commitments as Head of School of Physics have necessitated a reduction in the time he can devote to WFAU. His leadership responsibilities have largely been taken up by Mann, who was appointed to a lectureship associated with e-Science and wide-field astronomy in 2004, partly to underpin the University's long-term commitment to WFAU. Mann is now the WFAU PI, but Lawrence remains Co-PI, with a continuing strong role in the Unit, particularly as regards strategic discussions relating to future international collaborations. A WFAU Management Group comprising the Investigators, plus Hambly and Williams, meet regularly.

The other Investigators on the current WFAU rolling grant are Dunlop and Peacock. Dunlop is the Head of the Institute for Astronomy (IfA), and part of his role as a Co-I is to ensure that WFAU is well supported within the University: for example, he is currently seeking University funding for refurbishment of the former ROE Plate Library space, to produce further office space for WFAU. Peacock provides particular expertise relating to cosmological analyses of WFAU datasets – for example, the photometric recalibration of the SuperCOSMOS Sky Surveys dataset using SDSS and 2MASS was the result of his work to create a homogeneous all-sky galaxy catalogue – and his role as chair of the Science Project Oversight Committee of the PS1SC will prove very valuable for facilitating publication of public PS1 datasets. This proposal includes two new Co-Is – Ferguson and Rice – who will bring additional scientific expertise to bear on WFAU's programme. Ferguson leads a new stellar populations research group in the IfA, funded by an EU Marie Curie *Excellence* grant, and her expertise in the science requirements from studies of the Local Group will be an important resource, both for the existing WFCAM and VISTA Science Archives and for the new Synoptic Sky Surveys theme. Rice's expertise is in the study of protoplanetary disks. The bulk of his own research is theoretical, but his knowledge of the observational side of the discovery, and study, of exoplanetary systems will also be of great value to the Synoptic Sky Surveys theme, since exoplanets provide prime science drivers for PS1, Gaia and LSST.

### 2.2 The role of the Project Manager

Day-to-day management of WFAU is undertaken by Williams. Williams is currently funded at the 0.2 FTE level by the present WFAU rolling grant and at the 0.3 FTE level by the VDFS project, whose final extension is being subsumed into this rolling grant proposal: we therefore request that Williams be funded at the 0.5 FTE level from the start of the new grant.

The activities undertaken by the project manager include: financial planning and reporting; staffing tasks, such as appraisals of existing staff, plus recruitment of new staff, including grading of new posts and managing the recruitment process; liaison with Swindon office; and, currently, overseeing maintenance contracts for equipment. A characteristic of large, distributed projects is the necessity for quite formalised planning and reporting procedures, so it seems inevitable that the project management overhead will increase during the period of the new grant. For example, Williams is a member of the VDFS Management Team, overseeing the whole project, and to which he reports on the WFAU work monthly.

## 3 THE NEED FOR A WFAU TECHNICAL LEAD

Since its creation in 1999, the computational side of WFAU's work has changed radically. In the two years of WFCAM survey operations, WFAU have received from Cambridge ∼35 TB of data, which is more than was produced during the whole ten-year SuperCOSMOS plate-scanning programme. In the last five years, WFAU staff have progressed from having no experience of relational database management systems (RDBMS) to curating two multi-TB RDBMS-based science archives, with a third soon about to start ingesting data. Before the advent of the Virtual Observatory (VO), WFAU's sole computational interaction with the outside work was via webpages and simple file download via ftp and http, while now, as the VO matures into a production system, WFAU becomes one node in a globally-connected system.

This transition to date has been made possible by successful on-the-job re-training of existing staff, and the hiring of good staff with relevant IT experience, plus the dedicated funding available for VO development through AstroGrid. Several factors dictate that the next phase of this transition will not be so easy. Firstly, further VO development is likely to switch soon from being the task of a dedicated project, AstroGrid, to being just one of the activities undertaken by the data centres, who must be ready to work within the existing AstroGrid framework, and to take advantage of the opportunities it will present for future use of e-Science/Grid technologies. Secondly, our Gaia CU5 work will require our designing a data archive/repository system capable of interacting with the high-throughput iterative processing system, which adds design requirements we have not had to consider so far. Thirdly, our Gaia work for both CU5 and CU9 will require us to create and maintain good interfaces with the Gaia central database at ESAC, which is a requirement we have not faced before. Fourthly, the volumes of the datasets curated by WFAU will soon become so great that we ourselves shall have to maintain hardware for supporting compute-intensive *in situ* data analysis services, as well as our existing data storage hardware. Fifthly, the computational scale of LSST data analysis and archiving is two orders of magnitude greater than anything we have yet faced, and may well require a significant change to basic assumptions as to how our science archive systems work.

The rate of change of the computational scale, and complexity, of the WFAU programme will be at its maximum when we start significant work on the synoptic sky surveys theme. We propose, therefore, to appoint, at the same time as that theme starts, a WFAU Technical Lead, who will have overall responsibility for the computational side of WFAU work. We expect to recruit this person from the commercial IT sector, although qualified candidates are also likely to be found in academic supercomputing centres. What is clear is that we require someone with experience of leading significant distributed computing projects, with both the technical expertise required to ensure that WFAU takes full advantage of developing e-Science/Grid computing technologies (and their analogues in the commercial IT sector) and the personal qualities and experience needed to take over Williams' role as WFAU Project Manager when he retires. Such people do not come cheap, but the requirements of WFAU's future programme necessitate the presence within the Unit of someone who can take control of its computational activities, leaving Hambly to lead the astronomical side of WFAU project work, and raising the professionalism of WFAU's IT capabilities another notch.

## 4 KEY DELIVERABLES and MILESTONES

The key project management deliverables (D) and milestones (M) during the coming grant period are as follows:

| Date | Type | Details |
|---|---|---|
| Q4 2008?: | M | Release of AO for Gaia CU9 |
| Q? 2009: | D | Answer to AO for Gaia CU9 |
| Q4 2009: | M | Recruitment of Synoptic Surveys Scientist and Technical Lead |
| Q2 2010: | M | Recruitment of developer for Synoptic Sky Surveys theme |
| Q2 2010: | D | Next WFAU rolling grant proposal |

There will also be significant engagement by the Project Manager/Technical Lead in the development of a European LSST consortium, and the technical planning for LSST data management and analysis, but the timescales for these are TBC.

## 5 RESOURCES REQUESTED

We request funding for Williams as WFAU Project Manager at the 0.5 FTE level until his retirement in Q1 2010: this level of effort is the same as the sum of his current contributions to managing staff on the WFAU rolling grant and the VDFS, the completion of which is being rolled into this grant. We request the appointment of a new member of staff in Q4 2009, half of whose work will be a continuation of Williams' WFAU Project Manager line and half of which comprises the new Technical Lead post. This appointment date allows for approximately four months of overlap with Williams, making possible a smooth transition in advance of the next WFAU rolling grant renewal. We expect that some of the cost of this post will be recovered from other sources during the coming grant period, although, clearly, there is a limit to the work that the holder of this post can undertake, so the level of project management provision within the Unit will have to be kept under review.

In total, we request 4.4 d.s.y of staff effort for project management in the new grant. For travel and subsistence we request support at the level of £2K per d.s.y for attendance at meetings, plus £3K to cover the cost of recruiting the three new staff we intend to hire during the coming grant period. We also request £1K/FTE/year for miscellaneous consumables, plus £2K to cover the cost of a new workstation for the Technical Lead.

In summary:

(a) *Staff.* 4.4 d.s.y.
(b) *Travel and subsistence.* 4.4 $\times$ £2K/d.s.y + £3K = £11.8K
(c) *Consumables:* £1K/d.s.y + £2K = £6.4K
(d) *Maintenance:* none required
(e) *Equipment:* none required

– see the JeS–1 form for details.

## Section 5 Supplementary Information

**5(a) Publication summary table**

**5(b) Summary of grant holdings**

**5(c) Gantt charts for staff**

**5(d) Outreach plan**

**5(e) Knowledge Exchange plan**

## Section 5(a) Publication Summary Table

We provide below a table summarising the publications of the Investigators and Project Staff during the calendar years 2002-2006 inclusive.

| | Refereed Publications | Ref. Publ. (1st Author) | Technical Publications | Tech. Publ. (1st Author) |
|---|---|---|---|---|
| R.G. Mann | 22 | 1 | 18 | 4 |
| A. Lawrence | 19 | 0 | 18 | 5 |
| J.S. Dunlop | 56 | 3 | 0 | 0 |
| A.M.N. Ferguson | 24 | 3 | 0 | 0 |
| J.A. Peacock | 55 | 0 | 0 | 0 |
| W.K.M. Rice | 21 | 12 | 0 | 0 |
| | | | | |
| J. Bryant | 2 | 0 | 4 | 1 |
| R.S. Collins | 1 | 1 | 0 | 0 |
| N.J.G. Cross | 41 | 1 | 3 | 0 |
| N.C. Hambly | 29 | 2 | 8 | 2 |
| M.A. Read | 7 | 0 | 6 | 1 |
| L.G. Rimoldini | 0 | 0 | 0 | 0 |
| E.T.W. Sutorius | 4 | 0 | 6 | 0 |
| | | | | |
| P.M. Williams | 8 | 0 | 8 | 1 |

## Section 5(b) Summary of Grant Holdings

We provide overleaf a table showing all the PPARC/STFC grants currently held in the Institute for Astronomy.

## Current PPARC/STFC grants held in Institute for Astronomy

| From | Date to | Value | Grant Reference No | UoE Ref | Title | PI | Staff funded | FTE | |
|---|---|---|---|---|---|---|---|---|---|
| 10/1/04 | 9/30/07 | 100,436 | PPA/P/S/2003/00275 | R37851 | PPARC Fellowship | Dr Olivia Johnson | Dr Olivia Johnson | 100% | |
| 10/1/04 | 9/30/07 | 183,971 | PPA/A/R/2003/00274 | R37887 | Advanced Fellowship | Dr David Bacon | Dr David Bacon | 100% | |
| 10/1/04 | 3/31/08 | 923,390 | PP/C503062/1 | R39052 | VEGA: VDFS component | Prof A Lawrence | Dr Nigel Hambly | 70% | Development of Science Archives for UKIRT WFCAM and VISTA data. To be subsumed in new WFA RG to extend/complete project. |
| | | | | | | | Dr Ross Collins | 100% | |
| | | | | | | | Dr Nick Cross | 70% | |
| | | | | | | | Mr Mike Read | 100% | |
| | | | | | | | Dr Eckhard Sutorius | 100% | |
| | | | | | | | Dr Peredur Williams | 30% | |
| 10/1/04 | 12/31/07 | 232,835 | PP/D508147/1 | R39357 | AstroGrid2 in Edinburgh | Prof A. Lawrence | Ms Kona Andrews | 100% | Close coordination with WFA RG staff |
| | | | | | | | Mr Mark Holliman | 10% | |
| 10/1/05 | 9/30/08 | 120,021 | PP/C504819/01 | R39572 | Fellowship | Dr P. Norberg | Dr Peder Norberg | 100% | |
| 1/1/06 | 3/31/11 | 1,465,834 | PP/D00102 | R39873 | PDRA Rolling Grant | Prof J Dunlop | Dr Angela Mortier | 100% | Institute for Astronomy cosmology and large-scale structure rolling grant |
| | | | | | | | Dr Ben Panter | 50% | |
| | | | | | | | Dr Eric Tittley | 100% | |
| | | | | | | | Dr F. Simpson | 100% | |
| | | | | | | | Dr Michele Cirasuolo | 100% | |
| | | | | | | | Dr P Simon | 100% | |
| 4/1/05 | 3/31/10 | 1,401,776 | PP/C506764 | R39874 | Wide Field Astronomy | Prof A Lawrence | Mr Johann Bryant | 100% | Current Wide Field Astronomy rolling grant |
| | | | | | | | Dr Lorenzo Rimoldini | 100% | |
| | | | | | | | Dr Nigel Hambly | 30% | |
| | | | | | | | Dr Nick Cross | 30% | |
| | | | | | | | Dr Peredur Williams | 20% | |
| | | | | | | | Mr Mark Holliman | 30% | |
| 4/1/06 | 3/31/08 | 20,817 | PP/D/003040/1 | RA0006 | PATT travel | Prof J Dunlop | (no-one) | | |
| 1/1/07 | 6/30/11 | 1,104,516 | PP/D006503/1 | | Gaia data flow system: Edin | Prof A. Lawrence | (Hambly) | | Parts of Hambly, Collins, Cross and VEGA developer will be funded by this - see Gantt chart |
| | | | | | | | (Collins) | | |
| | | | | | | | (Cross) | | |
| | | | | | | | (VEGA developer) | | |

## Section 5(c) Gantt Charts for Staff

We provide a Gantt chart for project staff, in the specified format. We have divided it into two pages for clarity.

**ROLLING GRANT GANTT CHART**
**Dr R. G. Mann - University of Edinburgh (1/2)**

| **Existing WFA award PP/C506764** Prof A Lawrence | YEAR 2 | YEAR 3 | YEAR 4 Year 1 Apr-08 | YEAR 5 Year 2 Apr-09 | Year 3 Apr-08 | Year 4 Apr-09 | Year 5 Apr-10 | **REQUEST** |
|---|---|---|---|---|---|---|---|---|
| VDFS PP/C503062/1 RA3 Hambly | 70% | 70% | **20%** | **10%** | | | | (a) VDFS development **RA3 Hambly** |
| **WFA (f) Research** RA3 Hambly | 30% | 30% | 20% | 10% | | | | |
| | | | **30%** | **30%** | **30%** | **30%** | **30%** | (e) Research **RA3 Hambly** |
| | | | | **10%** | **20%** | **57.5%** | **70%** | (c) Synoptic Surveys **RA3 Hambly** |
| VDFS PP/C503062/1 RA1A Cross | 70% | 70% | **45%** | **10%** | | | | (a) VDFS development RA1A Cross |
| **WFA (f) Research** RA1A Cross | 30% | 30% | 20% | | | | | |
| | | | **30%** | **30%** | **30%** | **30%** | **30%** | (e) Research RA1A Cross |
| | | | | **10%** | **20%** | **57.5%** | **70%** | (c) Synoptic Surveys RA1A Cross |
| VDFS PP/C503062/1 RA1A Collins | 100% | 100% | **25%** | | | | | (a) VDFS development RA1A Collins |
| | | | | | | **75%** | **100%** | (c) Synoptic Surveys |
| VDFS PP/C503062/1 RA1A TBA | | 50% | **75%** | **25%** | | | | (a) VDFS development RA1A TBA |
| VDFS PP/C503062/1 RA2 Read | 100% | 100% | **50%** | **25%** | | | | (a) VDFS development RA2 Read |
| | | | **50%** | **75%** | **100%** | **100%** | **100%** | (b) Archive Operation Ra2 Read |
| VDFS PP/C503062/1 RA1A Sutorius | 100% | 50% | | | | | | |
| **WFA (a) Archive Ops** RA1A Sutorius | | 50% | 100% | 100% | | | | |
| | | | **100%** | **100%** | **100%** | **100%** | **100%** | (b) Archive Operations RA1A Sutorius |
| **WFA (a) Archive Ops** RA1A Bryant | 100% | 100% | 100% | 100% | | | | |
| | | | **100%** | **100%** | **100%** | **100%** | **100%** | (b) Archive Operations RA1A Bryant |

ROLLING GRANT GANTT CHART
Dr R. G. Mann - University of Edinburgh (2/2)

| **Existing WFA award PP/C506764** Prof A Lawrence | YEAR 2 | YEAR 3 | YEAR 4 Year 1 Apr-08 | YEAR 5 Year 2 Apr-09 | Year 3 Apr-08 | Year 4 Apr-09 | Year 5 Apr-10 | **REQUEST** |
|---|---|---|---|---|---|---|---|---|
| **WFA (b) Data mining apps** RA1A Rimoldini | 100% | 100% | 100% | | | | | |
| | | | **100%** | **100%** | **100%** | **100%** | **100%** | (d) Data anlysis services RA1A Rimoldini |
| | | | | **35%** | **70%** | **70%** | **70%** | (c) Synoptic Surveys (Sci) RA1A sp34 NEW POST |
| | | | | **15%** | **30%** | **30%** | **30%** | (e) Research RA1A sp34 NEW POST |
| | | | | | **100%** | **100%** | **100%** | (c) Synoptic Surveys (Dev) RA1A sp34 NEW POST |
| **WFA Systems Manager** RA1A Holliman | 26% | 36% | 34% | 26% | | | | |
| | | | **70%** | **70%** | **70%** | **70%** | **70%** | (f) System Manager RA1A Holliman |
| VDFS PP/C503062/1 RA2 Williams | 30% | 30% | | | | | | |
| **WFA Manager** RA2 Williams | 20% | 20% | 10% | 10% | | | | |
| | | | **50%** | **42%** | | | | (f) Manager RA2 Williams |
| | | | | **25%** | **50%** | **50%** | **50%** | (f) Manager RA2 sp34 NEW POST |
| | | | | **25%** | **50%** | **50%** | **50%** | (f) Technical Lead RA2 sp34 NEW POST |
| **WFA Secterarial** Gibson | 21% | 28% | 27% | 21% | | | | |
| | | | **50%** | **50%** | **50%** | **50%** | **50%** | (f) Secretarial/Clerical EU5 Dupin |
| | | | 15% | 15% | 15% | 15% | 15% | **PI Dr R. G. Mann** |
| | | | 5% | 5% | 5% | 5% | 5% | **co-I Prof A Lawrence** |
| | | | 5% | 5% | 5% | 5% | 5% | **co-I Prof J S Dunlop** |
| | | | 5% | 5% | 5% | 5% | 5% | **co-I Prof J A Peacock** |
| | | | 5% | 5% | 5% | 5% | 5% | **co-I Dr A M Ferguson** |
| | | | 5% | 5% | 5% | 5% | 5% | **co-I Dr K Rice** |

## Section 5d — Outreach Plan

WFAU continues to make major contributions in the area of public outreach locally, nationally and further afield. Wide-field astronomy is of direct appeal to the layman because of the aesthetic value of panoramic imaging observations. The legacy Schmidt photographs collection in the (now relocated) plate library forms the cornerstone of the annual site open days (two days every year), and WFAU staff prepare exhibition posters, computer demonstrations and light-table presentations of interesting film copies to engage public interest. These exhibitions are always a popular part of the site-wide open days, and over the course of the open weekend typically one to two thousand members of the public will visit. From the national and international standpoint, we note that our legacy photographic plate archives of digital scans from the (now completed) SuperCOSMOS scanning programme are publically accessible via the Internet, so digital images of any part of the sky are freely available for download. As the current generation of infrared sky surveys with WFCAM and VISTA enter the public domain following their proprietary periods, we anticipate supplementing legacy photographic public access and associated example panoramic images (e.g. the SuperCOSMOS Halpha Survey of the Galactic plane, see http://surveys.roe.ac.uk/ssa/hablock/hafull.html) with higher resolution, infrared colour images (e.g. the UKIDSS Galactic Plane Survey, see http://surveys.roe.ac.uk/data/mar/gps/gps.html). WFAU plans to continue these activities and play a full part in public outreach, both locally and internationally.

## Section 5e — Knowledge Exchange Plan

As regards knowledge exchange, we present no formal plan but would like the following to be noted. Firstly, the SuperCOSMOS scanning programme ended one year ago, and the facility needed to be decommissioned to allow redevelopment of its environment. At the core of the facility was a Leitz Precision Measuring Machine PMM654, adapted for plate scanning at sub-micron precision via housing in a temperature controlled class 100 clean room and the addition of a CCD optical scanning head. Following lengthy discussions with several interested parties, a home was found for the machine, its control electronics and the clean room. The equipment was recently relocated to a German company (Industrielle CNC und Messtechnik in Waldsolms) at no cost to STFC despite the considerable cost of dismantling and removal. The core optical scanning head components, designed and built in the ROE Technology Division were retained as an integral part of the system, vital to the new commercial application. Secondly, in our recent VDFS science archive development programme we have made extensive use of commercial database management system software - Microsoft SQL Server running on Windows 2003 Advanced Server Enterprise edition - through the generous support of Microsoft's Development Network Academic Alliance. During the course of this work, we have built up a mutually beneficial working relationship with Microsoft's Bay-Area research division, feeding back bug reports and optimisation techniques (e.g. scalable, billion-row binary loading short-cuts) to the SQL Server development team. Thirdly, our dealings with our local hardware supplier Eclipse Computing of Ayrshire have been most beneficial to both parties. We have developed low-cost, high performance and scalable database server solutions based on common 32-bit PC architecture at a fraction of the cost of higher-end 64-bit systems (cf. Sun Microsystems or IBM).