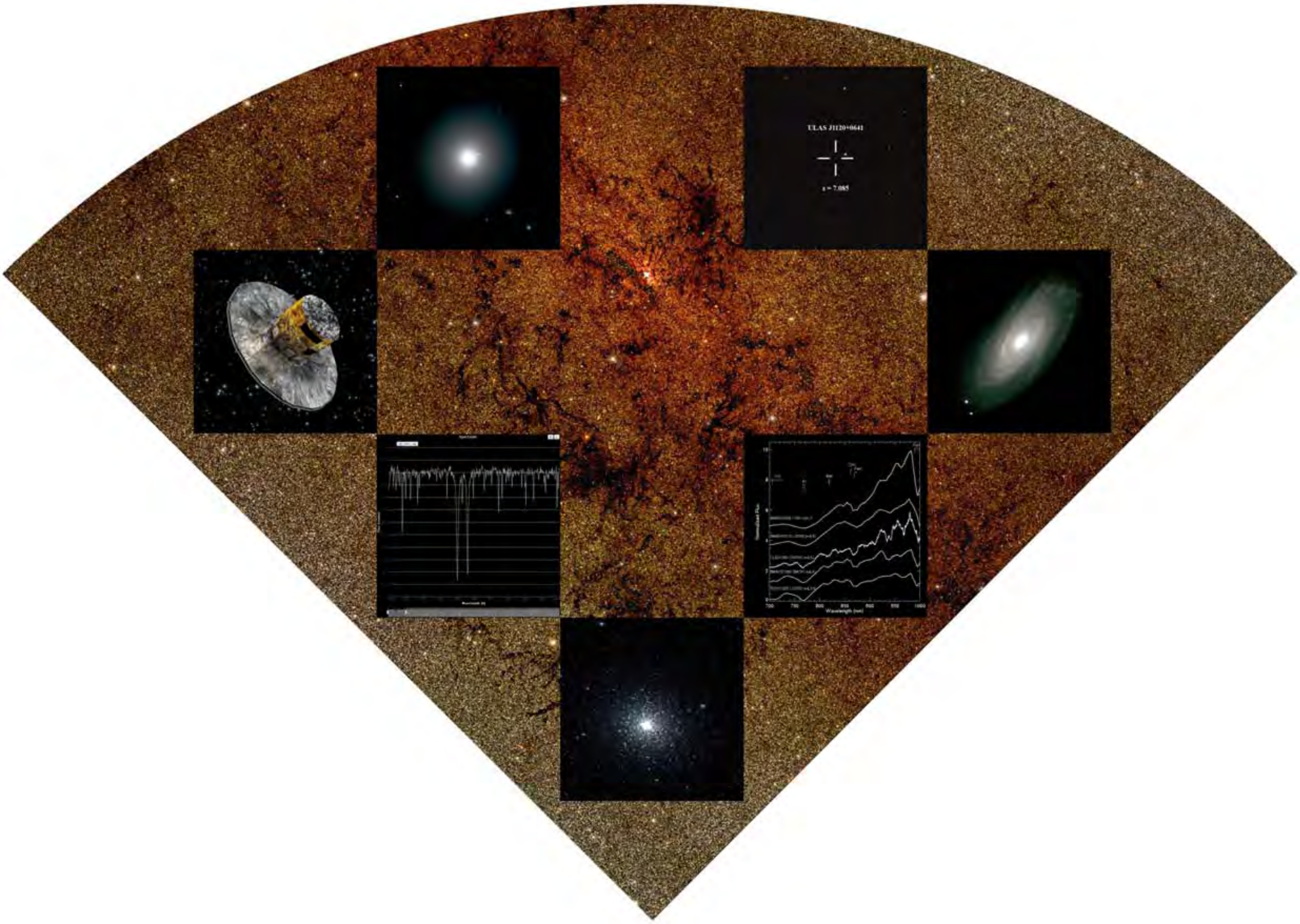

Wide-Field Astronomy at Edinburgh 2013-2016

R.G. Mann
A.M.N. Ferguson
N.C. Hambly
A. Lawrence
K.T. Noddle

Cover Image

The background image shows a mosaic of the Galactic Centre region constructed from data from the VISTA Variables in the Via Lactea (VVV) survey. The seven inset panels are (reading clockwise from the top-right) as follows:

- ULAS J1120+0641, the first quasar found at z>7, detected in the UKIDSS Large Area Survey;
- The UKIDSS view of the spiral galaxy M88;
- The optical spectrum (the middle line) of the new L subdwarf ULAS1350 identified photometrically by Lodieu and collaborators by searching a region of overlap between the UKIDSS LAS and the SDSS. It is the fifth L subdwarf to be found, and furthest known to date;
- The 47 Tucanae globular cluster, from the VISTA Magellanic Survey;
- A spectrum from the Gaia-ESO survey archive plotted interactively using the GES archive interface;
- The Gaia satellite; and
- M87 from UKIDSS.

Contents

Summary Science Case **4**

0.1 INTRODUCTION . . . 4

0.2 WFAU ACHIEVEMENTS 2009 – 2012 . . . 4

0.2.1 Achievements: Facilitating UKIDSS science through WSA operations . . . 4

0.2.2 Achievements: Supporting VISTA Public Survey PIs and Consortia . . . 4

0.2.3 Achievements: Preparing for Gaia exploitation . . . 5

0.2.4 Achievements: Leading Virtual Observatory developments in the UK . . . 5

0.2.5 Achievements: Publicising UKIDSS and VISTA science . . . 5

0.3 PROPOSED PROGRAMME Q2 2013 – Q1 2016 . . . 6

0.4 THE WFAU VISION FOR FUTURE SURVEY ARCHIVES . . . 6

Project 1: Enabling WFCAM exploitation **8**

1.1 BACKGROUND . . . 8

1.2 REPORT ON RECENT WORK: 2009 – 2012 . . . 9

1.2.1 Achievements: data releases . . . 9

1.2.2 Achievements: archive usage . . . 9

1.3 PROPOSED PROGRAMME OF WORK Q2 2013 – Q1 2016 . . . 10

1.3.1 Final developments . . . 10

1.3.2 Final data releases . . . 12

1.3.3 Legacy archive consolidation and operations . . . 12

1.4 KEY DELIVERABLES and MILESTONES . . . 12

1.5 RESOURCES REQUESTED . . . 13

1.6 ADDITIONAL REFERENCES . . . 13

Project 2: Enabling VISTA exploitation **14**

2.1 BACKGROUND . . . 14

2.1.1 The VISTA Data Flow System . . . 14

2.1.2 The VSA and the ESO SAF . . . 14

2.2 REPORT ON RECENT WORK: 2009 – 2012 . . . 15

2.2.1 Achievements: data releases . . . 15

2.2.2 Achievements: archive usage . . . 15

2.2.3 Achievements: complementary optical surveys . . . 16

2.2.4 Achievements: delivery of data to the ESO-SAF . . . 17
2.3 PROPOSED PROGRAMME OF WORK Q2 2013 − Q1 2016 . . . 17
2.3.1 Development . . . 17
2.3.2 Operations . . . 18
2.4 KEY DELIVERABLES and MILESTONES . . . 20
2.5 RESOURCES REQUESTED . . . 21
2.6 ADDITIONAL REFERENCES . . . 21

Project 3: Enabling Gaia exploitation **22**
3.1 BACKGROUND . . . 22
3.1.1 The Gaia Mission . . . 22
3.1.2 Preparing for exploitation of Gaia data . . . 22
3.1.3 Obtaining ancillary data to enhance Gaia exploitation . . . 23
3.2 REPORT ON RECENT WORK: 2009 − 2012 . . . 23
3.2.1 Achievements: GES end–user archive prototype . . . 23
3.2.2 Achievements: WFAU co-opted onto the Gaia Archive Preparation group . . . 23
3.2.3 Achievements: preliminary archive development programme . . . 25
3.3 PROPOSED PROGRAMME OF WORK Q2 2013 − Q1 2016 . . . 25
3.3.1 Final developments for the Gaia–ESO Survey science archive . . . 25
3.3.2 GES science archive operations . . . 25
3.3.3 Participation in Gaia DPAC CU9 . . . 25
3.4 KEY DELIVERABLES and MILESTONES . . . 27
3.5 RESOURCES REQUESTED . . . 28
3.6 ADDITIONAL REFERENCES . . . 28

Project 4: Enabling Pan-STARRS1 exploitation **29**
4.1 BACKGROUND . . . 29
4.2 REPORT ON RECENT WORK: 2009 − 2012 . . . 30
4.3 PROPOSED PROGRAMME OF WORK Q3 2014 − Q1 2016 . . . 30
4.3.1 Installation of a copy of the PS1 catalogue . . . 30
4.3.2 Supporting determination of optical counterparts of LOFAR surveys . . . 31
4.4 KEY DELIVERABLES and MILESTONES . . . 33
4.5 RESOURCES REQUESTED . . . 33

Project 5: Future archive infrastructure **34**
5.1 BACKGROUND . . . 34
5.2 REPORT ON RECENT WORK: 2009 − 2012 . . . 34
5.2.1 Achievements: a new interactive user interface . . . 35
5.2.2 Achievements: federated data access . . . 35
5.2.3 Achievements: a data annotation service . . . 37
5.3 PROPOSED PROGRAMME OF WORK Q2 2013 − Q1 2016 . . . 37
5.3.1 "Productisation" of prototype services . . . 38

5.3.2 User storage in the archive . . . 38
5.3.3 Replacing the AstroGrid Registry . . . 39
5.3.4 Authentication and authorisation . . . 39
5.3.5 Virtual Machines . . . 39
5.4 KEY DELIVERABLES and MILESTONES . . . 40
5.5 RESOURCES REQUESTED . . . 40

Cross-Project Resources **41**
6.1 INTRODUCTION . . . 41
6.2 Technical Leadership and Project Management . . . 41
6.3 Computing Systems Administration . . . 42
6.4 Administrative Support . . . 43
6.5 RESOURCES REQUESTED . . . 43

Appendix A: Summary Publication Statistics **44**

Appendix B: Summary of STFC grants held at the IfA **46**

Appendix C: Staff Gantt Chart **47**

Appendix D: Resource Summary **49**

Appendix E: Letters of Support **52**

Summary Science Case

0.1 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) over the next three years centres on the continuing operation of our state-of-the-art WFCAM and VISTA Science Archives, and the launch of a new archive to support the ambitious Gaia-ESO Spectroscopic survey. We will add value to these survey datasets both by integrating them with local copies of relevant external catalogues and by enabling them to be cross-queried with other data published to the Virtual Observatory and will provide archive functionality that aids effective science exploitation of the combined datasets. This will be achieved by continuing to develop world-class archive infrastructure, which underpins these major survey archives, drives the development of the global Virtual Observatory and paves the way for community exploitation of the next generation of sky surveys, from Gaia through the SKA precursors and Euclid to LSST and the SKA itself. The UK has long been a world leader in survey astronomy and maintaining WFAU as a world-class survey data centre will be crucial for ensuring that the UK community is best placed to exploit these powerful data sets as they come into being over the next decade.

0.2 WFAU ACHIEVEMENTS 2009 – 2012

This proposal seeks continued support for a programme funded previously through the Wide-Field Astronomy Unit's STFC Rolling Grant. At its last renewal, equipment and "VISTA Exploitation" grants were awarded separately, in addition to the main rolling grant, so this proposal consolidates support from three previous grants (ST/H00047X, ST/H000496 and ST/H004211) which together funded an average of 6.3 FTE of staff effort during the three year period from 1/7/2009. Sections 0.2.1– 0.2.5 summarise some of the principal achievements funded by those grants during this three year period.

0.2.1 Achievements: Facilitating UKIDSS science through WSA operations

The WFCAM Science Archive (WSA) continues to be heavily used in support of both the UK Infrared Deep Sky Survey (UKIDSS) and other PI programmes using WFCAM. More than 1000 astronomers from ESO member states are registered to access proprietary UKIDSS data through the WSA, and together they executed more than 500,000 queries on the WSA during this three-year period, obtaining result sets containing in excess of 2×10^{10} rows of data. As attested by the letters of support from the Director JAC and UKIDSS Project Scientist included in Appendix E, the WSA, with its sophisticated functionality and user-friendly interface, remains one of the main factors behind the great scientific productivity of UKIDSS, which has yielded almost 350 papers to date.

0.2.2 Achievements: Supporting VISTA Public Survey PIs and Consortia

While the basic design of the VISTA Science Archive (VSA) is derived from that of the WSA, the way that ESO manages its Public Surveys (PS) means that it has quite a different operating model. We currently support five of the six VISTA Public Surveys[1] in three ways: (i) we perform some of the survey Quality

[1] Support for the sixth VISTA Public Survey, UltraVISTA, is included in our proposed programme (see Section 2.3.1)

Control tasks and support the PS teams in undertaking the remainder; (ii) we prepare survey data products that the PSPIs are obliged to provide to ESO, and, for four of the surveys, we upload those files into the ESO Science Archive Facility (SAF) for them; and (iii) we provide the survey consortia with a science archive facility that greatly exceeds the capabilities of the ESO SAF and which provides the functionality they require to do their science. Despite its being relatively new – the first data were ingested into the VSA in Q2 2010 – the PS consortia have already executed more than 350,000 queries on the VSA, and the letters of support from the six VISTA PSPIs included in Appendix E emphasise how crucial is the VSA to the conduct of their surveys and their exploitation of the data coming from them.

0.2.3 Achievements: Preparing for Gaia exploitation

WFAU has been an important partner in the UK Gaia Data Flow System (GDFS), a national component of the Europe-wide Gaia Data Processing and Analysis Consortium, for more than five years, receiving support from PPARC, STFC and now the UK Space Agency to develop software to mitigate the mission-critical effects of radiation damage on the Gaia detectors and to undertake related core processing tasks. That work is totally separate from the WFAU science archive programme covered here, but Gaia is now coming within its scope in two regards: (i) we are developing the archive for the Gaia-ESO Spectroscopic survey (GES), which is one of two major public spectroscopic surveys launched in 2012 on the ESO Very Large Telescope (VLT) and which will provide an important dataset to enhance exploitation of the Gaia data itself; and (ii) we have an influential role in planning how the Gaia catalogues will be provided to the community. These developments make use of our science archive expertise and our detailed knowledge of the Gaia mission, and, as emphasised in the letter of support included in Appendix E from the UK PI for Gaia data processing and analysis, will help secure a leading role for the UK in the exploitation of the unique potential of Gaia.

0.2.4 Achievements: Leading Virtual Observatory developments in the UK

WFAU continues to be a leading player in the global Virtual Observatory (VO) movement, as both an early adopter of important new data access standards – WFAU was the first data centre to provide secure access to proprietary data through the International Virtual Observatory Alliance (IVOA) Table Access Protocol (TAP)[2] – and as an active player in IVOA standards discussions. We have (see Section 5.2.2) developed the first system for distributed queries across geographically distributed TAP services, which makes possible for the first time the major VO goal of distributed query processing. The use of WFAU's VO services continue to grow – with 600-700 unique users[3] accessing them per month, and more than 7.5 million queries being executed over the past six months – despite the fact that AstroGrid, the UK's national VO project, was terminated in 2009. This reflects a phase change in the wider VO world, in which developments are increasingly being driven not by dedicated VO projects, but by data centres, which are using the maturing IVOA standards suite to deliver their core services to their user communities.

0.2.5 Achievements: Publicising UKIDSS and VISTA science

An achievement of a very different kind was the international reaction to the press release that we published at the time of NAM 2012, based on the "billion star" mosaic we made from WFAU-hosted data from the UKIDSS Galactic Plane Survey and "VISTA Variables in The Via Lactea" (VVV) survey. Our zoomable image viewer[4] received more than 192,000 unique visits in the five days following the press release, while the webpage[5] providing a more detailed explanation had more than 15,000 unique pageviews, with an average duration of nearly four minutes. So, through a combination of a visually arresting image displayed using sophisticated web technologies, and several levels of description, we succeeded in our goal of providing material that caught the imagination of the general public – a related BBC News story was "Most Read" and "Most Shared" on their website on the day of its release – while also providing a fuller story for those interested in going beyond the pretty picture to the science.

[2] For a full list of VO services run by WFAU see `http://vo.roe.ac.uk/index.htm`.

[3] The limited logging in some of the software components means that we can only distinguish unique IP addresses, rather than separate user identities.

[4] `http://djer.roe.ac.uk/vsa/vvv/iipmooviewer-2.0-beta/vvvgps5.html`

[5] `http://www.ph.ed.ac.uk/news/milky-way-image-reveals-detail-billion-stars-27-03-12`

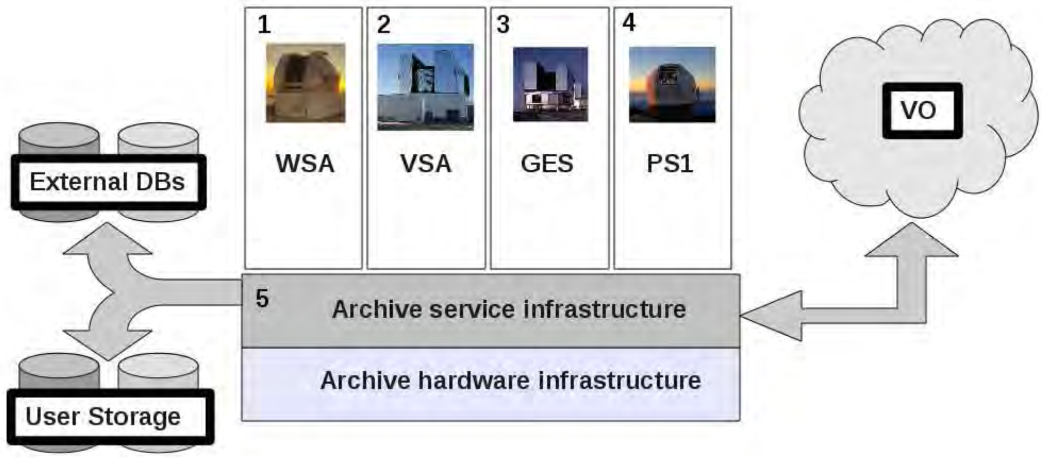

Figure 1: *A schematic representation of the proposed programme. Archives for WFCAM (Project 1), VISTA (Project 2), GES (Project 3) and PS1 (Project 4) rest on a common archive service infrastructure (Project 5), which is implemented on archive hardware infrastructure. The service infrastructure integrates local copies of external catalogues and user storage in the data centre, and provides access to the wider VO.*

0.3 PROPOSED PROGRAMME Q2 2013 – Q1 2016

Our programme for the next three years builds on these achievements, and centres on four projects facilitating exploitation of key sky survey datasets, namely:

- **Project 1: Enabling WFCAM Exploitation**
- **Project 2: Enabling VISTA Exploitation**
- **Project 3: Enabling Gaia Exploitation**
- **Project 4: Enabling Pan-STARRS1 Exploitation**

As illustrated schematically in Figure 1, these four projects are all underpinned by, and apply the results from, a fifth project

- **Project 5: Future archive infrastructure**

which develops the generic functionality required for scientific exploitation of archival sky survey data, and this, in turn, relies on a computational infrastructure which is one of a small set of **Cross-Project Resources** which provide the support necessary for the success of the five projects.

These six entities in bold font comprise the six strands of our proposed programme. The four science exploitation projects are well motivated by the scientific priorities of the UK community – we provide letters of support in Appendix E from 14 leaders of that community, covering all four of these projects – while the remaining two strands represent what we in WFAU know, from long experience of successful support of survey astronomy, is required to realise the scientific goals of our user community.

0.4 THE WFAU VISION FOR FUTURE SURVEY ARCHIVES

The programme described in the next five Sections is an ambitious combination of technology development and its application to meet externally-specified scientific goals. It is the meeting of these scientific goals that justifies the continued funding for WFAU's science archive programme, but it is through the technology development that WFAU makes its most unique contribution to UK astronomy. The work that WFAU staff undertake supporting astronomers performing survey science – as noted in Sections 1.2.2 and 2.2.2 our helpdesk system deals with ∼ 150 requests per year from WSA and VSA users – making them uniquely well placed to see how scientific goals translate into technological requirements and, as a consequence, to predict what technologies will be needed to meet the scientific objectives of the next generation of sky surveys and how astronomers will want to work together towards archieving those goals.

We already see the rise of multi-institutional (often multi-national) survey consortia with overlapping memberships, so that different people have access to different subsets of data within particular survey fields obtained by different sets of other people using a wide range of facilities. The importance of this mode of working means that it must be supported, but the flexibility and transience of its structures throws up two problems for the funders, and potential providers, of that support. Firstly, the web of consortium membership requires a correspondingly complicated set of authorization rules to enforce proprietary data rights in an environment in which someone may access one data set because of working in an ESO member state, a second by working in an institution that has bought into that particular survey and a third due to individual membership of a certain project team, and in which they need to share all those data with those collaborators who also have proprietary access to each. Secondly, the scientific value really resides in the coherent multi-wavelength data collection, not in the subset from any one survey or facility, and this causes problem for the long-term preservation of that scientific value. The *RCUK Common Principles on Data* policy[6] highlight the importance of preserving data in a form that facilitates discovery and re-use, while STFC's own *Scientific Data Policy*[7] emphasises that the choice of repository in which to store data should reflect the additional scientific value obtained from aggregration of related data. But what single institution has the expertise needed to curate the whole multi-wavelength collection? The consortium was formed to pool the expertise needed to undertake the multi-wavelength programme, but once it dissolves, that combination of expertise dissolves, too.

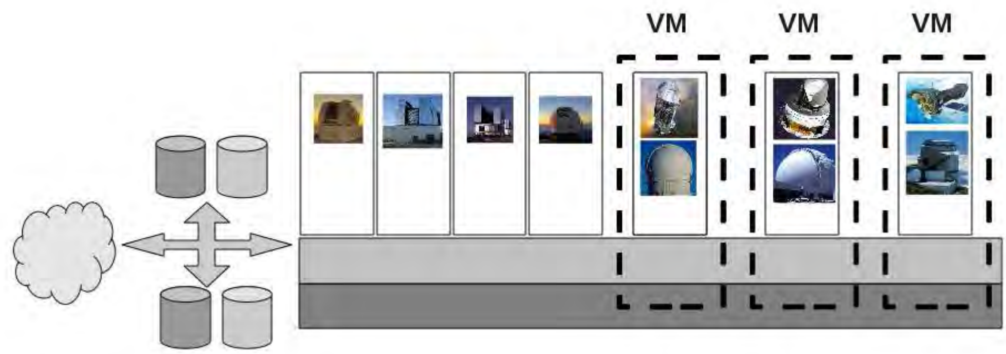

Figure 2: *A schematic representation of a future WFAU archive architecture within which a common archive infrastructure platform hosts WFAU-curated sky survery archives (shown on the left) and a set of virtualised archives, themselves hosted in Virtual Machines. These virtualised archives are operated by multi-wavelength consortia, and they benefit from the same archive service infrastructure as the core WFAU archives, have the same access to local database copies and user storage, and make use of the same access to the wider VO.*

The obvious solution to this problem is to provide an infrastructure within which multi-wavelength consortia can operate during their active lifetime, but which can be managed in such a way that the long-term preservation of their data collections can be ensured once the consortium has disbanded. One possible architecture for such a system is shown in Figure 2, and its relationship to Figure 1 is clear: WFAU is not currently in a position to support arbitrary multi-wavelength consortia in the manner that it now supports specific survey consortia, but solutions to most of the technical challenges to doing so are either in development or are certainly within grasp (see Sections 5.2 and 5.3), and our current proposal specifically targets the UltraVISTA and LOFAR survey consortia as testbeds for developing these capabilities. A strategic decision for STFC is whether it should fund WFAU (or other data centres) to undertake this work in future, or, if not, how else it can meet the objectives of its own stated policy on research data.

[6]`http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx`
[7]`http://www.stfc.ac.uk/About+STFC/37459.aspx`

Project 1: Enabling WFCAM exploitation

Staff involved

R. P. Blake
N. J. G. Cross
M. A. Read
E. T. W. Sutorius

1.1 BACKGROUND

The WFCAM Science Archive[1] (WSA, Hambly et al. 2008) has been operating successfully since early 2006. This proposed project covers the final few development tasks and continued operation of the WSA beyond ingestion of the final WFCAM data to the point where there is a final archive in a stable state that facilitates its long-term exploitation by the community. At the last review point, the panel assessing the WFAU rolling grant "...recognised the importance of the WFCAM/VISTA science archive operations to successful delivery and exploitation of the UKIRT UKIDSS programme..." and gave the WFCAM/VISTA components of the last proposal their highest priority ranking.

Ultimately, it is science exploitation that justifies investment in a system like the WFCAM Science Archive. The rate of publication of refereed papers from UKIRT[2] is at an all–time high, mainly due to the scientific output from UKIDSS[3] and this is due, to a significant extent, to the wide dissemination of the data via WFAU's tailor–made science archive system. Total citations to the refereed journal papers describing the system (Hambly et al. 2008; Cross et al. 2009) stand at 141, and archive usage (see later) remains at a very high level.

At the time of writing (July 2012), it appears that UK funding of UKIRT survey operations will cease at the end of Q3 2013. Until then, UKIRT/WFCAM will continue to produce $\sim$ 100 GB of survey data per night, mainly for the 'UKIRT Hemisphere Survey' (UHS[4]), plus smaller private survey programmes run by the CEOU[5] and KASI[6] institutes in Korea, possibly a few new PI programmes resulting from a final call for proposals from the UKIRT Board, and some final small hole–plugging observations for UKIDSS. The UHS observations will likely comprise the extension of the UKIDSS LAS/GPS/GCS J and K band coverage over as much of the accessible sky as possible. Pipeline processing/reprocessing of WFCAM data will likely continue for at least 6 months after cessation of observations, and the example of 2MASS[7] suggests that science exploitation will continue for many years provided that the science–ready data remain available online.

We request support at a constant, low level throughout the first half of the grant period for operations consisting of the day–to–day running of the science archives including transfer, ingest, curation and serving of science quality data and maintaining the hardware and software systems required for those tasks. We also seek a modest amount of final development effort to enhance the system in the light of requirements arising from a consultation with the user community at the recent 'Science from UKIDSS IV' meeting[8] held at IfA

[1] `http://surveys.roe.ac.uk/wsa/`
[2] `http://www.jach.hawaii.edu/UKIRT/news/UKIRT_Productivity_Rising.html`
[3] `http://www.ukidss.org/science/pubs/stats.gif`
[4] `http://wiki.astrogrid.org/pub/UKIDSS/SurveyCall/ukidss-uhs.ps`
[5] `http://bigbang.snu.ac.kr/`
[6] `http://www.kasi.re.kr/english/`
[7] `http://www.ipac.caltech.edu/2mass/`
[8] `http://www.roe.ac.uk/ifa/Meetings/UKIDSS12/programme.html`

Edinburgh in Q2 2012 and necessary to ensure the long-term exploitability of the final WFCAM archive.

1.2 REPORT ON RECENT WORK: 2009 − 2012

1.2.1 Achievements: data releases

Since Q1 2009, the WSA operations staff have prepared five science–ready UKIDSS data releases[9] for the community totalling ∼ 9 TB: Data Release 5 (Q2 2009; 1.7 TB); DR 6 (Q4 2009; 1.8 TB); DR 7 (Q1 2010; 2.2 TB); DR 8 (Q3 2010; 2.4 TB); and DR 9 (Q4 2011; 0.6 TB excluding the GPS). This latest Data Release incorporates database products from observations taken up to and including the end of Semester 10B: we emphasise that flat file images, quick–look jpegs and individual passband detection catalogues become available as soon as monthly pipeline processing is completed, so, at the time of writing, all data up to and including March 2012 are available through the WSA. Operations staff have also made survey–like prepared database releases for 62 private PI programmes[10] (in addition to enabling flat–file access for service observations). Static database products (UKIDSS data releases and survey–like databases for PI programmes) are made world public once the proprietary periods (18 and 12 months, respectively) are passed. The latest world accessible UKIDSS release is DR 8; with a few exceptions granted by the UKIRT Board, PI programmes up to the end of Semester 10B are accessible to the world astronomical community via the WSA. World–accessible releases are published[11] to the international Virtual Observatory via infrastructure developed by the AstroGrid project and maintained at WFAU.

1.2.2 Achievements: archive usage

More than 1050 individual users are currently registered for proprietary (18 month) UKIDSS database access in the WSA. These users are distributed over 105 distinct institutions, more than half of which are outwith the UK. Furthermore, 241 private PI programme registrations allow access by small PI–led teams to their proprietary (12 month) private datasets.

Figure 1.1 illustrates the level of archive usage over the reporting period (NB: these statistics exclude *all* Edinburgh access to avoid testing activities skewing the results, and there is some 'shot noise' caused by data releases.). 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, with clear implications for the support of VISTA exploitation, as discussed in Section 2.1.2. The top panels in Figure 1.1 show flat file numbers and volumes. The eyeballing of science frames by the Consortium survey heads during QC accounts for the majority of the JPEG library image downloads. In terms of data volume, a steady download rate of >300 GB/month is observed. The lower panels of Figure 1.1 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 data release type, shows a sustained high level of usage with >160,000 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 year on year.

A 'helpdesk' system is operated to support users of the archive. Over the reporting period, ∼250 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[12] while general SQL query solutions are included in the SQL cookbook[13] 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.

[9] `http://surveys.roe.ac.uk/wsa/releasehistory.html`
[10] `http://surveys.roe.ac.uk/wsa/nonSurveyStatus.html`
[11] `http://vo.roe.ac.uk/index.htm`
[12] `http://surveys.roe.ac.uk/wsa/qa.html`
[13] `http://surveys.roe.ac.uk/wsa/sqlcookbook.html`

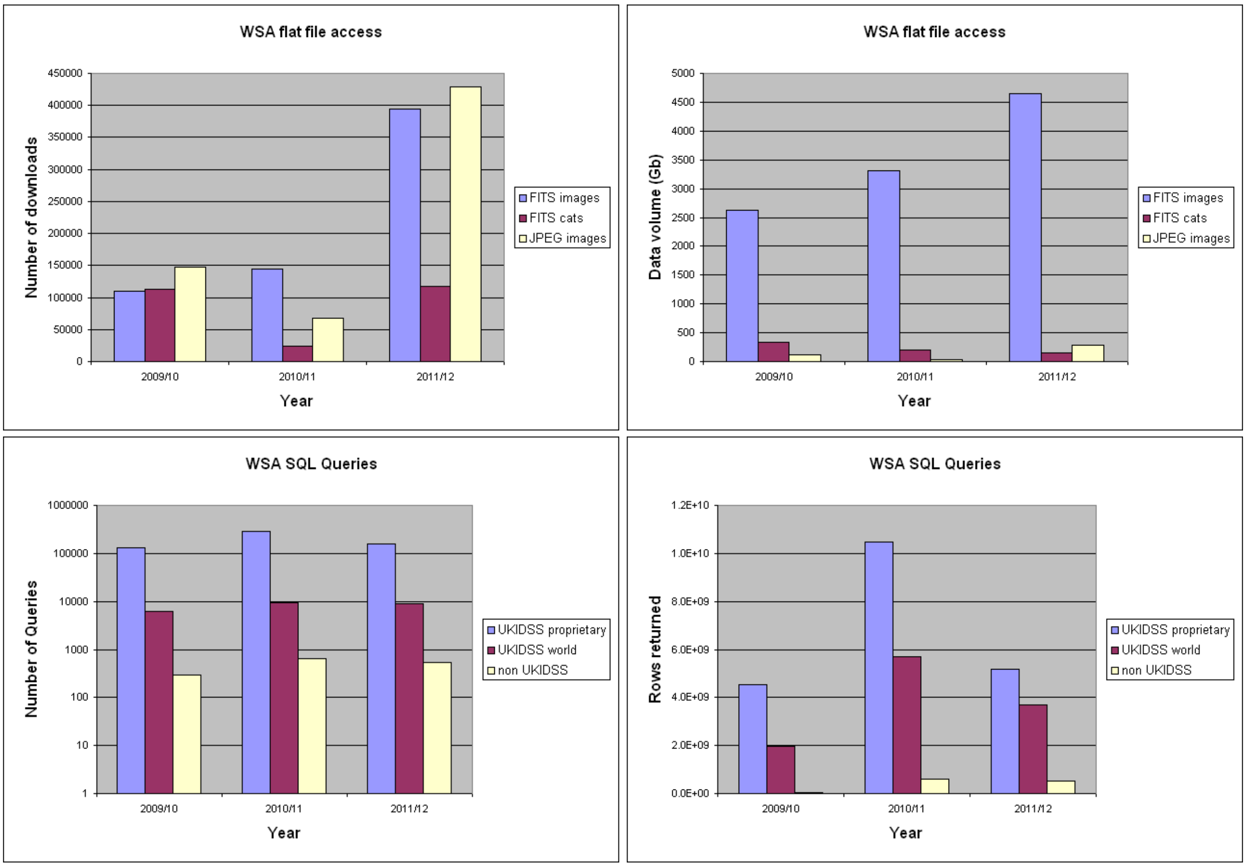

Figure 1.1: *WSA usage activity over the reporting period. Top panels: flat–file access (image/catalogue files and compressed JPEG images); bottom panels: SQL queries.* UKIDSS proprietary *refers to queries performed by registered UKIDSS/ESO users,* UKIDSS world *shows queries to world accessible releases and* non UKIDSS *records queries to databases of private PI programmes.* N.B. *these statistics exclude* all *Edinburgh access to avoid testing activities skewing the results, and there is some 'shot noise' caused by data releases.*

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; maintained and upgraded the archive software including patches and updates to third–party software; and, finally, have operated the system backup policy. Maintenance of the overall hardware systems has also included periodic upgrading of PC servers and LTO tape devices. Archive downtime is recorded[14] for the WSA – over the period Q1 2009 to Q4 2011 (1095 days) archive downtime is less than 20 days, i.e. archive availability has been $> 98\%$.

1.3 PROPOSED PROGRAMME OF WORK Q2 2013 – Q1 2016

1.3.1 Final developments

After consultation with representatives of the UKIDSS consortium in Q2 2012, a prioritised 'wish list' of survey release database and user interface enhancements was assembled as follows.

First priority enhancements

Cross-match facilities: all surveys
All surveys benefit greatly from the WSA cross-match facilities whereby complementary survey catalogues

[14] `http://surveys.roe.ac.uk/wsa/downtime.html`

are pre–joined via 'cross-neighbour tables' to enable cross-matching tuned at query time to specific science goals between, for example, the infrared and optical. This is already a significant facility – e.g. the WSA Schema Browser[15] lists 16 cross-neighbours tables for us with the the UKIDSS Large Area Survey (LAS) `source` table – and a top priority will be to continue integration relevant external catalogues in response to requests from the UKIDSS user community. Recent acquisitions of wide–angle catalogue datasets in the WSA include SDSS DR8 and the WISE all–sky data release and there will be further public SDSS and Pan-STARRS releases (see Section 4.3.1) to be incorporated during the grant period.

Stacking and source detection for the two J–band epochs: LAS
The pixel and catalogue tools required for this already exist in the curation software suite of the WSA, but there are some table schema, ingest application and user interface changes required to implement this in order to present the results to the end–users in science–ready form. One of the key science goals is of course to push surveys for rare objects (e.g. Warren et al. 2008; Mortlock et al. 2011) significantly deeper in the J band.

PSF–fitted object detection: GPS
This is covered in more detail in Section 2.3, since it is primarily motivated by VISTA requirements, but will provide significant scientific benefits when applied to the UKIDSS/GPS, too.

More rapid turn-around on eye-ball quality control: UHS
The reconfiguration of the UHS observing plan in the light of the decision to close UKIRT in Q3 2013 requires that basic QC operations are completed promptly enough to allow coverage gaps caused by QC deletions to be filled in the same observing season. This is quite demanding, given that WFAU receives data from CASU some 6-8 weeks after they are taken, so this will require some operational changes to expedite flat–file ingests and the generation of the quick–look jpegs that are crucial to the QC process. Needless to say, UHS is to benefit hugely from existing observations from UKIDSS–LAS/GPS/GCS in areal coverage, so there is also a new requirement to stitch seamlessly those catalogues (along with the associated image sets) into UHS.

Second priority enhancements

Bulk list-driven photometry: principally LAS
We have long offered a tool with which users can generate SDSS magnitudes or upper limits at the positions of catalogued UKIDSS sources and in consistent apertures, but there has been a long-standing request from the community for bulk list-driven photometry where, for example, SDSS magnitudes are published in the WSA for *all* LAS sources not detected by SDSS. This has been a nice-to-have for UKIDSS, but is a crucial requirement for VISTA, so we shall definitely address it during the grant period, as discussed in more detail in Section 2.3.

Recovery of saturated bright star photometry: GPS
Algorithms for this (Irwin 2009) have been implemented in the VSA from the outset, but needs to be applied retrospectively in the WSA where required.

Third priority enhancements

Re-processing: all surveys
A certain amount of reprocessing in the pipeline system is inevitable as enhanced features are added or as bugs are fixed. Pipeline reprocessed data are required to be retransfered, reingested and quality–controlled again, with older superseded data being deprecated. This is a considerable task from an operational point of view, and we anticipate a significant amount of reprocessing before the WFCAM pipeline finally ceases operations. Furthermore, some archive–end reprocessing (regeneration of database–driven products like deep stacks; recalibrations to produce the best relative zeropoints and hence variability information in time–sampled datasets like DXS and UDS) will be required to produce the best legacy value.

[15] `http://surveys.roe.ac.uk/wsa/www/wsa_browser.html`

1.3.2 Final data releases

At the time of writing, UKIDSS DR10 is being prepared and will appear during Q3 2012, with the GPS to be appended during Q4 2012; these releases will incorporate data up to and including Semester 11B observations. A final Data Release is required that will incorporate 12A/B and 'mop–up' observations taken during the UHS observing campaign and it has been agreed that this should take place towards the end of Q4 2013. This schedule also allows time to make the enhancements discussed above.

Currently, the requirements for UHS release(s) are still being discussed amongst the survey implementation team and the UKIRT Board. Obviously there needs to be at least one release; our baseline assumption is that a release (possibly with some proprietary restrictions) will be required after the first year of observations (Q3 2013), followed by a final, complete (as far as possible) release around one year after that.

Finally, while WFCAM/UKIRT remains in survey operations, there is a requirement to provide survey–like releases of science–ready, static database products for any smaller PI programmes undertaken (known colloquially as 'non–survey' programmes). These will consist of the Korean projects, and any new small programmes resulting from any new, final call issued by the UKIRT Board in response to reprioritisation following the announcement of the withdrawal of UK funding from JAC.

1.3.3 Legacy archive consolidation and operations

In order to preserve the digital heritage for continuing legacy value of WFCAM data, it is vital to maintain the WSA online facilities in the medium term. This requires a relatively small amount of effort (tenths of an FTE) given the continued VSA operations (Project 2) but it is important that this is explicitly stated and not overlooked. The resource level will taper over time, and is included in the overall operations resource level requested below.

The requirements for storage are estimated as follows for WFCAM data, based on experience with the WSA to date, and assuming cessation of survey observations at the end of Q3 2013:

	Pixel data	**Catalogue data**
Data volume per night (TB)	0.06	0.012
Nights	548	548
Total transfered (TB)	33	6.6
Reprocessed data (TB)	150	15
Total (TB)	183	21.6

Hardware Type	**3yr capacity total**
Pixel storage (TB)	183
Catalogue storage (TB)	21.6

1.4 KEY DELIVERABLES and MILESTONES

Milestones:

Q3 2013: UK funding for UKIRT/WFCAM survey operations ceases

Q2 2014: Pipeline processing of UK–funded survey observations ceases

Q4 2014: Final WSA legacy consolidation (residual operations only from now on)

Deliverables:

Q3 2013: UHS first year intermediate Data Release

Q4 2013: UKIDSS Final Data Release

Q3 2014: UHS second and final Data Release

1.5 RESOURCES REQUESTED

For final developments, we request staff resources at the level of 0.5 FTE from Q2 2013 to Q1 2014 (split equally between Cross and Read) tapering to 0.25 FTE (Read only) in Q2 2014 to Q3 2014 to cover the requirements. Development effort is then zero from Q4 2014.

The operational overhead on all of the above benefits greatly from an economy of scale with the corresponding VSA operations (see Project 2). We request staff resources at the level of 0.75 FTE (split equally between Read, Blake, and Sutorius) to cover this activity from Q2 2013 to Q3 2014. Operations effort is then zero from Q4 2014.

Residual system administration effort will be required to keep the WSA live and usable after the end of its operational phase, but this is modest, and is implicitly included in our SysAdmin request in Section 6.3.

Guideline T&S and personal computing is requested for development, operations and attendance of appropriate meetings (NAM, ADASS series, survey science workshops) to interact with the user community.

In Summary:

(a) *Staff.* 1.25 FTE from Q2 2013 to Q1 2014 falling to 1.0 in Q2 and Q3 2014 and zero thereafter
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE.
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE.
(d) *Equipment & maintenance:* please see Appendix D

N.B. should funds become available to extend the life of UKIRT beyond the end of September 2013, the annual cost of extending WSA operations would be approximately £130k per annum: $\sim$ £30k for (pixel+database) storage and $\sim$ £100k for operations staff effort.

1.6 ADDITIONAL REFERENCES

Cross N.J.G. et al., 2009, *Archiving multi-epoch data and the discovery of variables in the near–infrared*, MNRAS, 399, 1730

Hambly N.C. et al., 2008, *The WFCAM Science Archive*, MNRAS, 384, 637

Mortlock D. et al., 2011, *A luminous quasar at $z = 7.085$*, Nature, 474, 616

Irwin M.J., 2009, *Extending the dynamic range of WFCAM data*, In: UKIRT Newsletter, Autumn 2009, No. 25, p. 15

Warren S.J. et al., 2007, *A very cool brown dwarf in UKIDSS DR1*, MNRAS, 381, 1400

Project 2: Enabling VISTA exploitation

Staff involved

R. P. Blake
R. S. Collins
N. J. G. Cross
M. A. Read
E. T. W. Sutorius
S. Voutsinas

2.1 BACKGROUND

2.1.1 The VISTA Data Flow System

The VISTA Science Archive (VSA[1]) is the product of the second phase of the VISTA Data Flow System project (VDFS[2]; see Emerson et al. 2004; Irwin et al. 2004; Hambly et al. 2004). The VDFS provides a systems engineered end–to–end data flow, incorporating 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: first, the WFCAM Science Archive (WSA; see Project 1 previously) and, now, the VISTA Science Archive (VSA). At an average level of around 2.5 FTE for the past ten years, the resulting science archive systems represents the culmination of ~ 25 staff years of effort. The systems are routinely used for the archiving and dissemination of UKIDSS survey data (see Project 1) and for five of the six VISTA public surveys[3]: 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. Data from the sixth public survey, UltraVISTA, has not been included in the VSA so far, but this is set to change: see Section 2.3.1.

2.1.2 The VSA and the ESO SAF

The initial VDFS plans for the VSA were perturbed by the inclusion of VISTA in the in-kind contribution to the package effecting UK's accession to ESO. While the science archive system developed for the WSA has formed the basis of the VSA, there are many differences between their operating models, resulting from the way that ESO has decided to operate its public surveys. This centres on the uploading by the Public Survey PIs (PSPIs) of data products into the ESO Science Archive Facility (SAF[4]). This is not a science archive in the WSA/VSA sense, designed to support survey science, but simply a repository for flat-file data products, much in the way ESO archives data products resulting from small PI programmes on its other telescopes.

Our analysis of the community use of the WSA, presented in Section 1.2.2, argues strongly that such a repository is not suitable for supporting survey science: a community that has spent more than a decade doing their science on the SDSS SkyServer and the WSA expects to have the flexible querying power of an SQL interface to a high-performance database that allows them to extract small, targeted subsets of a large

[1] `http://surveys.roe.ac.uk/vsa`
[2] `http://www.maths.qmul.ac.uk/~jpe/vdfs/`
[3] `http://www.eso.org/sci/observing/PublicSurveys/sciencePublicSurveys.html#VISTA`
[4] `http://www.eso.org/sci/archive.html`

survey catalogue, and not simply a webform through which they have to download whole catalogues. It comes as no surprise, therefore, to read[5] several national representatives at the most recent ESO User Committee meeting noting that the SAF is little used for VISTA data, and reporting that colleagues are *"expecting* [the] *possibility for massive queries"*, are requesting that the archive *"provide custom sized cutouts and links to other surveys"* and are making *"suggestions to improve the user interface in order to allow queries on VISTA catalogues"* - i.e. to provide the basic access modes that have been in the VSA design since its inception, and have been on offer in its precursor, the WSA, for more than five years.

Another fundamental difference between the VSA and the SAF is that static data is simply deposited into the latter, while the VSA provides active support to the public survey consortia in both their quality control and science exploitation activities. Queries on the VSA can (to a large extent, but not completely) generate the data products that the PSPIs are required to upload into the SAF to ensure continuation of their surveys. This helps to minimise the duplication of effort that results from data from each VISTA public survey being prepared for publication through two very different archive systems, but duplication remains - it was suggested to ESO during the discussion of possible in-kind contributions to meet the "VISTA penalties" that the VSA could become the official ESO archive for VISTA data, in place of the very basic SAF, but ESO declined that offer. The letters of support from the six VISTA PSPIs and the Director of the VISTA Consortium included in Appendix E attest to the crucial importance of the VSA for the scientific success of the VISTA surveys, and that provides the ultimate justification for our request for support of this project.

2.2 REPORT ON RECENT WORK: 2009 – 2012

2.2.1 Achievements: data releases

Although based originally on UKIDSS/WSA, the mode of operations in the VSA has had to adapt to different requirements coming from the survey PIs. Many more frequent proprietary releases are made to the various survey consortia, which comprise reasonably large groups of European scientists interested in exploiting the data. Beginning in Q2 2010, the VSA operations staff have prepared 28 science–quality data releases[6] for the various survey consortia totalling ~ 30 TB. This has included deep stacks produced by the VIDEO team in the same way that the WSA ingested stacked images generated by the UKIDSS/UDS team.

World–public releases to the wider community beyond the survey consortia have received less emphasis so far than has been the case for UKIDSS, in part due to the distraction of delivering data to the ESO–SAF. Despite this, in Q1 and Q2 2012 world-public first data releases (DR1s) for the five surveys were made totalling ~ 10 TB. These science–ready, queryable survey release databases were made available through WFAU's initiative in pushing this important aspect of the public survey service remit. New features of the VSA relating to these releases are described at length in Cross et al. (2012).

2.2.2 Achievements: archive usage

Like the WSA, users primarily access the VSA in two modes: flat-file access and flexible and powerful SQL queries. Around 50,000 flat-file images and catalogues are being downloaded each year. Some 350,000 SQL queries have been executed during the reporting period, with nearly 6 billion rows being returned already in 2012. For reference, the figure reported[7] for the ESO-SAF is that 2007 VISTA files were downloaded between December 2011 and late April 2012, with the bulk by number being VHS catalogue files (of which $\sim 32,000$ were included in the VHS DR1), and the bulk ($> 80\%$) by volume being UltraVISTA images (which are ~ 4GB each).

In common with the WSA, a 'helpdesk' system is operated to support users of the archive. Over the reporting period, $\sim$175 requests for support have been received. A response is normally made within a few hours of receipt during normal working hours. Also, administration of the IT infrastructure on which the VSA is built is a critical aspect of operations. The same activities that are noted for the WSA are undertaken for the VSA; there is of course an economy of scale in that it is the same operations team for both archives.

[5] `http://www.alma.ac.uk/other/UC36_summary_v2.pdf`
[6] `http://surveys.roe.ac.uk/vsa/releasehistory.html`
[7] `www.eso.org/public/about-eso/committees/uc/uc-36th/UC36_SpecialTopic_PublicSurveys.pdf`

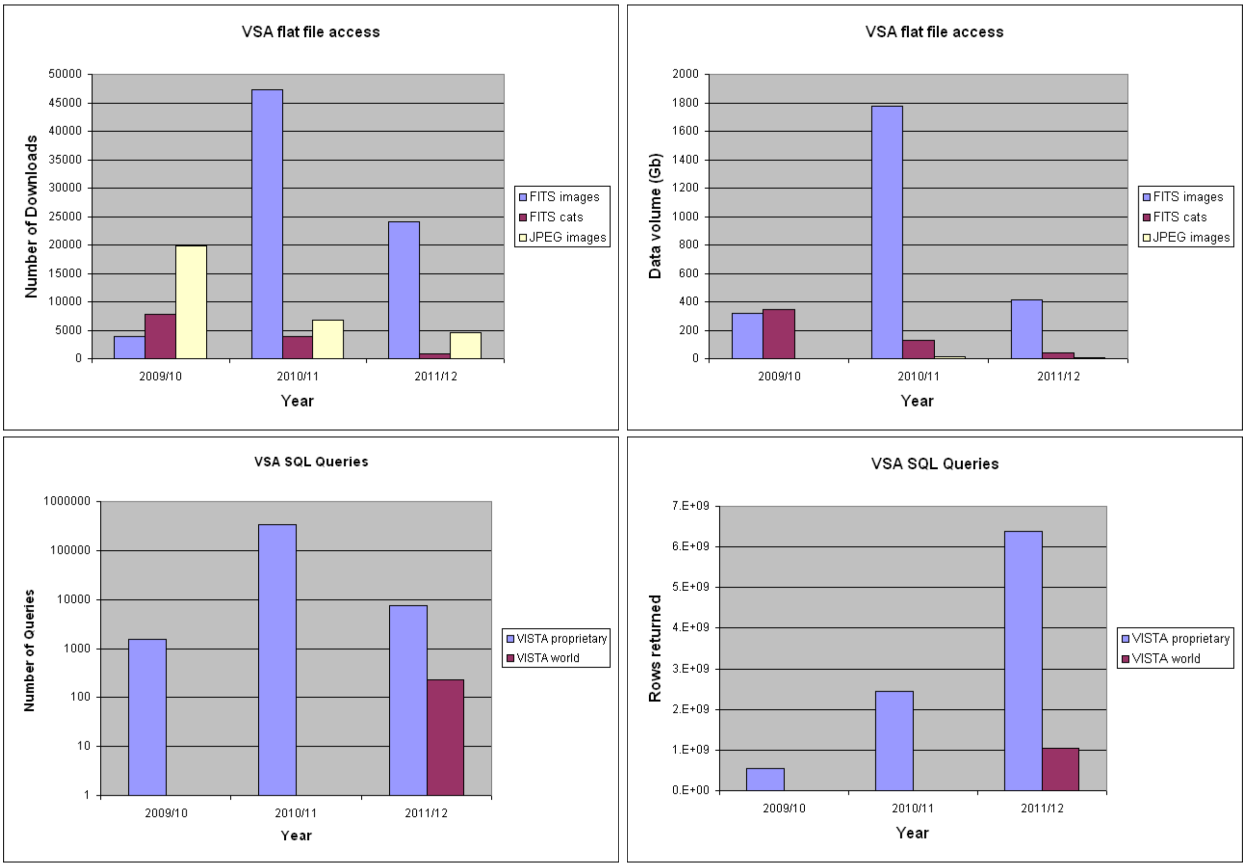

Figure 2.1: *VSA usage activity over the reporting period. Top panels: flat–file access (image/catalogue files and compressed JPEG images); bottom panels: SQL queries.* VISTA proprietary *refers to queries performed by registered VISTA consortia users,* VISTA world *shows queries to world accessible VISTA releases.*

Archive downtime is recorded[8] for the VSA – over the period Q1 2010 to Q4 2011 (730 days), downtime has been ~ 14 days, i.e. archive availability has been $> 98\%$.

2.2.3 Achievements: complementary optical surveys

Many of the usages described in the science cases for the VISTA public infrared surveys require complementary optical imaging (conversely, many usages of optical surveys require complementary infrared imaging). 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]: as for the WSA, cross-neighbour tables are automatically generated between the large area VISTA surveys and about a dozen external catalogues, currently.

The most substantial complementary optical surveys are those coming from the Public Surveys being conducted in the VLT Survey Telescope (VST). The Survey Management Plans (SMPs) for the VDFS-curated VISTA surveys included requirements for the integration of two out of the three of these, namely ATLAS[10] and KIDS[11]. WFAU was initially interested in these surveys solely from the point of view of meeting these cross-matching requirements, but we were requested by the ATLAS PI, Tom Shanks, to host an archive to meet the needs of the ATLAS consortium and, later, the wider community. This we are able to do at modest cost, given the existence of the basic VDFS science archive infrastructure and the resultant Omegacam Science Archive (OSA[12]) is taking shape, as the first VST data products emerge from the VDFS pipeline in

[8] `http://surveys.roe.ac.uk/vsa/downtime.html`
[9] `http://surveys.roe.ac.uk/vsa/sqlcookbook.html#crossmatch`
[10] `http://astro.dur.ac.uk/Cosmology/vstatlas/`
[11] `http://www.astro-wise.org/projects/KIDS/`
[12] `http://surveys.roe.ac.uk/osa/`

Cambridge.

2.2.4 Achievements: delivery of data to the ESO-SAF

VSA operations staff have helped to ensure continuation of the VMC, VVV, VIKING and VIDEO survey programmes by delivering to ESO–SAF database–driven stacked image products in Q2 2011 and corresponding deep catalogues and passband merged catalogues in Q2 2012. Access to the data is presently limited to simple, static web forms at ESO.

2.3 PROPOSED PROGRAMME OF WORK Q2 2013 – Q1 2016

2.3.1 Development

The insistence on the part of ESO that PIs deliver basic flat–file products to the ESO–SAF has ended up absorbing significantly more staff resource than was envisaged in our previous grant proposal, but it remains a high priority item for continuation of the surveys (the second highest priority item in the last grant renewal concerned the VDFS interface with the ESO–SAF). The phased approach of ESO in managing this process, and their tendency to issue detailed requirements only very close to the phased submission deadlines means that a small amount of effort must be reserved to cover this. We estimate that this task requires an average of 0.2 FTE during the period Q2 2013 to Q1 2016, starting higher than this and tapering to zero at the end.

The effort already expended since Q1 2010 in servicing this requirement and the changes required in the VSA due to the differences in survey management with respect to the WSA have resulted in some developments described in our previous grant remaining to be done. On the basis of interaction with survey PIs, the following development requirements have been identified:

List–driven photometry

Consistent photometry across an image set is important for the fitting of galaxy SEDs to yield photometric redshifts and stellar mass estimates. An on–demand list–driven photometry service has been implemented in the WSA[13] using a generalised list–driven photometry tool provided by CASU (e.g. Irwin et al. 2007), which provides consistent photometry at fixed positions using fixed circular apertures. However, for VIKING and VHS at least, there is a requirement to produce wholesale list–driven photometry between the infrared bands, and between the infrared and optical (VST) images, using a more sophisticated seeing–matched approach (e.g. Hildebrandt et al. 2012). The method is to model the PSF as a sum of Gauss–Hermite polynomials across the whole image, using the mathematics of shapelets, and use this to construct a convolution kernel which when convolved with the image will created an image with a Gaussian PSF of a fixed size all across the image. This is done in all filters at the same time, to create and image set with the same fixed width Gaussian PSF (with the width of the widest PSF anywhere in any of the images) on all images (in a pointing). Source extraction using the VDFS extractor developed by CASU, or SExtractor[14], may then be performed to produce catalogues with consistent multi–band photometry.

We estimate this work will require several staff months to adapt and integrate the existing codes. Tasks include defining new curation processes to perform housekeeping and enhancing the archive data model to keep track of the required metadata, for example to provide the Gaussianised images and the coefficients of the PSF and the convolution kernel to the end–user should they wish to perform further analysis (e.g. bulge/disk decomposition) on the images. We note that this work will also benefit the UKIDSS–LAS at least in the WSA (see Section 1.3.1).

PSF–fitted object detection

Highly crowded regions in the Galactic centre regions surveyed for the VVV present a significant problem for the VDFS standard source extraction software (Saito et al. 2012). We believe this problem can only

[13] `http://surveys.roe.ac.uk:8080/wsa/betaDataAnalysis.jsp`
[14] `http://www.astromatic.net/software/sextractor`

be fully resolved by application of a DAOPHOT–like solution in the affected areas, employing multiple, simultaneously PSF profile–fitting in an iterative detection/photometry process (Stetson 1987). The degree to which it is possible to automate such a process within a pipeline, and the feasibility of applying such a CPU–intensive process to 100s of GBs of image data, in extremely crowded regions of sky (source densities of > 1 million stars per square degree), requires some investigation by VDFS in collaboration with the VVV survey team, for whom this is a high priority (as expressed in the letter of support from VVV Co-PI Phil Lucas in Appendix E). We note that this work will also benefit the UKIDSS–GPS in the WSA (see Secion 1.3.1).

Scale–out to 10s of billions of rows

The recent world–public VVV release DB contains tables with up to ~ 10 billion rows; at the end of the survey programme, we anticipate having to deal with row sets of up to 10^{11} rows. Our use of a conventional relational database management system (RDBMS) as the backing store for both survey curation and dissemination is struggling to cope with these data volumes, and we propose to develop and deploy a new archive system based on the column–oriented RDBMS MonetDB[15] to solve these problems. This provides the same rich SQL interface for both curation applications and end–user querying, but optimises IO performance by storing data in column sets. This can bring huge speed gains in bulk data operations and trawl queries for databases containing tables with many attributes when the modifications and/or query predicates involve just a few attributes — a situation that is very common in the VSA. We note this work has synergies with the R&D programme described in Project 3 for contributions to the development of the Gaia end–user archive system.

Added value to imaging datasets

The small size (~ 2 sq. deg.) of the UltraVISTA field might make it seem the VISTA survey least likely to benefit from the WFAU science archive systems, which are largely tailored to supporting wide-angle surveys. The ESO SAF file download figures quoted in Section 2.2.2 above show that most UltraVISTA exploitation is image-based, rather than based on default catalogue data products, but the limitation of the SAF for supporting this research is that users have to download full UltraVISTA images, whereas, in most cases, they will be happy with postage stamp images of the immediate surrounds of objects of interest.

We propose to support that mode of access to UltraVISTA and to study with the UltraVISTA Co-PI, Jim Dunlop, how to add value to the UltraVISTA dataset through integration of complementary imaging datasets from other wavebands. Dunlop is currently investigating how to do this with Subaru, CFHT and Spitzer data, to define consistent, deblended multi–wavelength source lists from the optical through to the mid–infrared. This is a specialised R&D activity outwith the scope of the standard VDFS processing, since there is a significant angular resolution mismatch between the ground–based optical, near–infrared and space–based mid–infrared images. Cross-matching individually generated catalogue lists is simply not good enough to extract the science, which includes robustly measured optical to mid–infrared SEDs and photometric redshifts. Dunlop has EU FP7-funded resources to derive the optimal methods for doing this[16] but not to publish the outputs, so WFAU can make these valuable new data products available to the community through inclusion in the VSA.

2.3.2 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 these posts are 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.

[15] `http://www.monetdb.org/Home`

[16] One method would be PSF Gaussianisation described previously, but, while this may be optimal for the analyses to be conducted on the wider surveys, it may not be the best method for specific analyses to be run on UltraVISTA.

General archive curation

General archive curation 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 larger timescale tasks, e.g. preparing release database products for publishing online and managing VISTA data products egress to the central ESO repository. 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 is set using a manually instigated data modification script for a host of survey data quality issues. Our experience of working with the UKIDSS consortium and VISTA public 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 specialised and experienced staff to work on local QC implementation in order to maintain survey data quality in released database products: our last WFAU rolling grant proposal identified a Survey Quality Assurance Scientist role and Mike Read continues to fulfil that role.

User support

Archive operations requires provision of user support. The access point to WFAU archives is 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.

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 RAID6 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 all the archive hardware and network infrastructure; and finally managing downtime, e.g. liaising with users to minimise the impact on of scheduled maintenance of the archive system.

The requirements for storage expansion are estimated as follows for VISTA data:

	Pixel data	**Catalogue data**
Data volume per night (TB)	0.2	0.04
Nights	1095	1095
Total transfered (TB)	219	44
Reprocessed data (TB)	200	50
Total (TB)	419	94

Hardware Type	**3yr capacity total**
Pixel storage (TB)	419
Catalogue storage (TB)	94
Processing (CPUs)	64

and for VST–ATLAS/KIDS:

	Pixel data	**Catalogue data**
Data volume per night (TB)	0.05	0.02
Nights	1095	1095
Total transfered (TB)	55	22
Reprocessed data (TB)	165	44

Hardware Type	**3yr capacity total**
Pixel storage (TB)	165
Catalogue storage (TB)	44

2.4 KEY DELIVERABLES and MILESTONES

Deliverables (subject to shifting requirements from public survey PIs and ESO survey administration particularly with regard to delivery of products to the ESO–SAF):

Q2 2013: 4th delivery of flat-file survey products to ESO–SAF

Q3 2013: Consortium survey DB releases for all VISTA and VST surveys

Q4 2013: 5th delivery of flat-file survey products to ESO–SAF

Q1 2014: Consortium survey DB releases (excluding VVV and VST–ATLAS)

Q2 2014: 6th delivery of flat-file survey products to ESO–SAF

Q3 2014: Consortium survey DB releases for all VISTA and VST surveys

Q4 2014: 7th delivery of flat-file survey products to ESO–SAF

Q1 2015: Consortium survey DB releases (excluding VVV and VST–ATLAS)

Q2 2015: 8th delivery of flat-file survey products to ESO–SAF

Q3 2015: Consortium survey DB releases for all VISTA and VST surveys

Q4 2015: 9th delivery of flat-file survey products to ESO–SAF

Q1 2016: Consortium survey DB releases (excluding VVV and VST–ATLAS)

World–public releases of VISTA and VST–ATLAS survey databases follows around one year after the respective survey consortium releases. Note that survey consortium DB releases for all VISTA surveys except the VVV take place at 6–monthly intervals while VVV and VST–ATLAS releases take place at 12–monthly intervals.

2.5 RESOURCES REQUESTED

For developments, we request staff resources at the level of 0.5 FTE from Q2 2013 to Q1 2015 (Voutsinas) tapering to 0.25 FTE from Q2 2015 to Q1 2016 (Collins) to cover the requirements.

For operations, we request a unifrom 2.25 FTE over all 3 years for Blake (0.75 FTE from Q2 2013 rising to 1.0 FTE at Q4 2014), Collins (0.25 FTE from Q2 2013 to Q3 2014), Read (0.5 FTE from Q2 2013 rising to 0.75 FTE at Q4 2014), Voutsinas (0.25 FTE from Q2 2013 to Q3 2014), and Sutorius (0.5 FTE throughout).

Travel and subsistence is requested for operators to enable them to attend conferences such as NAM, JENAM and ADASS[17] for the purposes of providing 'roadshow' demonstration of archive functionality and for face–to–face interaction with the user community. Personal computer equipment is a necessity for both development and operations and is requested at guideline rates.

In Summary:

(a) *Staff.* 2.75 FTE from Q2 2013 reducing to 2.5 FTE at Q2 2015
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE
(d) *Equipment & maintenance:* please see Appendix D

2.6 ADDITIONAL REFERENCES

Cross N.J.G. et al., 2012, A&A, in press

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

Hambly N.C. et al., 2004, In: Optimizing Scientific Return for Astronomy through Information Technologies, Proceedings of the SPIE, 5493, 423

Hildebrandt H. et al., 2012, MNRAS, 412, 2355

Irwin J. et al., 2007, MNRAS, 375, 1449

Irwin M.J. et al., 2004, In: Optimizing Scientific Return for Astronomy through Information Technologies, Proceedings of the SPIE, 5493, 411

Saito R.K. et al., 2012, A&A, 537, 107

Stetson P.B., 1987, PASP, 99, 191

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

Project 3: Enabling Gaia Exploitation

Staff involved

N. C. Hambly
R. S. Collins
N. J. G. Cross
M. A. Read
E. T. W. Sutorius
S. Voutsinas

3.1 BACKGROUND

3.1.1 The Gaia Mission

Gaia is an ESA Cornerstone mission, due for launch in Q3 2013 and with a planned lifetime of five to six years. It will conduct whole–sky astrometric, photometric and spectroscopic surveys, visiting each field an average of ~ 70 times, and generating a combined dataset comprising $\sim 10^{12}$ observations. Gaia's principal goal is the generation of a high precision 3D map of $\sim 10^9$ 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 trigonometric distances to Cepheids and RR Lyrae stars, and fixing the cosmological reference frame through precise astrometry of distant quasars.

3.1.2 Preparing for exploitation of Gaia data

Gaia's data products fall into three broad categories with respect to the mission timescale. 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 and ultimately, 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. What is now emerging (O'Mullane & van Leeuwen, 2012) is the nature and schedule for a third category of intermediate data products, to be released 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 about 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 obviously these will be of great interest to the scientific community.

As part of the PPARC/STFC-funded Gaia Data Flow System (GDFS) project 2007–2012, and latterly the UK Space Agency Gaia Post Launch Support grants 2012–2015, WFAU is playing a key role in developing core processing algorithms and software for the real–time and offline astrometric and photometric processing pipelines for Gaia. This work is undertaken within Coordination Unit 5 (CU5) of the Gaia Data Processing and Analysis Consortium (DPAC). Within DPAC, public access to all Gaia catalogues will be the responsibility of CU9, which has yet to be activated. The original 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 that perfectly fit WFAU's Science Archive philosophy and working practices: 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[1], who confirmed that CU9 will be the subject of separate Announcement of Opportunity, to be issued by ESA. The AO for formation of CU9 is expected before the end of 2012, and in the meantime the DPAC Executive instigated the Gaia Archive Preparation (GAP) group, a precusor to CU9. As described in the letter of support from Gerry Gilmore, the UK PI, Gaia data processing and analysis activity, included in Appendix E, work on CU9 is Gaia exploitation, and so support for it is to be sought from STFC and not from UKSA. He also notes that WFAU is expected to take a leading European-scale role as the CU9 team develops, given its existing expertise in data release and science archive activities.

3.1.3 Obtaining ancillary data to enhance Gaia exploitation

The importance of the Gaia mission to European astronomy is underlined by the fact that following a call for large public survey programmes in 2010, ESO awarded some 300 nights of VLT time to a ground–based high–resolution spectroscopic observing campaign to complement the space–based observations (Gilmore et al. 2012). This unprecedented spectroscopic survey will augment Gaia's on–board low–resolution spectrophotometry and high–resolution Ca II triplet spectra for $V < 17$ with a subsample of 10^5 stellar spectra from FLAMES on UT2 (using both GIRAFFE and UVES) down to $V \sim 19$.

3.2 REPORT ON RECENT WORK: 2009 – 2012

The Gaia group within WFAU has established a key role in providing Gaia core processing algorithms, including fundamental work[2] on critical calibration issues like the radiation damage problem, background estimation and handling of the quasi–stable electronic bias in the CCD readout amplifiers. The group is now firmly integrated into the CU3 core processing development teams, and has gained considerable knowledge of Gaia's highly complex instruments and mission. We list here our additional achievements in relation to GAP/CU9 and Gaia exploitation.

3.2.1 Achievements: GES end–user archive prototype

As a result of the success of the VDFS Science Archive concept (see Projects 1 and 2 previously), the Gaia–ESO consortium requested the services of WFAU in creating an end–user archive system for the survey: see letter of support from GES Co-PI Gerry Gilmore in Appendix E. Starting in Q4 2011, WFAU followed a VDFS–like design process and has prototyped a GES science archive[3] (e.g. see Figure 3.1) that will be used to disseminate survey data products for consortium science exploitation, and ultimately for the world–public research community. WFAU VDFS developers and operations staff have adapted VDFS software and procedures to archive early pipeline processed observations (which started at the end of Q4 2011).

3.2.2 Achievements: WFAU co-opted onto the Gaia Archive Preparation group

In Q2 2011, WFAU contributed to the coordinated UK-wide response to the ESA call[4] for Letters of Interest in participating in GAP activities, which led to Nigel Hambly being invited to join GAP. He quickly proposed that a requirements gathering exercise based on 'usage scenarios' should be followed (again, following the VDFS–style approach) in order to kick–start the developments. This was done via a series of meetings and a community–wide call for inputs, and has resulted in the generation of the key document, 'Gaia data access scenarios summary' (Brown et al., 2012).

[1] T. Prusti, priv. comm, March 2007
[2] see for example http://www.rssd.esa.int/index.php?project=GAIA&page=picture_of_the_week&pow=153
[3] http://surveys.roe.ac.uk/ges/
[4] http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=48463

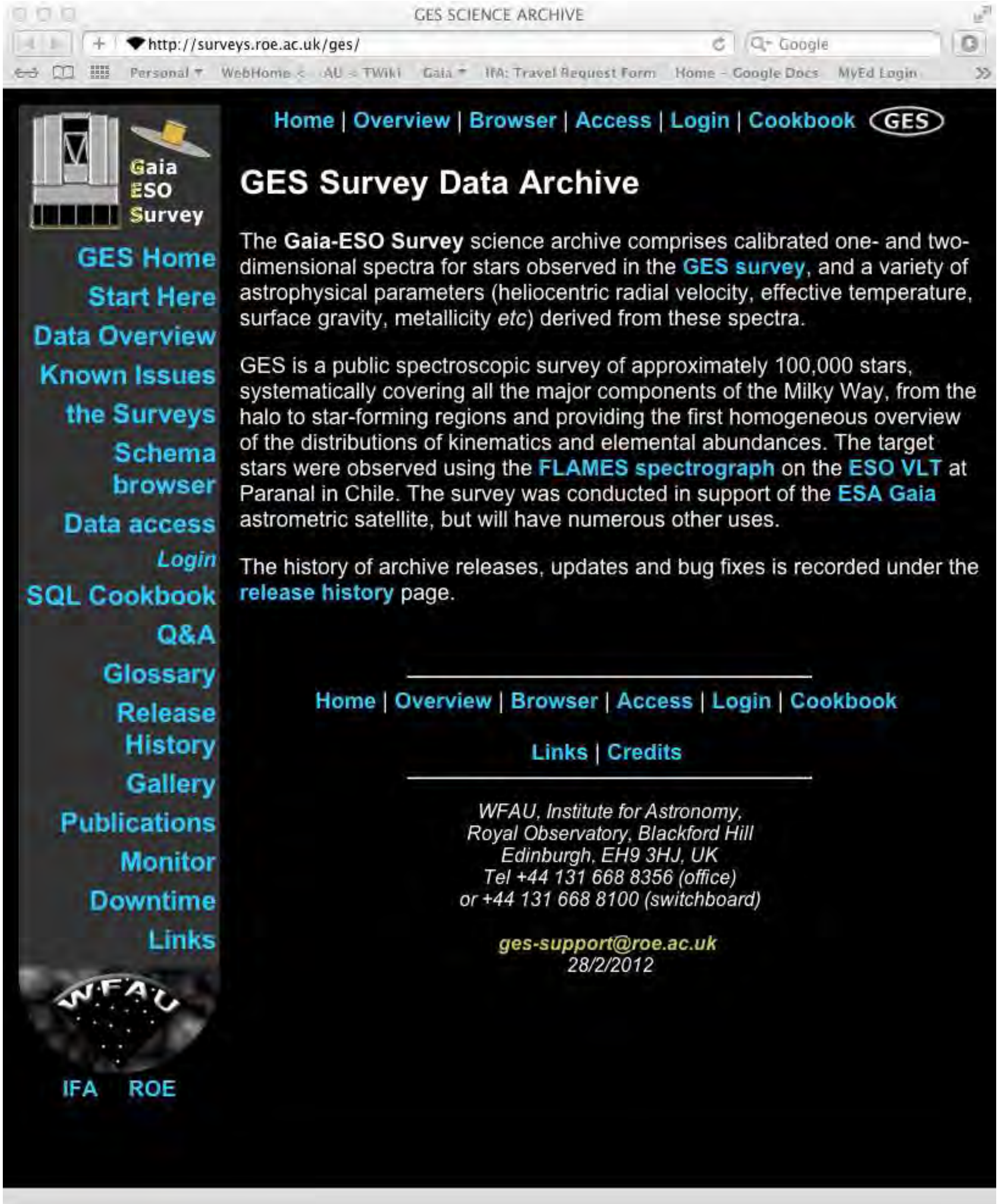

Figure 3.1: *Homepage of WFAU's science archive prototype for the Gaia–ESO Survey*

3.2.3 Achievements: preliminary archive development programme

In collaboration with European partners in Spain, Holland and France, WFAU developed the Gaia European Network for Improved User Services proposal (a.k.a. GENIUS[5]) in response to the European Commission FP7-2012-Space call. Although very highly rated that proposal was unfortunately not funded. However the workplan formulated therein is being used by GAP as the basis for planning a coordinated response to the ESA AO for CU9 participation. Hence WFAU's interests and expertise have been already knitted into a cohesive plan for CU9 amongst interested research institutions within Europe.

3.3 PROPOSED PROGRAMME OF WORK Q2 2013 – Q1 2016

We propose a programme of work to service Gaia exploitation in the early phases of the mission by (i) final development of the Gaia–ESO Survey archive system; (ii) GES archive operations; and (iii) contributions to DPAC CU9. *N.B.* This work is independent of, but benefits greatly from, our separately–funded work on core data processing for DPAC under the auspices of CU5.

3.3.1 Final developments for the Gaia–ESO Survey science archive

In contrast to the extant VDFS archives, GES curation requires the handling of a more richly structured dataset in that it consists of i) instrumentally–corrected 2d spectrograph image science frames (and the corresponding 2d calibration frames used in the correction process); ii) 'multi–spectrum' 2d frames, where rows correspond to different spectra and columns correspond to dispersion elements; iii) tabular datasets of measurements (e.g. equivalent widths) from spectral features; iv) derived information like atmospheric parameters and element abundancies; and v) ancilliary data like survey input catalogues, line lists etc. This requires careful organisation in the archive to facilitate diverse end–user science usage, and some requirements only become clear after several iterations with the user community. Furthermore, the interface applications (from simple webforms and browsing visualisation through to more sophisticated facilities based on freeform SQL accesss and VO–enabled interfaces) need to be developed and enhanced based on end–user feedback.

Prototype facilities have been developed and deployed: we now request resources for Voutsinas and Collins to complete the curation and interface applications for the GES science archive. We propose also to deploy an archive to support the UK component of the HARPS-N[6] consortium. They are obtaining radial velocities for planets from 2d spectroscopic data taken on by the HARPS-N instrument on the Telescopio Nazionale Galileo (TNG[7]), and their requirements for data archiving and publication are a subset of what is already being provided for GES. So, given the modest data volumes involved ($\sim$ 10TB over five years), we are able to provide at minimal cost to this grant an archive that has high value to their consortium: see the letter of support from Andrew Collier Cameron, the UK PI of the HARPS-N consortium in Appendix E.

3.3.2 GES science archive operations

Again following the VDFS model, a relatively small amount of staff effort is required for day–to–day operations for the GES science archive. This task benefits greatly from the experience in operating the VDFS archives and will be done by the same staff. Tasks include transfering the data from the pipeline processing operational repository (at the Cambridge Astronomical Survey Unit), ingesting the data and metadata and ensuring their consistency, making periodic releases to the survey consortium, and preparing static release products and their documentation for the world astronomical community.

3.3.3 Participation in Gaia DPAC CU9

(*N.B.: data access facilities are to be physically hosted at the European Space Astronomy Centre near Madrid, so this proposal is for staff resources to enable UK contributions to the developments; we make no bid for*

[5] `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/`
[6] `https://plone2.unige.ch/HARPS-N`
[7] `http://www.tng.iac.es/`

hardware resources. Furthermore, this plan of work has been coordinated with the relevant European partners as part of the unfunded GENIUS FP7 bid, and as such represents a viable, integrated contribution from the UK).

Databases and query engines will be at the core of the Gaia archive system. We propose a plan of work devoted to the implementation of demonstration elements for the Gaia archive system, in part based on specific application of the generic technologies being investigated and developed under Project 5. Moreover, the technology choices and the design of the systems will be carefully based on the real user needs, and demonstration system(s) will support the advanced tools and activities defined by GAP/CU9.

Database systems evaluation

The current generation of terabyte–scale, billion–row survey catalogues are served to the community using conventional relational technology (e.g. Hambly et al. 2008 and references therein). This brings many advantages for applications design, and, more importantly, users have become accustomed to the power provided by Structured Query Language (SQL) access to the full relational schema for a complex database, rather than being restricted to certain access patterns and query types implemented as webforms.

The success of these systems suggests that a similar conventional relational database management system (RDBMS) should be considered as the baseline for Gaia, but several newer developments in database technology may be particularly relevant to Gaia and should be studied, too. Firstly, in recent years there has been a growing realisation that many scientific databases – which typically feature wide tables, within which a small number of columns are very popular – are well suited to column-oriented databases, which minimise the data I/O needed for queries on small numbers of columns, which are typical in scientific database workloads. The leading open–source column–oriented database is MonetDB[8], and initial tests[9] have shown that multi–TB sky survey data sets do indeed perform very well when implemented in MonetDB. Secondly, the Gaia project at ESAC is using a high–performance object–oriented database, Intersystems Caché[10], for the persistence of Java objects in the core processing systems and, since it also supports an SQL interface, it would be advisable to see whether this same system can support the user scenarios developed through GAP/CU9 in order to assess whether a single DBMS can be used for all of Gaia.

Instances of these three DBMSs will be implemented on identical hardware (probably through the use of virtual machines; see Project 5) and loaded with a testbed data set of sufficient size that a realistic query workload can be executed upon it. Different DBMSs are likely to perform best for different types of query: as noted in the Science Requirements Specification for the *Interrogator* component of the Gaia archive (Tapiador 2011a) *'It is assumed . . . that there will not be a single technology that will out–perform the others in all the different use cases foreseen . . . there will probably be more than one solution in operation . . . '*, and that . . . *'there might be cases where two or more back–ends are put in place . . . so that different queries might be redirected to the most appropriate back–end in terms of performance or other service.'* In addition to performance against this benchmark workload, it will be important to assess how easy it will be to load data from the Gaia Main Database (MDB) into each candidate for the archive DBMS - i.e. in the language of the 'Gaia Catalogue and Archive Software Requirements and Specification' (O'Mullane 2009), we need to prototype the implementation of an *Ingestor* for each system.

In summary, this work will consist of i) definition of a testbed data set (most likely involving GUMS simulated data[11]) and benchmark query workload from Brown et al. (2012); ii) deployment of the testbed data set in a conventional RDBMS (probably Microsoft SQLServer, as used by WFAU for its existing multi-TB sky survey archives), a column-oriented DBMS (MonetDB) and an object-oriented database (Intersystems Caché); iii) running a controlled set of benchmarks; iv) designing and prototyping an *Ingestor* to extract data from the Gaia Main Database (MDB) into each of these candidate DBMSs; v) documentation, including a detailed report and recommendations on the outcome for CU9, ensuring that what is prototyped here is compatible with requirements and the ESAC infrastructure hosting the Gaia archive.

[8] http://monetdb.cwi.nl/
[9] see http://www.scilens.org/CaseStudies/AstroScilens
[10] http://www.intersystems.com/cache/
[11] http://gaia.am.ub.es/GUMS-10/

Aspects of DBMS interface design

The Gaia mission will produce a wide variety of data products, leading to a complex archive. A crucial issue for the exploitability of the Gaia data set is, therefore, an archive interface that supports a sufficiently rich range of functionality and is sufficiently easy to use for users to do their science with it effectively. WFAU's prototyping of server–side infrastructure employing Web 2.0 technologies (see Project 5) for the delivery of an intuitive, but richly-functioned user interface to sky survey archives with a complicated schema appears promising for Gaia: functionality like making schema information readily available to users as they develop their queries, and, even, using code completion to help write them, can make archive use much more effective. What is most important is that the functionality prototyped is that prioritised by scientists, and that the testbed developed here helps the user community to further refine their expressed requirements. For example, while GAP has successfully engaged the Gaia user community via a call for 'usage scenarios' (Brown et al. 2012), iteration of requirements with these key consumers has not been considered so far. We propose to use a benchmark system for an initial deployment as a testbed for the community to further assess its requirements.

Use of VO infrastructure and Data Centre Coordination

With the Table Access Protocol (TAP[12]) in combination with Edinburgh's innovative *TAPFactory* (see Project 5), the VO provides users with the means of querying any combination of TAP–published catalogues on–the–fly. In practice, however, a naive spatial crossmatch query executed between distributed multi-TB data sets will remain expensive, given network speeds. There remains, therefore, a need for coordination between data centres to support the cross-correlation between large data sets. The aim here is to assess through quantitative analysis and, where possible, direct experimentation, the optimal configuration of the multi–wavelength data sets required for the scientific exploitation of Gaia - e.g. for which external catalogues will it be best to co-locate a copy with the Gaia archive, for which should "cross-neighbour" tables be pre-computed to facilitate queries between data sets that remain geographically separated, and for which can cross-matches be performed on–the–fly with sufficient speed. We propose a focused programme of VO–enabled applications development work on the tools required for Gaia exploitation, building on the generic work undertaken in Project 5.

3.4 KEY DELIVERABLES and MILESTONES

Milestones (including some outwith the period of this grant proposal period for context):

Q4 2012: ESA AO for participation in Gaia Coordination Unit 9

Q2 2013: WFAU participation in pan–European CU9 resourced via this grant commences

Q3 2013: Gaia launch (at the time of writing, scheduled for the 1st week of September 2013)

Q2 2015: First release of Gaia data via ESAC central access system[13]

Q4 2015: Second release of Gaia data

Q4 2016: Third release of Gaia data

Deliverables:

Q3 2013: First annual 'advanced data product' release of GES data[14]

Q1 2014: Gaia archive interface demonstrator

Q2 2014: DBMS benchmarking report for the Gaia archive

Q3 2014: Second annual release of GES data

Q1 2015: SW prototype contributions to Gaia archive system

Q3 2015: Third annual release of GES data

[12] http://www.ivoa.net/Documents/TAP/
[13] according to O'Mullane & van Leeuwen (2012)
[14] from the Gaia–ESO Survey Management plan

Q1 2016: Final SW module contributions to Gaia archive system

Q3 2016: Fourth annual release of GES data

3.5 RESOURCES REQUESTED

Support for overall scientific leadership for the elements of this project is requested at the level of 0.15 FTE for N. Hambly over the entire grant period. We request a further 0.85 FTE of Hambly over the period Q2 2015 to Q1 2016, together with the following mixture of development and operations effort: Collins – 0.1 FTE at Q2 2013 rising through 0.25 FTE at Q2 2014 to 0.5 FTE at Q4 2014, and continuing at that level to Q2 2016; Cross – 0.1 FTE from Q2 2014 to Q1 2015; Read – 0.25 FTE from Q4 2014 to Q1 2016; Sutorius – 0.25 FTE from Q2 2013 to Q2 2014 and 0.5 FTE thereafter to Q1 2016; and Voutsinas – 0.25 FTE from Q2 2013 to Q1 2015.

Using GDFS/PLS–experienced staff (Collins, Cross & Hambly) is an extremely cost–effective way of providing UK contributions to Gaia exploitation activities, since the steep learning curve associated with this highly complex mission has been ascended already as part of the ground–segment developments.

In Summary:

(a) *Staff.* 0.75 FTE at Q2 2013 rising through 1.0 FTE (at Q2 2014) and 1.75 FTE (at Q4 2014) to 2.25 FTE (at Q1 2015)
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE
(d) *Equipment & maintenance:* please see Appendix D

3.6 ADDITIONAL REFERENCES

Brown A.G.A., Arenou F., Hambly N.C., van Leeuwen F., Luri X., Malapert J.-C., O'Mullane W., Tapiador D., Walton N.A., 2012, *Gaia data access scenarios summary*, Gaia DPAC document GAIA-C9-TN-LEI-AB-026, available from `http://www.rssd.esa.int/SYS/docs/ll_transfers/project=PUBDB&id=3125400.pdf`

Hambly N.C. et al., 2008, *The WFCAM Science Archive*, MNRAS, 384, p. 637

Gilmore G. et al., 2012, *The Gaia–ESO Public Spectroscopic Survey*, The Messenger March 2012, No. 147, p. 25

O'Mullane W., 2009, *Gaia Catalogue and Archive Software Requirements and Specification*, Gaia DPAC document GAIA-C9-SD-ESAC-WOM-033, available from `https://gaia.am.ub.es/Twiki/pub/GENIUS/ReferenceDocuments/Gaia_Catalogue_and_Archive_Software_Requirements_and_Specification.pdf`

O'Mullane W., van Leeuwen F., 2012, *Release scenarios for the Gaia archive*, Gaia DPAC document GAIA-C9-TN-ESAC-WOM-066

Tapiador D., 2011, *Interrogator Software Requirements and Specification*, Gaia DPAC document GAIA-C9-SP-ESAC-DTP-001, available from `https://gaia.am.ub.es/Twiki/pub/GENIUS/ReferenceDocuments/GAIA-C9-SP-ESAC-DTP-001-D.pdf`

Project 4: Enabling PS1 Exploitation

Staff involved

R. S. Collins
S. Voutsinas

4.1 BACKGROUND

PS1, the prototype for the Panoramic Survey Telescope & Rapid Response System (Pan-STARRS) is a survey system centred on a dedicated 1.8m telescope on Haleakala, Maui, with a 7 sq. deg. FOV and the world's largest CCD camera (1.4×10^{12} pixels). Initially a prototype for a larger system – PS4, with four such telescopes – it is a very powerful survey facility in its own right. Construction of PS1 was funded primarily by the US military, and it is being operated by the PS1 Science Consortium (PS1SC[1]).

The PS1SC is conducting a 3.5 year survey programme. The main survey, taking ∼55% of the time, is covering 3π steradians ($\delta > -30°$) to $r \sim 24$ (plus corresponding depths in g,i,z and y), while a Medium Deep Survey will cover ∼ 80 sq. deg. in ten fields (see Table 4.1) 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 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.

Name	Alternative name	RA (J2000)	Dec (J2000)
MD01	XMM-LSS/UKIDSS DXS/VVDS 02h	02 23 30	-04 15 00
MD02	CDFS/GOODS/GEMS	03 32 24	-27 48 00
MD03	Lynx	08 42 22	+44 19 00
MD04	COSMOS	10 00 00	+02 12 00
MD05	Lockman/UKIDSS DXS	10 47 40	+58 05 00
MD06	M106/NGC 4258	12 20 00	+47 07 00
MD07	DEEP2 Field 1/Extended Groth Strip	14 14 49	+53 05 00
MD08	ELAIS N1/UKIDSS DXS	16 11 09	+54 57 00
MD09	SA22/UKIDSS DXS/VVDS 22h	22 16 45	+00 17 00
MD10	DEEP2 Field 3	23 29 15	-00 26 00

Table 4.1: *Field centres for the PS1 Medium Deep Fields. All overlap with existing multi-wavelength survey fields (as indicated by the alternative name), with the exception of MD03, which covers the Lynx radio survey field, and MD06, which includes the spiral galaxy M106/NGC4258, which contains a water maser.*

PS1 data is currently only available in the UK to researchers from Durham, Edinburgh and QUB, which have

[1]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; the NASA Near Earth Object Program; and a UK consortium composed of Durham, Edinburgh and Queen's Belfast.

joined the PS1SC, but it is certain to become a prime resource once it becomes public, being a somewhat deeper and much wider version of the SDSS, with all the scientific possibilities that opens up. As discussed in Section 4.3, however, the case for WFAU's implementing a copy of the PS1 catalogue is based specifically upon the additional science that it will enable when integrated with survey data from two other STFC-funded projects, namely radio surveys from LOFAR and the UKIRT Hemisphere Survey.

4.2 REPORT ON RECENT WORK: 2009 − 2012

The last WFAU rolling grant award contained ~ 1 FTE of staff effort for three years to develop a PS1 archive, probably in collaboration with the SDSS archive group at Johns Hopkins University (JHU). The intention was that this development would take advantage of the access Edinburgh astronomers have to proprietary PS1 data, so that a complete archive could be built in Edinburgh ready for the date at which the data become public.

In the event, the PS1 project has been delayed by problems with both the telescope and the data reduction system. These problems have now generally been overcome, but full survey operations did not start until April 2010, so that the 3.5 year survey programme is now due for completion in Q4 2013. In the light of these delays, it was decided not to devote WFAU effort to development of a PS1 archive during the reporting period, although contact has been maintained with the PS1 data management team: Hambly and Mann have both been members of relevant PS1 working groups. The JHU team was contracted by the PS1 project to deliver an archive for installation in Honolulu to service the requirements of the PS1SC, and, subsequently, an agreement has been made with STScI for a PS1 legacy archive to be implemented within the STScI MAST[2] archive system. This is due to made public in Q4 2014, and our proposed programme of work to enable exploitation of PS1 data by the UK community now centres on adding value to a copy of that archive in Edinburgh through combination with other surveys datasets.

4.3 PROPOSED PROGRAMME OF WORK Q3 2014 − Q1 2016

The scientific motivation for having a copy of the PS1 data in Edinburgh is to enable multi-wavelength analyses made possible by the addition of optical data to existing northern hemisphere survey datasets. The foremost of these will be the combination of PS1, UHS and WISE data, but the provision of optical counterparts for radio surveys from LOFAR[3] (and, potentially, Apertif[4]) will greatly enhance what can be learnt from them, as well as providing an excellent testbed for the development of data management and analysis tools for the future southern hemisphere pairing of LSST[5] and SKA[6]. We see no strong motivation for obtaining a copy of the PS1 pixel data (which are likely to total ~ 1.6PB), so all that is discussed here concerns PS1 catalogue data.

4.3.1 Installation of a copy of the PS1 catalogue

Given the current arrangements for archiving PS1 data, the most cost-effective way to establish a PS1 archive in Edinburgh is to take a copy of the database files to be created by MAST at STScI. This should present no major technical challenges, since the MAST sky survey databases are hosted in the same database management system (Microsoft SQLServer) as used by WFAU, and we have recently successfully exchanged UKIDSS and GALEX database files with MAST. However, some development work is likely be required to ready the PS1 catalogue data for integration with other WFAU datasets, but that is difficult to quantify until the design of the MAST system is finalised. We assume that it will be easier for MAST staff to help with the transfer of PS1 data to Edinburgh once their PS1 archive has gone live, so we have scheduled our own work to start in Q3 2014, but we would investigate starting it earlier were that possible for our MAST colleagues.

[2] http://archive.stsci.edu/
[3] http://www.lofar.org
[4] http://www.astron.nl/general/apertif/apertif
[5] http://www.lsst.org
[6] http://www.skatelescope.org

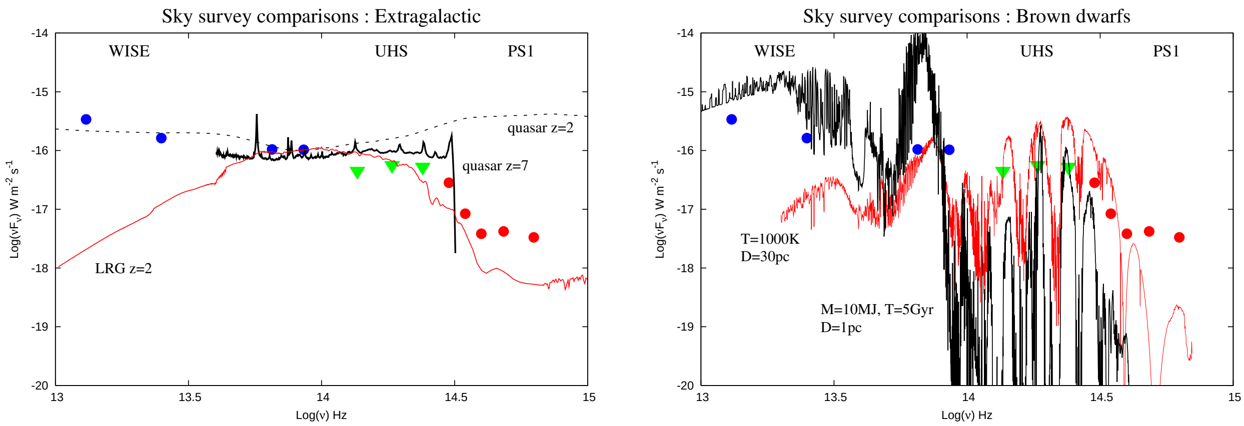

Figure 4.1: *Spectral energy distributions of various objects compared to the 5σ sensitivities of the WISE (filled circles, from Wright et al 2010), UHS (triangles) and PS1 (open circles, Chambers, private communication) surveys. The left hand frame compares extragalactic objects. The "quasar z=2" curve uses the mean quasar continuum SED from Elvis et al (1994) redshifted to $z = 2$ and scaled so that $i_{\rm AB} = 20.1$, similar to the faintest quasars in the SDSS catalogue. The "quasar z=7" curve is a model, but scaled to represent the observed flux of ULAS J1120+0641 (Mortlock et al 2011) at 1.1 µm. The "LRG z=2" curve is a model ellipical galaxy SED redshifted to $z = 2$ and scaled so that $K_{\rm point} = 18.0$, corresponding to the brightest radio galaxies at $z = 2$, with $K_{\rm total} = 17.5$. The right hand frame compares two model brown dwarf spectra, from Burrows et al (2003, 2006). The solid line (lower curve at low frequency) is for an object with effective temperature of 1000K and surface gravity of 4.5, placed at a distance of 50pc. The dashed line (upper curve at lower frequency) is for an object with mass of 10 Jupiter masses and age 5 Gyr, placed 1 pc away.*

The PS1+UHS+WISE combination

The principal motivation for bringing a copy of the PS1 catalogue to Edinburgh is to combine it with the UHS and WISE datasets to generate a hemispheric survey that stretches two decades in wavelength from the optical g band through to the mid-IR. As shown in Figure 4.1, the sensitivities[7] of the three surveys are well matched, and the power of their combination is that different types of source will be found in different bands, thereby enabling classification of sources that would be not possible with data from one passband alone. For example, relatively warm brown dwarfs can be seen in UHS and PS1 to large distances, but will not be detectable in WISE. The UHS-PS1 combination will provide proper motions, and the yJK colours needed for the separation of types. On the other hand, the coolest and nearest brown dwarfs will be invisible in PS1, but detectable in WISE and UHS. Likewise, a $z = 2$ quasar will be seen in all three surveys, whereas a $z = 7$ quasar will be seen only in UHS and the PS1 y band and will need both PS1 and UHS data to separate them out from T-dwarfs and lower-z quasars. In between these two extremes, the PS1+UHS+WISE combination should yield a sample of $\sim 10^4$ Luminous Red Galaxies at $z \sim 2$, making possible the detection of Baryon Acoustic Oscillations at much higher redshift than possible with SDSS-III BOSS spectroscopy (at $\bar{z} \sim 0.7$).

4.3.2 Supporting determination of optical counterparts of LOFAR surveys

The second motivation for having a copy of the PS1 catalogue is to aid exploitation of the LOFAR radio surveys. LOFAR is a new-generation radio telescope operating at low radio frequencies, below 240 MHz, one of the few largely unexplored regions of the spectrum. With its unique design of stations of simple antennas, that are 'beam-formed' in software to look in a particular direction, it offers a combination of sensitivity and sky coverage that make it more than an order of magnitude faster than any other radio telescope for large-area radio surveys. Over a 5-year period beginning in 2013, LOFAR will carry out deep multi-frequency surveys of the entire northern sky, as well as ultradeep pointings over hundreds of square degrees. With

[7] This figure is taken from the UKIDSS Extension proposal submitted to the Director JAC in March 2011. The UHS sensitivities are for the full four-year survey. Closure of UKIRT in September 2013 would yield a UHS that is 0.3 magnitudes fainter in both bands, but this difference in depth does not alter the conclusions presented in the text as to the discriminatory power of the PS1+UHS+WISE combination.

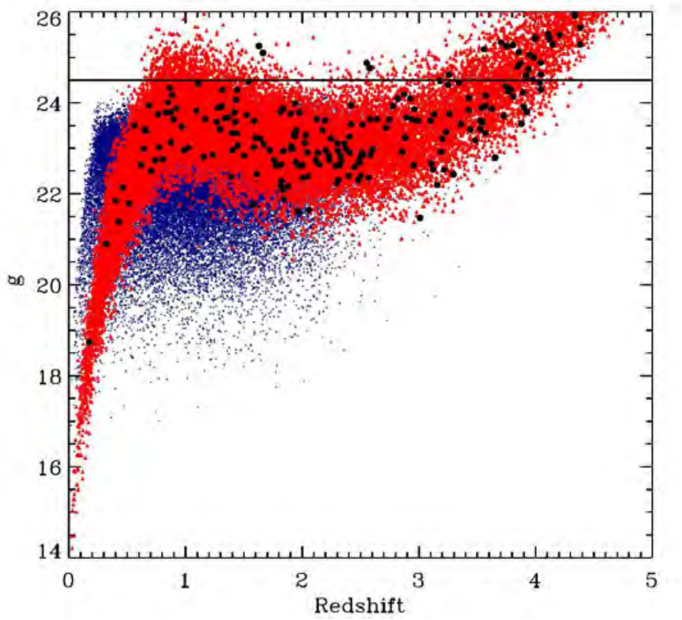

Figure 4.2: *A simulated g-magnitude vs redshift diagram for radio sources expected to be detected over 1 square degree of LOFAR's deeper tier surveys (from Best et al 2008). The blue points are star-forming galaxies, the red are FR1-like radio AGN, and the black are FR2-like radio AGN (over 100 sq. deg., to increase the numbers). The horizontal line is the g band detection limit from the Pan-STARRS1 Design Reference Mission: the real 3π survey is likely to be slightly shallower, but it remains true that almost all the LOFAR survey sources should be detected by PS1.*

the inclusion of international stations, sub-arcsecond angular resolution is achievable at the highest LOFAR frequencies.

The scientific goals of these surveys are wide-ranging, from detailed investigations of our own galaxy, through studying the evolution of star-forming galaxies and AGN across a wide range of cosmic epoch, to large-scale structure and cosmological studies. A crucial aspect of all of these goals is the ability to move beyond simply the radio catalogue, to identify the optical counterparts of the radio sources, classify and categorise these, and determine fundamental properties such as estimates of redshift, mass, etc. The combination of the PS1+UHS+WISE dataset discussed above is the best available to achieve this, and will be sufficiently deep to provide identifications and colour information for the majority of the detected radio sources (as shown in Figure 4.2).

The determination of optical counterparts for LOFAR survey sources will be supported through a two-stage programme. The first will import and make available the LOFAR Survey catalogues, and produce simple PS1-LOFAR cross-neighbour tables, in the manner already provided for SDSS and UKIDSS, for example. This will provide users with a set of potential optical counterparts for each radio source, freeing them from needing to undertake the conceptually simple (but computationally expensive) spatial matching operation between the two catalogues and enabling them to concentrate on studying the more complicated cases where there are several optical sources within the error radius of a LOFAR source. The second phase will support that process without requiring the user to download data, by providing them with virtual machines (VMs) within which they can run more sophisticated probabilistic cross-matching algorithms that include non-spatial attributes. This will provide an ideal testbed for the VM deployment capabilities described in Section 5.3.5, and, in so doing, also constitute a very valuable prototype for the kinds of cross-matching services that will need to be provided within the data centre for future surveys such as LSST and SKA.

4.4 KEY DELIVERABLES and MILESTONES

Milestones:

Q4 2013: End of PS1 operations

Q3 2013: Start of LOFAR surveys

Q4 2014: Launch of MAST PS1 legacy archive

Deliverables:

Q3 2015: Working copy of PS1 archive with UHS cross-neighbours

Q4 2015: LOFAR-PS1 cross-neighbours table

Q1 2016: Facility for running generalised LOFAR-PS1 cross-matching

4.5 RESOURCES REQUESTED

(a) *Staff.* 0.25 FTE from Q4 2014 split 40/60 between Voutsinas & Collins
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE
(d) *Equipment & maintenance:* please see Appendix D

Project 5: Future archive infrastructure

Staff involved D. Morris

5.1 BACKGROUND

As discussed in earlier Sections of this proposal, one of the strengths of the WFAU approach to science archive development is that, where possible, we develop generic systems that are then configured for specific archives, rather than creating each anew as a standalone entity. This has been the philosophy of the VDFS project from the outset, enabling the basis of the VSA to be produced from the same software as the operational WSA, and allowing us to provide archives for VST/ATLAS and HARPS-N at minimal marginal cost, given our investment in the WSA/VSA and GES archives, respectively.

In this Project we discuss the next steps in the development of that generic archive infrastructure. These will be undertaken within an environment in which many archives are published using VO protocols and most astronomers are using client-side tools (e.g. TOPCAT[1] for tabular data, SPLAT[2] for spectra and Aladin[3] or GAIA[4] for images) that use those protocols extensively under-the-hood. This user community will also have greater expectations for the interactivity of the services provided to them; used to Web 2.0 systems outside work, they will come to expect ("require") the same level of personalisation and the same ready means of internet-enabled collaboration in their research as in their personal lives.

While we detail in Section 5.4 the schedule for rolling these new services out onto our existing archives, we emphasise that this work is also planned with an eye to the next generation of survey archives (e.g. first the SKA pre-cursors, then LSST and SKA itself), which will be so large that an increasing amount of science will be done *in the archive*. The current paradigm – defined by the SDSS SkyServer, and followed by the WSA and VSA – implicitly assumes that, while users will specify the data subsets they want to analyse through SQL queries on the survey database, the size of those datasets will be sufficiently modest that they can readily download them and analyse them on their desktop. This will not be possible for much longer and it is important to gain experience in supporting that analysis in the archive now, before the absence of that support becomes the bottleneck in the process of scientific exploitation of survey datasets.

All the enhancements described below are proposed in response to a definite identified need – whether one raised explicitly by our users, one demonstrated implicitly through the active use of analogous services elsewhere or one revealed through our own analysis of how our existing services are used. In each case, there is an existing prototype or proof-of-concept to learn from, so we can assess the size of each development task.

5.2 REPORT ON RECENT WORK: 2009 – 2012

We have been prototyping most of these intended new infrastructural services over the past couple of years, largely supported through other sources - e.g. MSc projects, FP7 grants, a JISC initiative - and this exercise has shown how we can greatly enhance the usability and functionality of our archive interface without major investment in people or machines. By drawing upon and extending emerging technologies provided by the

[1] `http://www.star.bris.ac.uk/~mbt/topcat/`
[2] `http://star-www.dur.ac.uk/~pdraper/splat/splat.html`
[3] `http://aladin.u-strasbg.fr/`
[4] `http://star-www.dur.ac.uk/~pdraper/gaia/gaia.html`

Virtual Observatory, by the wider e-science community (e.g. OGSA-DAI[5]), and through using generic Web 2.0 tools and services (e.g. the web.py[6] framework) we can provide users with tools enabling them to integrate and analyse data in more efficient ways, thereby increasing the science that can be extracted from existing data resources.

5.2.1 Achievements: a new interactive user interface

The rapid rise of AJAX[7] and similar technologies to support Web 2.0 systems is blurring the traditional boundary between what can be run through a web browser and what requires a desktop application. These same technologies that support richly featured social networking sites can be applied to offer a more interactive browser-based front end to astronomical archives, which are better matched to astronomers' workflows than the more static websites that provide access to most astronomical archives.

Analyses[8] of the logs of the queries executed on the WSA show how many of its users build up complex SQL queries iteratively – starting with a basic query, looking at the result set it generates, then extending or amending the query to bring the results more in line with requirements. At the moment, the WSA interface displays a subset of the query results as a table, but often what the user really wants is a quick graphical summary of a result set through a scatter plot or a histogram. Currently, that requires downloading the result set into a locally-installed copy of TOPCAT or some similar desktop application. The datasets involved in this iterative process are usually small – whether limited through a spatial selection or by use of the SQL `top` clause – so a more efficient workflow would be provided by including simple data plotting functionality within the archive interface, so that the iterative process of running a query, checking the results and running a tweaked query requires nothing more than flicking between two browser tabs. (More involved statistical analysis or analysis of larger datasets will still require TOPCAT, and we are curently investigating the use of the IVOA WebSAMP[9] standard as a way of communicating between the browser and an instance of TOPCAT running on the client.)

In Figure 5.1 we show a screenshot from a prototype of such a browser-based tool, intially developed by Stelios Voutsinas as part of his MSc project, supervised by Keith Noddle, with further development funded by the last WFAU rolling grant. A user has executed a query (from Section 4.2.1 of Hambly et al 2008) on the WSA to select candidate members of the Sigma Orionis cluster and plotted a (z v $z - J$) colour-magnitude diagram of the stars in the field using the Table Viewer tool in the new interface, which opens in a new tab in the user's browser. The cluster members trace a tight locus (due to their common distance) to the right of the looser cloud of other stars in the field and the separation of the two populations in this plot allows the user to work out the colour-magnitude cut that must be added to the original query to isolate the cluster members. This additional clause may then just be added to SQL query text box that has remained open in the first browser tab and executed, to yield a clean sample of cluster members whose selection has been made solely using the browser. (*N.B.* this particular plot is a combination of a scatter plot and a density plot: individual data points are plotted in sparsely-populated regions, but are replaced by density values — in user-specified grid cells —where the density is high, so that the morphology of the main locus can be shown in addition to the positions of individual outliers from it.)

5.2.2 Achievements: federated data access

One the principal goals of the VO has always been to facilitate the cross-querying of distributed catalogues. Somewhat remarkably, this has only recently become possible in a standards-compliant fashion. Malik et al (2002) presented a prototype of a such a service more than ten years ago, but it relied on features specific to Microsoft's .Net framework and only worked for Microsoft SQLServer databases with a particular schema, and so had limited applicability, despite obvious power.

The appearance of a generalised version of this service has been slowed by the delay in the agreement of the Table Access Protocol (TAP), which is the IVOA standard for querying tabular data (typically held in relational databases). TAP is potentially a game-changer for the VO: WFAU took (through Keith Noddle) a

[5] `http://www.ogsadai.org.uk/`
[6] `http://webpy.org/`
[7] `http://en.wikipedia.org/wiki/Ajax_(programming)`
[8] e.g. F. Asger, University of Edinburgh MSc dissertation, in preparation
[9] `http://www.star.bristol.ac.uk/~mbt/websamp/`

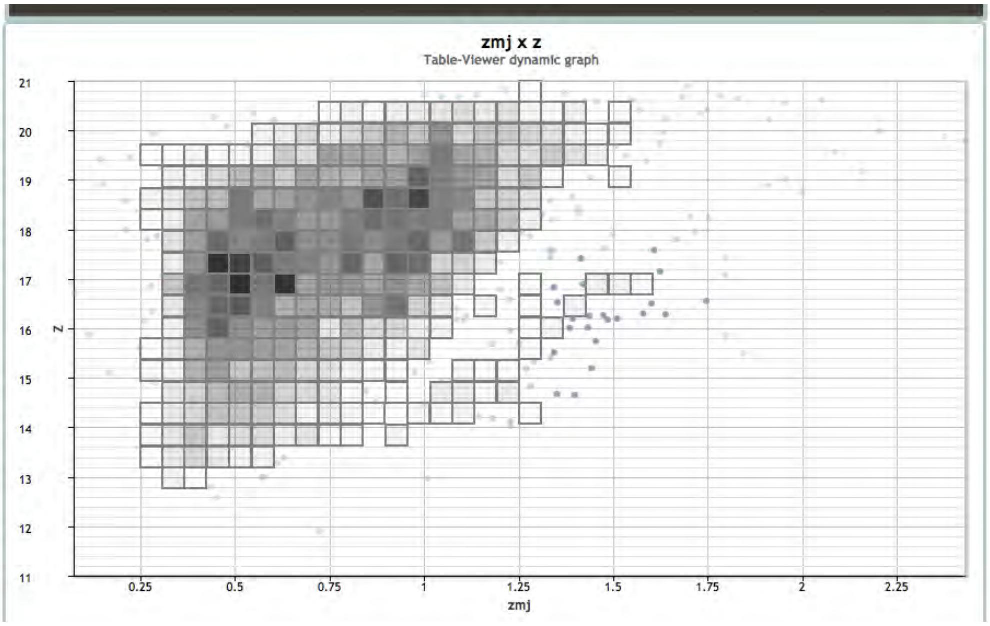

Figure 5.1: *A (z v $z - J$) colour-magnitude diagram of the stars in the Sigma Orionis field, generated with the Table Viewer tool in the prototype Web 2.0 interactive browser-based user interface, using a result set generated from querying the UKIDSS DR7 database from a different tab in the same browser. The cluster members trace the tight locus to the right in this plot of the amorphous cloud of other stars in the field. This particular plot is a combination of a scatter plot and a density plot: individual data points are plotted in sparsely-populated regions, but are replaced by density values (in user-specified grid cells) where the density of points is high, so that the morphology of the main locus is shown in addition to the positions of individual outliers from it.*

pivotal role in its development and has been at the forefront of its adoption – for example, we are currently the only data centre using an authenticated variant of TAP to provide access to proprietary data, which is a key step in having the VO deliver what the community really needs.

More importantly, we have devised the first TAP-based service for querying multiple databases, working with our Edinburgh colleagues at EPCC to develop *TAPFactory* (Hume et al 2011), which uses EPCC's OGSA-DAI middleware framework to combine multiple TAP services and present them to the user as a single TAP service. This is illustrated in the panel (a) of Figure 5.2, where two databases, A and ANN (the meaning of this latter name will be explained in Section 5.2.3) are exposed through TAP services and OGSA-DAI has created the uppermost TAP service from them through the TAPFactory mechanism. A user can now send a single query to this new TAP service, and have OGSA-DAI's Distributed Query Processor (DQP) decompose it into subqueries that it submits to each of the original TAP services, before gathering the results from each of these and combining them into a final result set that is returned to the user. The TAP services accept queries in Astronomy Data Query Language[10] (an astronomy-specific variant of SQL) and they issue queries to data resources in whatever dialect of SQL they use. In this way, we have been able to produce the VO's first, generic, standards-compliant distributed query processing service.

This work was funded intitally by the Euro-VO AIDA[11] and ADMIRE[12] FP7 projects, while latterly Stelios Voutsinas has been been supported by the last WFAU rolling grant to incorporate this facility into his new archive interface, through which users can now issue queries against databases distributed anywhere in the world, so long as they have been published to the VO using TAP. Federating TAP services this way, and making them accessible through a simple, yet powerful, user interface like this is a major milestone in the realisation of the VO concept.

[10] http://www.ivoa.net/Documents/latest/ADQL.html
[11] http://cds.u-strasbg.fr/twikiAIDA/bin/view/EuroVOAIDA/WebHome
[12] http://www.admire-project.eu/

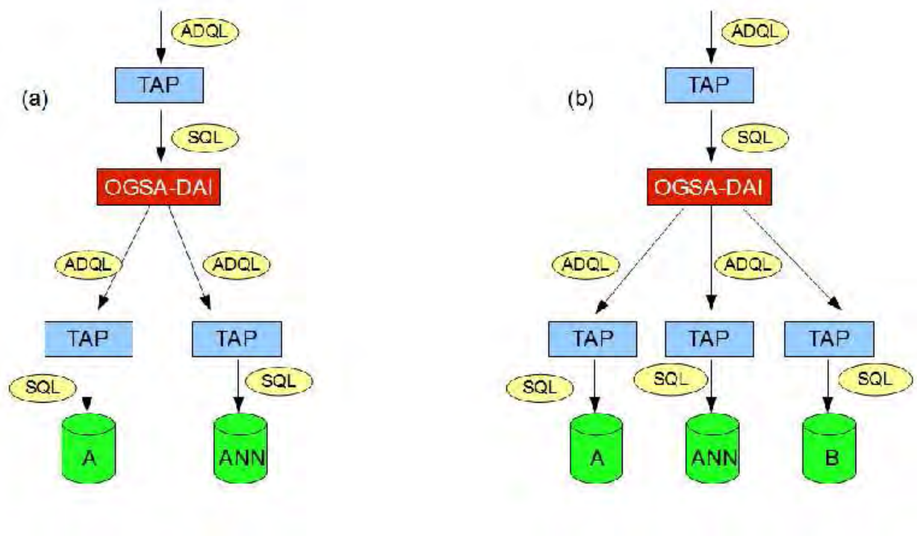

Figure 5.2: *The use of the AstroDAbis service in conjunction with a TAPFactory service based on OGSA-DAI. In (a) a user is able to cross-query an annotation database (ANN) with another database, A, while in (b) the annotations in ANN are being used as a join index in a cross-match between A and a third database, B.*

5.2.3 Achievements: a data annotation service

A striking feature of the data held in most sky survey archives is that is very low-level: measurements of positions, fluxes, shapes, etc. Some classification information (e.g. star/galaxy separation) may be present, but most of the scientific interpretation of the data is to be found in journal papers describing and using the data, not in the archive itself. The difficulty of linking automatically from the literature to the archives means that it is impossible for archive users to include this interpretative information in their queries - e.g. extract from the WSA all sources listed as likely to be obscured AGN in some MNRAS paper.

Analogous problems in other disciplines, notably the life sciences, have been solved via third-party annotation services, in which *anyone* can publish, in a queryable format, some comment about an entry in a database over which they have no control. In the JISC-funded AstroDAbis[13] project (Gray et al 2011) we have developed a similar annotation system for astronomy, whereby users can publish TAP services that contain annotations on entries in databases published to the VO through TAP. This is illustrated in Figure 5.2. As an example of what this enables, consider the case where A is the WSA, and ANN is an annotation database into which has been loaded the table of obscured AGN from the aforementioned fictitious MNRAS paper. Panel (a) shows the situation where a user is able to run a distributed join query between A and ANN and so extract from the WSA further near-IR attributes of these candidate obscured AGN not included in the original paper, and so investigate their properties further. The WSA itself contains a cross-neighbour table between UKIDSS tables and the SDSS, so, if database B in panel (b) is a copy of SDSS, the user can now perform a three-way join between A, B and ANN and, so, extract any SDSS spectroscopic redshifts that exist for this sample of candidate obscured AGN. Clearly, to do this without access to the annotation database would require the user to perform a number of data download and manipulation operations.

5.3 PROPOSED PROGRAMME OF WORK Q2 2013 – Q1 2016

In the previous Section we have shown how we have used modest investments of resources – often funded by other sources – to design and prototype many of the building blocks of the vision outlined in Section 0.4. The task for this Project over the next three years is to turn that vision into a reality, transforming these prototypes into robust, reliable, scalable and maintainable services that can be seamlessly integrated into

[13]http://code.google.com/p/astrodabis/

our existing archive services (and that interoperate smoothly with the rest of the VO) and to supplementing them, where necessary, with further services to complete the picture. Many of the services developed here will have a wider applicability, and our intention is to make the source code for them available to interested parties, in a similar manner to what our colleagues at the Canadian Astronomy Data Centre have done with their OpenCADC[14] initiative.

5.3.1 "Productisation" of prototype services

The work described in the previous Section has delivered services that are still in prototype, or, at best, in early adoption phase. To make them ready for production use will require some re-engineering and hardening work, plus the extension of some functionality. For example, the OGSA-DAI DQP is capable of making sophisticated execution plans for distributed queries, but to do so requires more information about the data resources than is currently required by the TAP standard. We are currently pushing in the IVOA for the extension of the TAP standard to support this additional metadata, and will extend our existing TAPFactory software in parallel with the developing standard so that the VO community can fully exploit what DQP can offer.

Our new interactive user interface is a great advance of what is currently offered, but will inevitably need some modification in the light of user feedback. As it becomes used in earnest we are also likely to put more effort into automating decisions as to which aspects are done server-side and which are done in the browser: the current system uses a combination of the two approaches, depending on the size of the dataset involved, but this will need further refinement.

5.3.2 User storage in the archive

As discussed in Section 5.2.1 above, we see from the WSA query logs that users develop sophisticated queries iteratively. Our new interactive user interface is the first stage in supporting that mode of work, but, while it can aid the iterative refinement of queries, each query is still executed on the full database table whereas, often, the iteration is refining selection criteria applied to a modest subset of the table defined by an initial broad cut. In that case, it would be more efficient to store the results of the initial, broad selection and experiment with more restrictive constraints through queries of that subset, rather than querying the full table anew each time.

Enabling that workflow is the first motivation for offering WFAU archive users a service akin to the MyDB personal database within the SDSS CasJobs system. That system is well used[15] with more than 4000 active MyDBs currently in use. The storage implications of offering this service are modest – more than 95% of the databases remain at the initial size of 0.5GB, but some are increased significantly to meet the requirements of a small number of power users – but the benefits to the user community are significant. Users can also upload data into their MyDB and use them in join queries with SDSS databases, and storage in a MyDB provides a user with a means of persisting a result set and of sharing it with collaborators, through use of appropriate access control mechanisms.

Seeing the success of the SDSS MyDB system, WFAU has long wished to offer a similar facility to its archive users – and has been asked by them to provide one – but we have been reluctant to do so before it could be done in an IVOA-compliant manner. The advent of TAP, and our work on TAPFactory, provide support for the inclusion of MyDB-like tables in distributed queries in a standards-compliant manner, and allow them to explored using TOPCAT, while the loading of data into such tables would be accomplished through a long-planned extension to the IVOA VOSpace standard[16]: VOSpace currently only offers users access to remotely held flat files, but it was always envisaged that it would be extended to include remote tables, and we have concluded that the best way to drive the IVOA to adopting this extension to the VOSpace standard is to deploy a demonstrably powerful new service that requires it. We have a design for this new standards-compliant VOSpace-meets-MyDB system, and the first new development task of this Project will be to produce a working service implementing that design.

[14] `http://code.google.com/p/opencadc/`
[15] MyDB usage statistics kindly provided by A. Thakar, Johns Hopkins University.
[16] `http://www.ivoa.net/Documents/VOSpace/`

5.3.3 Replacing the AstroGrid Registry

As VO services emerge across the astronomical community it is vital these services can be located and exploited. Locating these services is the role of the IVOA Registry. AstroGrid, the UK Virtual Observatory project, played a leading role in the development of the Registry Standard and even today, the AstroGrid Registry is the premier Registry service in the VO world. However, since the demise of AstroGrid, little maintenance of that component has taken place and as new standards and services emerge, it is showing its age. In order that we can continue to make fullest use of VO services across the world it is vital we have an up-to-date, functional Registry – for example, the VOSpace-meets-MyDB system outlined in Section 5.3.2 will only work if users are able to publish their personal databases as TAP services and have them registered so that query tools, such as TOPCAT and our own TAPFactory, can locate them . We propose to build such a service, utilizing the best of the AstroGrid Registry component and adding the missing elements.

5.3.4 Authentication and authorisation

WFAU currently serves proprietary data to more than 1000 registered users, but the way that this is done cannot be readily extended to cover the new situations of users making join queries between a proprietary sky survey data and their own (or a collaborator's) MyDB-like personal database, nor of joins executed between two proprietary survey archives. It is necessary, therefore, for us to overhaul our authentication and authorisation infrastructure to support these new use cases.

Authentication and authorization are two distinct activities even though they are often conflated. Authentication is the process by which users verify who they are. Once done, usability demands that they are not asked again to undertake this authentication. This principle is known as "Single Sign-on" and the IVOA has a comprehensive standard covering single sign-on, called the Credential Delegation Protocol (CDP[17]). The CDP allows a client program to delegate a user's credentials to a service such that that service may make requests of other services in the name of that user. We propose to implement this standard across WFAU archive services.

Once a user's identify has been authenticated, they still need to be authorized to allow them access to restricted resources. Traditionally the management of user authorization has been simplified by the implementation of User Groups and/or Access Control Lists. Even then administration becomes complex with implied authorization (e.g. membership of a group within a group) being very difficult to track but which may be either be too permissive or too restrictive. Whilst Groups and ACLs are important, significant research undertaken in the JISC-sponsored AGAST[18] project has shown how new semantic technologies can be used to greatly enhance the use and administration of authorization requirements.

We propose to implement the mechanics of authorization supplemented by the sophistication of AGAST to delivery authorization control for our archives. This is a clear investment in an expanding future that will make the task manageable rather than one that grows out of control and at worst, becomes a barrier to legitimate data access. We can share this technology with the wider community by making it open source as required.

5.3.5 Virtual Machines

Virtualisation has been a goal of IT almost since the start of modern computing. The advantages were – and are – clear: expensive software is no longer tied to specific hardware or host operating systems, it runs within its own well-defined environment unentangled with the host. The ability to seamlessly move this guest environment – even in a running state – across host platforms is another critical advantage of virtualisation: running software is no longer confined to a single host, it can be transparently moved across hosts to address issues of performance or even host instability. Given even modest servers can host multiple guests simultaneously, the advantages are clear.

Virtualisation is in widespread use throughout industry. The technology has matured significantly of late with most – if not all – modern processors having hardware support built in. In addition modern operating systems also have built-in support for virtual machines (VMs) making the deployment of these systems a

[17] http://www.ivoa.net/Documents/CredentialDelegation/
[18] http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/agast.aspx

routine operation. Industry evaluations have shown that once running, VMs often run with near bare-metal performance thus giving all the advantages of visualization with almost none of the downsides.

WFAU have used VMs for some time to provide one-off services, but recent investigation into state-of-the-art Virtualisation technologies allied to deployment of VMs on Solid State Disks (SSDs) has shown that a VM can boot from scratch to running in as little as three seconds. This brings yet another key advantage: we can create VMs to perform specific tasks and once complete, shut them down with very little overhead. This means that physical machines can be used for multiple, concurrent, disparate tasks without the need for complex reconfiguration or software infrastructure bloat; hardware utilization can be greatly increased.

We propose to complete our evaluation of virtualisation technologies and to reconfigure a significant proportion of our existing compute resources into a Virtual Machine "farm" where we run as many of our services as possible in guest environments. We will also be able to offer VM guests to authorized users allowing them to execute analysis software in an environment they control running next to the data they wish to investigate. This has all the benefits of moving the software to the data (data volumes will make this inevitable within the near future) whilst at the same time protecting our computing resources from errant software bugs - not to mention greatly improving our hardware utilization. Our colleagues in CADC already do something similar with great support from their user community in the CANFAR[19] project, whose success augurs well for what is proposed here.

5.4 KEY DELIVERABLES and MILESTONES

Deliverables:

Q4 2013 : Auth/Auth libraries and components complete

Q2 2014 : Productisation of existing prototypes complete

Q1 2015 : Registry complete

Q1 2016 : MyDB-like system deployed and fully functional

5.5 RESOURCES REQUESTED

We request 1 FTE of staff effort for this project for the full three-year period. Our intention is to employ Dave Morris, one of the former AstroGrid developers to undertake this work. He has been working for WFAU part-time for the past year, designing and prototyping much of the functionality outlined here, and he is ideally suited to this post, having been long associated with VO developments, and, in particular, having been one of the main designers of the VOSpace standard.

(a) *Staff.* 1.0 FTE throughout (Q2 2013 to Q1 2016; Morris)
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE
(d) *Equipment & maintenance:* N/A

[19] `http://canfar.phys.uvic.ca/`

Cross-Project Resources

Staff involved

K. T. Noddle
M. S. Holliman
Pooled Admin Support

6.1 INTRODUCTION

The Projects described in the previous five Sections constitute an ambitious programme of research and development for the next three years. For that programme to be successful, the grant-funded RAs undertaking that work must be well supported in a number of ways: they must receive both scientific direction and technical guidance, to ensure that they are doing what the WFAU user community requires and that they are doing it using sensible techniques and technologies; they must have project management support to ensure that they work effectively towards clearly defined goals in a timely fashion, and that they report their progress as required to satisfy funders and other key stakeholders, such as survey PIs; they require well designed and well maintained computing infrastructure of a specification that meets their needs; and they need basic administrative support so they can concentrate on the work they are funded to perform.

It is the role of the grant PI and Co-Is to provide the scientific direction for the grant-funded RAs, but they are not IT professionals and so cannot provide the best technical guidance. During the course of the last WFAU rolling grant we initiated a Technical Lead role to provide that guidance, and, as discussed below, that has proven to be very successful, thanks to our appointment of Keith Noddle, the former AstroGrid Project Manager into a post that combines that role and the duties of Project Manager. WFAU's substantial and specialist computing infrastructure is managed by Mark Holliman, who has managed a smooth evolutionary change in the technologies delivering that infrastructure, so that it has supported a $>$ 10-fold increase in storage volume over the past seven years, while increasing in performance, maintaining a very high ($>$ 98% availability) and keeping costs low. In the following three subsections we provide justification for the continuation of these two posts, together with the award of modest quantity of pooled administrative support.

6.2 Technical Leadership and Project Management

In the last WFAU rolling grant case we argued for the establishment of a new WFAU Technical Lead (TL) role, in addition to the existing Project Management role, motivated by the realisation that the coming few years would see a radical increase in the importance of the computational side of WFAU's work. This has proven to be correct and this transformation continues apace: WFAU's data holdings currently total just over 0.5PB, while the programme proposed here would see that increase to nearly 2PB over the next three years; integrating WFAU archives into the developing VO requires both mastery of cutting-edge distributed computing technologies and authoritative participation in IVOA debates to steer standards definition in the right direction; and meeting the rising expectations of users for powerful, interactive archive interfaces needs not only familiarity with the latest web frameworks and toolkits, but also the judgement to select amongst them.

All this requires a very skilled and highly experienced IT professional, and WFAU has been lucky that in Keith Noddle it has such a person. Since his arrival in 2009 he has made a great contribution to WFAU's work: the prototype services described in Section 5.2 all bear his mark, as does a less visible overhaul in the software engineering practices followed in WFAU. This is based on the Redmine[1] tool, which supports planning and reporting with a small overhead, making it ideal for the modern agile development practices that appropriate for the way that WFAU works. Redmine integrates well with code repositories, continuous integration servers and has a native trouble-ticketing system. This means design, development and bug fixing are fully traceable, greatly assisting in maintaining high software quality values, and this will prove invaluable as the developments described in this proposal come to be rolled out into operational services.

Other aspects of the Project Management (PM) role continue to be busy, too. In addition to the usual PM activities of financial planning and reporting, staff appraisals, etc, two aspects of the WFAU science archive programme are particularly demanding of PM effort. The first is hardware procurement, where legal requirements necessitate a mini-tendering exercise for each purchase, which is very time-consuming, given the scale of the WFAU hardware procurement plan. Secondly, the range of astronomical and computational skills needed for different aspects of WFAU projects are to be found to varying degrees in different staff members, so we move staff between projects as needed to ensure that each has the correct expertise to hand at all stages. This ensures the most effective use of RA effort, but has a PM overhead associated with it, when staff are moving between a number of projects and it is necessary to balance the requirements and schedules of each: e.g. RAs covered by this proposal are also working part-time on the Gaia Post-Launch Support project and will contribute to WFAU's Euclid Science Data Centre (SDC) activities as these ramp up during the course of the new grant period.

At the last grant renewal, we were awarded a 0.75 FTE TL/PM post and we request the continuation of that post at the same level throughout the period of the new grant, to cover the range of responsibilities outlined above. We have costed this post as continuing to employ Keith Noddle, but it is possible that not all this work will be peformed by him: he has another role leading the Euclid SDC, but continuing uncertainty over the level of UK Space Agency support for the Euclid Science Ground Segment over the next few years means that it is unclear how much of Keith's time will be committed to that during the period of the new grant. We have an alternative plan for the provision of the necessary level of expertise should Keith not be available to fulfil this role at the full 0.75 FTE level.

6.3 Computing Systems Administration

At the moment WFAU science archive development and operations use a total of 38 servers. These are a mixture of NAS (Network-attached storage) boxes for bulk storage of image data, curation servers for generating advanced data products from the output of the VDFS nightly pipelines run by CASU, database servers for storing catalogues and image metadata, and webservers that support the user interfaces to the WFAU archives. The procurement plan, outlined in Appendix D, needed to support the proposed programme will add a further 17 servers, while increasing the disk storage from 0.5 PB to almost 2PB (the new machines will have larger capacity than the old ones), so there is a very significant sysadmin role to fill supporting the WFAU computing hardware.

In addition to simply managing the extant WFAU hardware Mark Holliman also undertakes R&D to ensure that WFAU continues to use the most appropriate hardware. In recent years this has led to the adoption of new technologies for networking, file and database storage, all introduced in a carefully-managed way, so that data centre operations have not been adversely affected, while the scale of WFAU data holdings have increased by more than a factor of ten. Holliman's current R&D activities focus on solid state drives (SSDs). These are currently much too expensive to be used for all WFAU's database storage needs, but their very high I/O performance, compared to standard hard disks, motivates consideration of their use for storing the most intensively-accessed parts of databases, such as index files.

As well as delivering a robust, scalable hardware platform that is relied upon to support 24/7 data centre operations and to provide data reliably to WFAU's 1000+ registered users, Holliman is also responsible for all of WFAU's VO services. He is currently in the process of moving all the core AstroGrid service instances from Leicester (where effort is no longer available) to Edinburgh, and, thanks to his efforts, WFAU is at the forefront of deployment of TAP services internationally. Holliman also represents WFAU at IVOA meetings, where his authoritative contributions, based on practical experience of running VO services, is always well

[1] `www.redmine.org`

received and is influential in setting the direction of VO standards discussions.

The last WFAU rolling grant included a 0.7 FTE system manager post. In practice, this work has occupied Holliman full-time during the period of the grant, and we request that his post continue to be funded at the 1 FTE level for the period of the new grant, to ensure that the WFAU programme continues to be supported by computational infrastructure of the necessary scale and resilience. We note that the AGP guideline figure for this proposal would be about 0.8 FTE of computing officer effort, but few, if any, research groups operate computing systems of the scale and complexity as WFAU's nor that are relied upon by such a large user community, so a full-time post is entirely justified.

6.4 Administrative Support

The AGP guidance implies a guideline figure of 0.6 FTE of administrative support effort, given the FTE level of RAs and applicants on this proposal. However, we are only requesting 0.2 FTE of Directly Allocated pooled administrative effort in support of our proposed programme. This will primarily cover travel administration and HR activities, since the substantial computing hardware procurement programme will be undertaken by the TL/PM.

6.5 RESOURCES REQUESTED

(a) *Staff.* 1.0 FTE (Holliman) and 0.75 FTE (Noddle) throughout Q2 2013 to Q1 2016, plus 0.2 FTE pooled administrative support
(b) *Travel and subsistence.* annual AGP guideline amount of £2K per FTE
(c) *Personal computing:* annual AGP guideline amount of £2K per FTE
(d) *Equipment & maintenance:* N/A

Appendix A: Summary Publication Statistics

The following table lists the ADS publication statistics for the applicants for the calendar years 2007-2011 inclusive. In total, the five applicants published 118 refereed papers during that five year period and these have received a total of 4624 citations. N.B. Lawrence was Head of School during the period 2003-2008.

Name	Total refereed papers	First-author papers	Citations
Ferguson, A.M.N.	36	0	1073
Hambly, N.C.	42	1	1871
Lawrence, A.	17	2	1425
Mann, R.G.	34	0	1723
Noddle, K.T.	0	0	0

Annette Ferguson was appointed to the lecturing staff at the IfA in 2005 and was promoted to Readership in 2007. Her award of a Marie Curie Excellence Grant (2006-2010) enabled a new research group to be built at the IfA, specialising in galaxy archaeology. She co-led the INT Wide-Field Camera Surveys of M31 and M33 and is a key Co-I in the follow-on PAndAS CFHT/Megacam Survey, the analyses of which have demonstrated some of the most compelling evidence yet for ongoing galaxy assembly in the Local Group. She is on the Steering Group for Gaia-ESO, a 300 night VLT/FLAMES Public Spectroscopy Survey, and is actively involved in the Gaia GREAT ESF and ITN networks. She serves on a number of national and international committees, including the ESO OPC, the Astronet Wide-Field Spectroscopy Working Group and the ESO Public Surveys Panel. She was awarded the Annie Jump Cannon Award from the AAS in 2003 and a Caroline Herschel Distinguished Visitorship at STScI in 2011.

Nigel Hambly joined the IfA in 1999 as a research fellow and founder member of the Wide Field Astronomy Unit. Since then, he has undertaken a leading role in project work for the legacy Schmidt all–sky SuperCOSMOS Sky Surveys, the state–of–the–art current generation of infrared sky surveys for UKIRT/WFCAM (UKIDSS) and VISTA as chief architect of the VISTA Data Flow System Science Archives, and latterly as a key member of the Data Processing and Analysis Consortium for the ESA Cornerstone astrometric sky survey mission Gaia. Since joining the IfA, his project–related research work has generated 85 refereed papers with nearly 3000 citations.

Andy Lawrence is the Regius Professor of Astronomy at Edinburgh, with a strong track record in observational cosmology, survey astronomy, and the phenomenology of active galaxies. He has 31 papers in the > 100 citation class. He is particularly well known for the receding torus model of AGN unification, and for defining and leading the very successful UKIDSS survey, which has generated 349 refereed papers and over 7500 citations so far, with the rate per year still increasing. He is also a leader of the world-wide Virtual Observatory initiative, and has been PI of several related projects and chair of the International Virtual Observatory Alliance.

Bob Mann was appointed to a lectureship in the IfA in 2004 and promoted to Senior Lecturer in 2008. The scientific leadership of WFAU has been his principal research activity for more than five years, with a particular focus on the novel application of computer science techniques and e-Science technologies in survey astronomy. He was a Research Leader at the National e-Science Centre and led the ‘Next Generation Sky Surveys’ Theme at the e-Science Institute, which will be reported upon in a forthcoming special issue of ‘New Astronomy Reviews’. He is a Planck Scientist, a member of the Editorial Board of the International Journal of Digital Curation and one of the two Coordinating Editors of *Astronomy and Computing*, a new journal to be launched in late 2012. His 76 refereed journal papers have yielded $>$5400 ADS citations.

Keith Noddle joined the Wide Field Astronomy Unit in 2009 as Technical Lead/Project Manager. He has introduced industry standard software development practices that are helping deliver improved scientific access to WFAU-held archives and beyond. He leads the UK Science Data Centre for the Euclid Dark Energy space mission and is an active core member of the Euclid System Team, helping to develop the Ground Segment architecture and working practices. He is also a member of EuroVO management team, regularly organising and chairing European-wide Virtual Observatory conferences.

Appendix B: Summary of STFC astronomy grants held at the IfA

The following table lists the STFC grants currently held at the Institute for Astronomy on 01/07/12.

STFC Ref.	Scheme	Title	Start Date	End Date	Principal Investigator	Funded Posts	Amount
ST/H00047X/1	Standard	Wide Field Astronomy Equipment	01/04/09	31/03/14	Dr R. Mann		£235,737
ST/H000496/1	Standard	Wide Field Astronomy	01/06/09	31/03/14	Dr R. Mann	1 PI, 3 Co-Is, 9 PDRA	£1,994,582
ST/G007039/1	Advanced Fellowship	Exploring the epoch of first light and the cosmic history of galaxies	01/10/09	30/09/14	Dr M. Cirasuolo	1 fellow	£406,184
ST/J001422/1	Standard	Astronomy and Astrophysics at Edinburgh	01/04/12	31/03/15	Prof J. Peacock	1 PI, 19 Co-I, 8 PDRA	£2,998,045

Appendix C: Staff Gantt Chart

We present below in Figure 9.1 a staff Gantt chart in a format close to that specified in the AGP guidelines, and, in Figure 6.5 a consistent one in a possibly more comprehensible format.

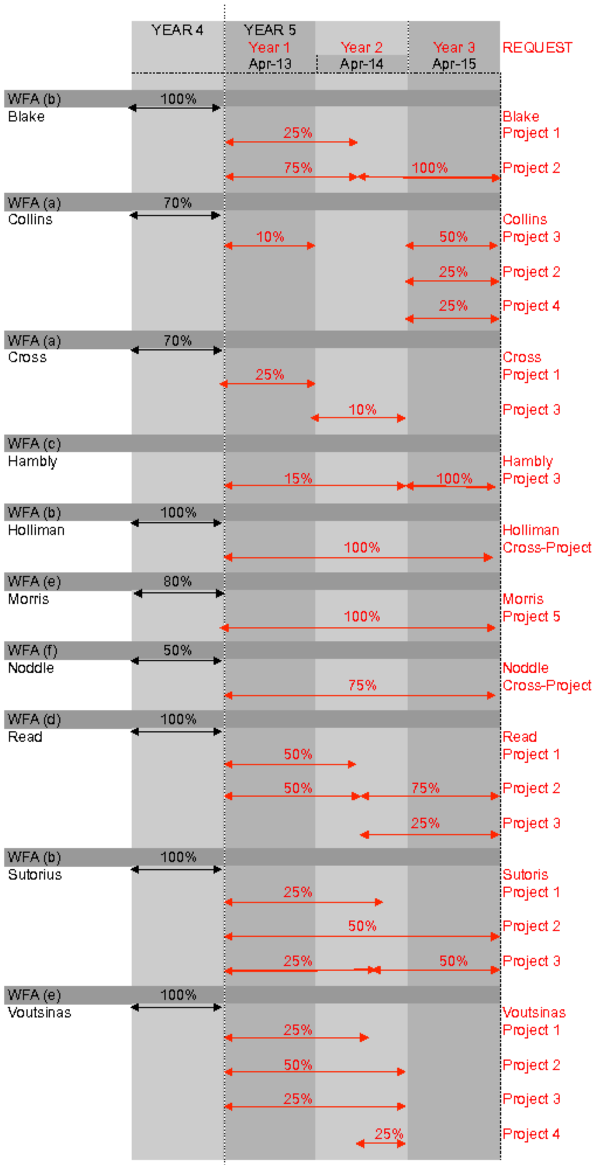

Figure 9.1: Staff Gantt chart. Projects on the old WFAU Rolling Grant do not match the projects proposed here exactly, and the fractions of staff on different projects varied significantly over time, so the Year 4 column records the total effort from each member of staff on that grant.

	Calendar year	2012	2013			2014				2015						Total staff-years by Project					
	Quarter		Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1		1	2	3	4	5	
	Grant Year	*(Year 4)*	Year 1 (2013/14)				Year 2 (2014/15)				Year 3 (2015/16)					WFCAM	VISTA	Gaia	PS1	FAI	CPR
Hambly	3 Gaia Exploitation	0.00	0.15								1							1.3			
Blake	1 WFCAM Exploitation	1.00	0.25													0.375					
	2 VISTA Exploitation		0.75						1								2.625				
Collins	2 VISTA Exploitation	0.70	0.25								0.25						0.625				
	3 Gaia Exploitation		0.1				0.25		0.5									0.975			
	4 Pan-STARRS1 Exploitation										0.25								0.25		
Cross	1 WFCAM Exploitation	0.70	0.25													0.25					
	3 Gaia Exploitation						0.1											0.1			
Holliman	Cross-Project Resources	1.00	1																		3
Morris	5 Future Archive Infrastructure	0.80	1																	3	
Noddle	Cross-Project Resources	0.50	0.75																		2.25
Read	1 WFCAM Exploitation	1.00	0.5													0.75					
	2 VISTA Exploitation		0.5						0.75								1.875				
	3 Gaia Exploitation								0.25									0.375			
Sutorius	1 WFCAM Exploitation	1.00	0.25													0.375					
	2 VISTA Exploitation		0.5														1.5				
	3 Gaia Exploitation		0.25						0.5									1.125			
Voutsinas	2 VISTA Exploitation	1.00	0.75						0.5								1.375				
	3 Gaia Exploitation		0.25															0.5			
	4 Pan-STARRS1 Exploitation								0.25										0.125		
																1.75	8	4.375	0.375	3	5.25

Figure 9.2: The current and proposed disposition of the WFAU staff mentioned in this proposal: the projects on the old WFAU Rolling Grant to do not match the projects proposed here exactly and the fractions of staff on different projects varied significantly over time, so the figures for 2011/12 are not completely straightforward to interpret. For example, Noddle is listed as 0.5 FTE on the fomer WFAU rolling grant for 2012/13, but his average level on that grant since his appointment is 0.7 FTE. Hambly is covered at the 0.15 FTE level on the IfA Consolidated Grant for three years from 1 April 2012.

Appendix D: Resource Summary

Resource Summary Table

Grant year	**01/04/13**	**01/04/14**	**01/04/15**
Staff FTE	**%FTE**	**%FTE**	**%FTE**
Project 1 : Enabling WFCAM exploitation			
1.1 M.A.Read	50	50 (6)	-
1.2 E.T.W.Sutorius	25	25 (6)	-
1.3 R.P.Blake	25	25 (6)	-
1.4 N.J.G.Cross	25	-	-
Project 2 : Enabling VISTA exploitation			
2.1 M. A. Read	50	50(6) / 75 (6)	75
2.2 E. T. W. Sutorius	50	50	50
2.3 R. P. Blake	75	75 (6) / 100 (6)	100
2.4 R. S. Collins	25	25(6)	25
2.5 S. Voutsinas	75	75(6) / 50 (6)	-
Project 3 : Enabling Gaia Exploitation			
3.1 N. C. Hambly	15	15	100
3.2 R. S.Collins	10	25(6)/50(6)	50
3.3 E. T. W. Sutorius	25	25 (6) / 50 (6)	50
3.4 S. Voutsinas	25	25	-
3.5 M.A.Read	-	25 (6)	25
3.6 N.J.G.Cross	-	10	-
Project 4 : Enabling PS1 Exploitation			
4.1 R. S. Collins	-	-	25
4.2 S. Voutsinas	-	25(6)	-
Project 5 : Future archive infrastructure			
5.1 D. Morris	100	100	100
Cross-Project Resources			
6.1 K. T. Noddle	75	75	75
6.2 M. S. Holliman	100	100	100
Applicant FTE			
R.G.Mann	25	25	25
A.Lawrence	10	10	10
A.Fergusson	10	10	10
Total Applicant FTE	*0.45*	*0.45*	*0.45*
Total PDRA FTE	*7.50*	*7.50*	*7.75*
Total Research FTE	*7.95*	*7.95*	*8.2*
Other Directly Allocated			
Pooled Admin Support	20	20	20

Resource Summary Table (cont'd)

Grant year	**01/04/13**	**01/04/14**	**01/04/15**
Cost			
Infrastructure Technicians (3 yrs - £27040)			
Travel & Subsistence Total (3 yrs £66500)			
Academic conf. Travel (all projects, £2k/FTE)	£16,000.00	£16,000.00	£16,000.00
Other Directly Incurred Total (3 yrs -£282217)			
Laptops / PCs (all projects, £2k/FTE)	£16,000.00	£16,000.00	£16,000.00
Computing hardware (see Purchasing Plan below)	£87,966.30	£115,355.60	£30,894.80
Exceptions Total (3 yrs £0)			

Notes

We have used AGP guideline figures of £2k per FTE per annum for Travel & Subsistence and for personal computing, taking a value of 8 FTE to get round numbers. For the sake of clarity we have not subdivided these by Project.

Computer Purchasing Plan

WFAU receive data at a predictable rate. We have modeled this rate on a quarter-by-quarter basis, using the per-project figures in the case and, in Table 10.1 below present a proposed hardware purchase schedule that matches this rate: *N.B.* we have allowed an additional $\sim$ 10% capacity for spare disks, enabling us to maintain the required level of data availability in the archive, and have also costed for the LTO-5 tapes that we use for data back-up. From this we have determined what volume of pixel and database storage we require in a given quarter and have scheduled computer purchases and upgrades so that we track the required capacity. This has the advantage that, by delaying a procurement until it is needed, we can benefit from the 'Moore's Law'-like decrease in hardware prices over time. The translation of capacity into estimated cost is made using recent quotes from our usual supplier, as follows:

- **Database cluster node**. 9 TB disk (formatted), 64 core 2.6 GHz CPU, 96 GB RAM, 10Gbps networking. £ 9,200, plus £ 450 per annum maintenance.
- **Network-attached storage (NAS) ndoe**. 20 TB disk (formatted), ethernet networking, £ 9,765, plus £ 800 per annum maintenance.
- **Disk upgrades**. NAS upgrade: 10 $\times$ 3TB 7200RPM 6 Gbps SATA disk - £ 2,670. DB upgrade: 6 $\times$ 3TB 7200RPM 6 Gbps SATA disk - £ 1,602. Spare disk set: 3 $\times$ 3TB 7200RPM 6 Gbps SATA disk - £ 801.

With these costs, and the purchase schedule shown in Table 10.1, we obtain (after merging a few of the smaller orders) the set of hardware orders listed in the *Other DI costs* section of our Je-S form.

	2013			2014				2015				2016	
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Total
Pixel	40.19	80.38	165.57	250.76	340.24	429.72	519.20	608.68	673.16	737.64	802.12	866.6	866.6 7
Catalogue	12.57	25.14	42.71	60.28	77.85	95.42	266.895	433.37	447.345	461.32	480.295	494.27	494.27
Purchase Schedule													
NAS	2			1	1	1		1		1			
NAS Disk Upgrade		2	3	3	3	3	3	3	2	2	2	2	
DB Node	2					1	3	3	1				
DB Disk Upgrade		1	1	2	1	1							
Processing Node					1								
Spare Disk	4	2	3	3	4	3	4	5	2	2	2	2	
NAS Capacity	40	90	165	260	355	450	525	620	670	740	790	840	840
No. of NAS	2	2	2	3	4	5	5	6	6	7	7	7	7
DB Capacity	18	32	46	74	88	153	306	459	510	510	510	510	510
No. of Nodes	2	2	2	2	2	3	6	9	10	10	10	10	10
Moore's Factor	0.8	0.8	0.7	0.7	0.6	0.6	0.5	0.5	0.4	0.4	0.3	0.3	
Cost	£33,003	£6,835	£8,411	£16,367	£18,361	£21,723	£27,245	£32,528	£8,547	£6,683	£2,083	£2,083	£183,867

Table 10.1: Defining the hardware purchase schedule

Appendix E: Letters of Support

We enclose letters from the following 14 leaders of the WFAU user community in support of the four science exploitation projects proposed in this Case:

1. Prof Philip Best, Project Leader, LOFAR-UK (*Project 4*)
2. Dr Maria-Rosa Cioni, PI, VMC (*Project 3*)
3. Prof Andrew Collier Cameron, UK PI, HARPS-North Consortium (*Project 3*)
4. Prof Gary Davis, Director, Joint Astronomy Centre (*Project 1*)
5. Prof Jim Dunlop, Co-PI, UltraVISTA (*Project 2*)
6. Dr Simon Dye, Project Scientist, UKIRT Hemisphere Survey (*Project 1*)
7. Prof Jim Emerson, Director, VISTA Consortium (*Project 2*)
8. Prof Gerry Gilmore, UK PI, Gaia data processing and analysis activity (*Project 3*)
9. Dr Matt Jarvis, PI, VIDEO (*Project 2*)
10. Dr Phil Lucas, PI, UKIDSS GPS and Co-PI, VVV (*Projects 1 & 2*)
11. Dr Richard McMahon, PI, VHS (*Project 2*)
12. Prof Tom Shanks, PI, VST ATLAS (*Project 2*)
13. Dr Will Sutherland, PI, VIKING (*Project 2*)
14. Prof Steve Warren, Project Scientist, UKIDSS (*Project 1*)

Institute for Astronomy
Blackford Hill
Edinburgh EH9 3HJ
email: pnb@roe.ac.uk
tel: +44-31-6688358

July 4, 2012

To whom it may concern

LOFAR-UK support letter for WFAU grant

I am writing in my role as Principal Investigator of the LOFAR-UK project, with regards to the grant application being submitted by the Wide Field Astronomy Unit (WFAU) in Edinburgh. Part of that grant application involves the cross-matching of LOFAR Survey data with the Pan-STARRS1 dataset, which will produce a very useful tool for UK (and worldwide) astronomers.

LOFAR is a next-generation software-driven low-frequency radio interferometer, operating between 30 and 240 MHz. It is composed of 48 "stations" of receivers, spread around Europe, the signals from which are digitised on-station as phased-arrays and transported through optical fibres to a central digital processor. The core of the array is located in the Netherlands, and most of the stations are located there, but additional international stations are located in each of Germany, France, Sweden and the UK. The UK station is located at the Chilbolton Observarory, and is funded and operated through a collaboration between LOFAR-UK (a consortium of over 20 Universities and Research Institutes) and STFC. The LOFAR array is now largely installed and, although commissioning work is on-going, a Cycle-0 open call for proposals has recently been issued.

LOFAR will be a powerful instrument for a wide range of scientific topics, and a large fraction of its resources in the first 5 years of operation (2013-2017) will be focussed on 6 key science projects. The largest of these, and the one with the most UK scientific involvement, is the LOFAR Surveys Key Science Project; this will carry out large-scale multi-frequency radio surveys of the entire northern sky, much deeper than any existing large-area radio survey. Maximal exploitation of these surveys will only be possible through multi-wavelength cross-identifications of the detected radio sources to identify and characterise the host galaxies. The Pan-STARRS1 survey will be the deepest optical survey over the sky area mapped by the LOFAR surveys, and will be publically available on the relevant timescale.

The WFAU proposal to provide a user-friendly catalogue archive to the LOFAR Survey dataset, coupled with cross-neighbour tables to match sources between this and the Pan-STARRS dataset will be a very valuable tool to the many tens of astronomers across the UK involved in work on the LOFAR surveys. The aim to expand this to allow users to run their own cross-matching algorithms without the need to download data is also very welcome: this is an ambitious aim, but is a development that will be absolutely essential for the ever-larger radio surveys that will be produced by the Square Kilometer Array (SKA) pathfinder arrays, and ultimately the SKA itself.

On behalf of LOFAR-UK, I am very happy to give full support to this proposal.

Yours,

Professor Philip Best
PI of LOFAR-UK

Dr Maria-Rosa Cioni (PI of VMC)
University of Hertfordshire, Physics Astronomy and Mathematics
College Lane, Hatfield AL10 9AB
m.cioni@herts.ac.uk

July 4, 2012

Dear member of the evaluation panel,

with this letter I would like to support the application for the renewal of the STFC grant from the Wide Field Astronomy Unit (WFAU) in Edinburgh for "Enabling VISTA exploitation" during the period April 2013 - March 2016.

The operational support provided by WFAU to the VISTA survey of the Magellanic Cloud system (VMC - PI Cioni) is an essential component of the data processing and availability of the data both to the VMC team and to the world wide community. In particular, WFAU is in charge of the stacking of deep VISTA tile images and the extraction of the point sources as well as the preparation of the final data producccts for their delivery to ESO according to a specific timeline.

The promptness and excellent quality of the WFAU work has allowed the VMC team to make immediate use of the VMC data. This has resulted in six scientific publications in refereed journals of which one currently under review by a referee. These studies have addressed planetary nebulae, pulsating variable stars, the star formation history and quasars behind the Magellanic system as well as presenting the VMC survey. All but one have made use of the VMC data provided by WFAU via the VISTA Science Archive (VSA).

The VMC survey is at present 21% complete and during next three years the completion rate is expected to reach $\sim 60\%$. This data, which represents a considerable fraction of the VMC survey, will be timely exploited only after the WFAU analysis and archival steps are executed. The services provided by WFAU to the VMC survey team include the positional match between VMC sources and external catalogues. This multi-wavelength aspect of the VMC data analysis is necessary for various investigations, such as the proper motion (in combination with 2MASS data).

The WAFU support is therefore essential for a successful exploitation of the VMC survey data as provided at present but also for its continuing development. Additional modes of querying the data, for example, something analogous to the SDSS 'MyDB' system, represent a key step in the growth of the VSA that would meet the expectations of large database users, enabling them to do their science more effectively.

Please do not hesitate to ask me for further information where necessary.

Sincerely

Maria-Rosa Cioni

University of St Andrews
from first to foremost

600 YEARS
1413 – 2013

Professor Andrew Collier Cameron, FRSE
School of Physics and Astronomy

STFC
Polaris House
Swindon

3 July 2012

To whom it may concern,

Letter of Support for WFAU renewal proposal

I am writing as UK PI of the HARPS-N Collaboration in support of the "Enabling Gaia Exploitation" line within the WFAU grant renewal proposal, through which we hope to see an archive deployed to facilitate exploitation of the unique HARPS-N dataset by several UK Universities.

HARPS-N is a high-precision radial-velocity spectrometer, partly funded by the Universities of Edinburgh, St Andrews (both with SUPA-II input) and Queen's Belfast, and partly built at UKATC. It was commissioned on the 3.5-m Telescopio Nazionale Galileo on La Palma, in spring 2012. In May 2012 the HARPS-N Science Team began science operations on a five-year, 80N/year guaranteed-time programme to characterise exoplanets detected by NASA's Kepler satellite and to detect low-mass exoplanets around nearby bright stars. The HARPS-N Science Team has also just closed negotiations on an EU FP7-SPACE proposal "ETAEARTH" that will be exploiting synergies with both Kepler and with Gaia, with co-investigators at Edinburgh, St Andrews, Queen's Belfast, and Warwick.

Whilst the data volumes involved in our project are very modest by WFAU's usual standards (less than 10 TB over 5 years), the provision of a robust archive capable of supporting the scientific exploitation of the data by the HARPS-N Collaboration, and, potentially, subsequent public release, is not a trivial task, hence my request that WFAU provide this service for us. I gather that this can be done for minimal cost, given the investment that WFAU is making in developing the archive for the much larger Gaia-ESO Spectroscopic Survey. It is good to see that development effort expended on one project can benefit another in such a cost-effective manner, through the careful design of generic archive infrastructure.

With best regards,

Yours sincerely

A. Collier C.

Andrew Collier Cameron
(UK Co-PI, HARPS-North Collaboration)

North Haugh, St Andrews, Fife KY16 9SS, Scotland
T: +44 (0)1334 463147 F: +44 (0)1334 463104
E: acc4@st-andrews.ac.uk W: star-www.st-and.ac.uk/~acc4/

Joint Astronomy Centre

660 N. A'ohōkū Place, University Park
Hilo, Hawaii 96720-2700, USA

http://www.jach.hawaii.edu/

Tel: +1 808 961 3756 (Central)
Fax: +1 808 961 6516
+1 808 969 6591 (Directorate Office)

Professor Gary R. Davis *BSc MSc DPhil FRAS AMRI DSc(Hon.)*
Director, Joint Astronomy Centre
James Clerk Maxwell Telescope
United Kingdom Infrared Telescope

Tel: +1 808 969 6504 (Direct)
Email: g.davis@jach.hawaii.edu

Dr R G Mann,
IfA Wide-Field Astronomy Unit,
by email.

21 June 2012

Dear Bob,

I write to confirm my support for your proposal to extend WFAU operations. UKIRT is now the most productive telescope in the world, and UKIDSS is one of the most successful astronomical surveys ever undertaken. These two statements speak to the high standard of UK astronomy generally, and of UKIRT and UKIDSS specifically, and neither of them would have been possible without the operational support provided by WFAU.

Specifically, all UKIRT data are now processed at CASU and archived at WFAU, making these two units integral parts of UKIRT's operational system. As long as WFCAM on UKIRT is collecting data, the services of these two units will be essential. I believe it is now recognised within STFC that the funding of CASU and WFAU must be tied to the facilties they are supporting. With STFC's recent decision to cease support for UKIRT operations on 30th September 2013, it is entirely appropriate for WFAU's operational support to be extended for some period beyond this to ensure that the final night's data are fully captured.

My thanks to you for the excellent service you provide to UKIRT's user community, and I trust this support will ensure that your resources are appropriately extended.

Best regards,

Professor Gary Davis,
Director JAC.

INSTITUTE for ASTRONOMY

School of Physics & Astronomy
University of Edinburgh
Royal Observatory
Edinburgh
EH9 3HJ

26th June 2012

To whom it may concern

I write as the ESO PI of the UltraVISTA survey in support of the proposal by the Wide Field Astronomy Unit (WFAU) at Edinburgh to expand their delivery of VISTA survey data products to include UltraVISTA and the key associated optical and (ultimately) mid-infrared datasets.

First, while UltraVISTA, as the smallest and deepest of the VISTA surveys, is delivering final Y, J, H, K_s images to ESO which can, in principle, be downloaded in their entirety by interested users, the reality is that many people interested in UltraVISTA (and in the COSMOS field in general) would prefer to be able to interrogate the images from a catalogue perspective, and then download the relevant (and much more manageable) postage stamp sub-images of the regions of interest. The WFAU already has proven expertise in this area, and this is not a service that the ESO archive will ever provide.

Second, ESO has no intention of serving the associated optical and Spitzer data in the COSMOS field which, for much research, needs to carefully combined with the UltraVISTA near-infrared imaging (e.g. for the derivation of complete spectral energy distributions, photometric redshifts etc). Most of the associated CFHT and Subaru and Spitzer data is now public, and can be downloaded from individual telescope archives, or in some cases from the COSMOS website, but there is no single, properly validated source for the required multi-frequency data. Here at Edinburgh we are already in control of all the key supporting datasets from a research perspective, but are not in any position to serve this to the public in an effective and efficient manner. Clearly this is an area in which the WFAU can apply their expertise to offer enormous added value for those interested in using UltraVISTA.

Thirdly, in collaboration with groups at Rome and Paris, I have recently secured FP7 funding specifically directed at comparing, improving and developing methods used to deconfuse Spitzer IRAC data for effective combination with the higher-resolution optical–near-infrared data delivered by the ground-based survey telescopes. We have the funding for the development work, but not to then serve the resulting data products to the community. By ultimately incorporating the Spitzer data (S-COSMOS and SEDS) available within the UltraVISTA field, WFAU could deliver a uniquely-powerful multi-frequency interface for the (large number of) astronomers within the UK and across the world who are already showing huge interest in UltraVISTA.

Yours sincerely

Professor James Dunlop *BSc Hons, PhD, FRSE, FInst Phys*

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

School of Physics & Astronomy
University park
Nottingham
NG7 2RD

Colin Vincent,
Polaris House,
North Star Avenue,
Swindon SN2 1SZ

Monday June 25th 2012

Dear Dr Vincent,

Re. Funding of archiving operations for the UKIRT Hemisphere Survey

I write on behalf of the UKIRT Hemisphere Survey (UHS) Consortium in strong support of the grant application by the Wide Field Astronomical Unit at the Royal Observatory of Edinburgh to maintain archiving operations of UKIRT data until September 2014.

The UHS is an ambitious extension of the hugely successful UKIRT Infra-red Deep Sky Survey (UKIDSS). The survey will provide invaluable data covering almost the entire northern hemisphere at infra-red wavelengths. A significant fraction of the success of UKIDSS is due to the archiving facilities operated by WFAU which provides a crucial conduit for the mining and dissemination of data products within the astronomical community. So that the UHS enjoys the same high impact and legacy value as that experienced by UKIDSS, it is crucial that WFAU continues its role in providing a platform for the access of UHS data.

Specific reasons for continued support from WFAU include:

1. The database, data ingest procedures and user interface have been honed over many years with UKIDSS data, culminating in a highly efficient facility and one that users are familiar with. UHS is an extension of UKIDSS and as such its data products will adhere to exactly the same protocols. The archive operated by WFAU is therefore immediately ready for ingest and staging of UHS data.

2. In addition to providing archive facilities, WFAU have developed a series of quality control tests for UKIDSS which are applied to the data prior to release. These tests are well established, rely on the existing data base set up and are directly applicable to the UHS data.

3. Of paramount importance is that UKIDSS data are merged with the new UHS data so that the user experiences one complete product, rather than having to search separate databases. In addition to WFAU already hosting the UKIDSS data, the unit has the necessary expertise and experience to provide this facility.

The UHS will run from late June 2012 until the end of September 2013. After surveying has completed, data processing in Cambridge will continue for a further six months. After this, a further six months will be required by WFAU to allow for data ingest, quality control, final release and minor software development to further enhance the archive user interface. This takes the total request for funding for WFAU up until the end of September 2014.

Kind Regards,

Dr Simon Dye
UHS Project Scientist

From: Prof. J .P. Emerson,
Professor of Astrophysics
Tel: +44 (0)20 7822 5040
Mobile: +44 (0)794 127 1548
Fax: +44 (0)20 8981 3522
Email: j.p.emerson@qmul.ac.uk

QUEEN MARY
UNIVERSITY OF LONDON

Astronomy Unit
School of Physics & Astronomy
Mile End Road
LONDON
E1 4NS

To whom it may concern

27 June 2012

Dear STFC grants assessors,

Re: WFAU and VISTA Science Archive

I understand that the Wide Field Astronomy Unit's (WFAU) grant covering their work on the VISTA Science Archive (VSA) is coming up for renewal and so am writing to be sure that the Panel appreciates the importance of continuing to support this essential work.

VISTA dominates the data flow volume out of from ESO's Cerro Paranal Observatory and the VISTA Data Flow System (VDFS) consisting of a pipeline at the Cambridge Astronomical Survey Unit (CASU) and a science archive at WFAU was designed and implemented to handle this, and also WFCAM data. Both components (pipeline and archive) are essential and both were shown to work well with WFCAM data before being hit, some years later than originally expected, with the deluge of VISTA data. They have survived this well, with necessary adaptations, and are providing the only practical way to efficiently exploit VISTA's large surveys. The VISTA survey teams (all led by UK PIs or co-PIs) rely on VSA, especially for the surveys with the larger areas, for their data products and for the ability to efficiently extract new science from the vast databases. It is crucial for exploiting the investment in VISTA and turning it into science that support for the VSAs work be continued at the level that allows it to continue to operate effectively as net data volume continues to increase, and to address the evolving needs of users and respond to ESO's requirement on data products.

The panel should be aware that although PIs are required to submit data products (mostly from VSA) to ESO for the ESO archive this does not offer the capabilities to do any thing with the data other than search and download it. Thus VSA is essential to continued science exploitation of VISTA data, and, in accordance with the assurances given by PPARC when the VISTA project was started, should continue to be well enough funded to enable to community to continue to extract science efficiently, and to improve its ability to respond to reasonable new user requirements which arise as the data itself throws up new questions.

Yours Sincerely

Jim Emerson (VISTA Consortium Director)

Professor Gerard Gilmore
Professor of Experimental Philosophy

Dr Bob Mann
WFAU Edinburgh

UNIVERSITY OF CAMBRIDGE

28 June 2012

Institute of Astronomy

Dear Bob,

You ask that I clarify the intended role of WFAU in both the current Gaia-ESO Survey, and in the on-going UK Space Agency Supported Gaia UK Data Processing and Analysis efforts, with specific reference to those aspects which should be supported by STFC.

I write as co-PI for the Gaia-ESO Survey, with lead responsibility for some 60% of the total project, and as UK PI for the Gaia UK Data Processing Consortium work.

The Gaia-ESO Survey is a 300night VLT+FLAMES Spectroscopic Public Survey targetting a representative sample of some 100,000 stars across all Galactic Stellar Populations, with high-quality high-resolution spectroscopy for every star. This is by a long margin the largest spectroscopic Survey ever attempted on a large telescope - the telescope time allocation alone has notional equivalent value some 30Meuro. The Survey started in January 2012, is well underway, and is going well. The Survey plan requires that WFAU Edinburgh has sole responsibility for archiving all data, has sole responsibility for release of all science-verification-ready data to the over 300 members of the Survey Consortium for verification analyses, and has sole responsibility for long-term archival access to the data, with associated enhanced-value products and data-query tools, for the international community. That is, WFAU has earned the resposnsibility to become the single international data portal for what is ESO's (and the world's) largest large-telescope Survey, leading into the Gaia era. This work requires STFC support, and will place the UK in a global-leading position as we enter the decade of Gaia.

The UK Gaia Data Processing Efforts over the decade following Gaia launch (anticipated in September 2013) have two parts. The first is core mission data processing. This will happen at the Cambridge Gaia Data Centre, and is UKSA funded. The second, which will be initiated by an ESA Announcement of Opportunity for the so-called "CU9" activities, is anticipated in mid 2013, and covers the efforts required to release the processed Gaia data for science analysis. This science exploitation work is to be funded by STFC. For this data-release/archiving work, WFAU is anticipated to play a leading European-scale role, through a European-wide consortium currently being formed. WFAU has ideal qualifications to lead the UK activity in this science-critical aspect of Gaia, and when supported will ensure a prominent leading role for the UK is the paradigm-changing Gaia mission science releases.

As PI of the two relevant programmes I am fully aware of the WFAU plan, obligations, and ambitions. All are fully consistent with wider-level plans. All wil allow leading UK presence in high-impact science. All are proposed at a sensible level, based on proven expertise.

STFC support for these activities has the strongest support from the PI-teams which I lead.

Yours sincerely,

Gerry Gilmore

Gerry Gilmore
UK PI, Gaia data processing and analysis activity

ioa
Madingley Road
Cambridge CB3 0HA.

Telephone: +44 (0)1223 337506
Fax: +44 (0)1223 339910
E-mail: gil@ast.cam.ac.uk

9

University *of* Hertfordshire

Dr Matt J. Jarvis
Centre for Astrophysics Research
Science & Technology Research Institute
University of Hertfordshire
College Lane
Hatfield AL10 9AB
UK

Telephone: +44-1707-285251
E-mail: *m.j.jarvis@herts.ac.uk*

20th June 2012

To whom it may concern,

As the principal investigator of the VISTA Deep Extragalactic Observations (VIDEO) Survey I would like to emphasize the key role that WFAU occupies for this survey. Although the ESO Science Archive provides a data base for the VISTA Surveys this archive is almost impossible to access in an efficient way for the users. As such this causes a considerable hindrance for the full scientific exploitation of the VISTA survey products. The VIDEO Survey team therefore relies on WFAU to provide a readily accessible data base that users from across the global astronomical community can access, using similar tools as already implemented for the highly successful UKIDSS. These not only allow access to the pixel data but also value added information which allows the users to link the VIDEO survey data to imaging data taken at other wavelengths, most critically for VIDEO, from *Spitzer* and ground-based optical telescopes.

The VIDEO team also relies on WFAU to liaise with the ESO archive team to ensure that the VIDEO data and catalogues are fully consistent with the ESO requirements and to send the formated tables and imaging to the ESO archive within the ESO timelines. If WFAU were unable to do this work it would add a significant burden to the STFC-funded VIDEO PDRA whose job already consists of performing the data quality control, deep stacking of the individual VISTA paw prints, catalogue extraction, and combining with other data sets to obtain photometric redshift information.

On behalf of the VIDEO survey team, I therefore fully support the WFAU's grant application.

Yours faithfully

M. Jarvis

Dr Matt J. Jarvis

Letter of Support for STFC grant application by the Wide Field Astronomy Unit

To whom it may concern,

I am writing as the head of the UKIDSS Galactic Plane Survey (GPS) and the co-PI of the VISTA Variables in the Via Lactea (VVV) ESO public survey to express my support for the current STFC grant application by WFAU. Specifically I am supporting the funding request for the "Enabling WFCAM exploitation" and "Enabling VISTA exploitation"

The work that WFAU does is essential to the ongoing production of data releases for the GPS and VVV. It is important to note that only 51% of GPS data has been released at the time of writing and considerable effort is still required from WFAU not only to provide the remaining 49% but also to archive profile-fitting photometry (which doubles the source counts in the more crowded fields) and to produce and archive difference images from 2 epoch data. These difference images are essential for variability and proper motion work. These work packages are detailed in WFAU's request for support.

Moreover, VVV is at a very early stage with respect to data releases. The First Data Release has been produced, using a sophisticated SQL database model that enables rapid and flexible interrogation of the variablility data. However, the huge volume of synoptic data in crowded Galactic bulge fields is severely taxing the existing systems, therefore requiring extensive development work by WFAU to speed up future releases. WFAU have agreed to archive new data products supplied by the VVV team and CASU, specifically profile fitting photometry and the results from a precise time series photometry algorithm that are essential to the survey goals, provided they have sufficient resources to do so.

Yours faithfully,

P. Lucas

Philip Lucas,
Reader in Astrophysics,
Centre for Astrophysics Research,
University of Hertforshire

Dr Richard G McMahon
Reader in Observational Astronomy
www.ast.cam.ac.uk/~rgm

UNIVERSITY OF CAMBRIDGE

Institute of Astronomy

Phone: 00-44-1223-337519 (direct)
Email: rgm@ast.cam.ac.uk

1-Jul-2012

Dear Bob,

I am writing I am writing in my capacity as the PI of the VISTA Hemisphere Survey (VHS) to strongly support your STFC grant application for enabling VISTA exploitation.

VHS is the highest ranked ESO Public Survey on VISTA and is the flagship all hemisphere near IR survey in the southern celestial sphere. VHS has been initially awarded around 300 clear nights over a 5-year period on VISTA. When completed, VHS will cover ~18,000deg^2 and will result in coverage of the whole southern celestial hemisphere (20,000deg^2) to a depth 30 times fainter than 2MASS/DENIS in at least two wavebands (J and K_s) to median 5σ point source limits (Vega mags) of J = 20.2 and K_S = 18.1. In the South Galactic Cap, ~5000 deg^2 will be imaged deeper and also include H band producing median 5σ point source limits (Vega mags) of: J = 20.6; H = 19.8; K_s = 18.5. In this 5000deg^2 region of sky the STFC funded Dark Energy Survey (DES) will provide deep multi-band optical (grizY) imaging data. The remainder of the high galactic latitude (|b|>30°) sky will be imaged in YJHK for 60sec per band to be combined with ugriz waveband observations from the VST ATLAS survey.

VHS is a public survey and the data is made public to the whole ESO community and scientific exploitation of the data depends on the science archive elements of the VDFS in Edinburgh.

The medium term scientific goals of VHS include:

- The discovery of the lowest-mass and nearest stars
- Deciphering the merger history our own Galaxy via stellar galactic structure
- Measurement of large-scale structure of the Universe out to z~1 and measuring the properties of Dark Energy
- Discovery of the first quasars with z>7 for studies of the baryons in the intergalactic medium during the epoch of reionization

In addition the VHS survey will provide essential multi-wavelength support for the ESA Cornerstone missions; XMM-Newton, Planck, Herschel and GAIA.

Yours sincerely

Richard McMahon

Dr Richard McMahon

Durham
University

Shaped by the past, creating the future

27th June, 2012

Astronomy Grants Panel,
STFC

Dear AGP members,

Letter in support of WFAU in respect of VST ATLAS archiving.

VST ATLAS is an ESO public survey made using the VLT Survey Telescope and I am the PI of the ATLAS project. The survey is aimed at being a Southern SDSS survey with an area of ~4500deg^2 targetted to be done in *ugriz* in the next 2 years ~1000deg^2 has already been observed. Basic data reduction is being done at the Cambridge Astronomical Surveys Unit (CASU), directed by Mike Irwin. Archiving for ATLAS will then be done at the Edinburgh Wide Field Astronomy Unit (WFAU) directed by Bob Mann. Since WFAU are also archiving the VISTA Hemisphere Survey (VHS) *YJHK* near-ir imaging data they are in an ideal position to make a survey over a large fraction of the Southern sky at high galactic latitudes that includes complete *ugrizYJHK* coverage. This will be a unique database for extragalactic and galactic astronomy. I therefore strongly recommend that support for WFAU be maintained in the current grants round to allow them to set up the VST ATLAS archive and then support the science exploitation that will flow from it over the next 5 years.

Yours faithfully,

(Tom Shanks,
Professor of Physics).

Science Site, South Road, Durham DH1 3LE

Telephone +44 (0) 191 334 3721 Fax +44 (0) 191 334 3645
www.durham.ac.uk

Astronomy Unit
Queen Mary, University of London
Mile End Rd
London , E1 4NS , UK

July 4, 2012

Dear Sir/Madam

Letter of support for VISTA Science Archive grant

I am writing in support of the application for grant renewal of the Wide Field Astronomy Unit, Edinburgh, specifically the activities relating to VISTA and the VIKING survey.

The VIKING survey (VISTA Kilo-degree Infrared Galaxy Survey) is one of the six large Public Surveys ongoing with the VISTA telescope, in which the UK invested 37 million pounds from JIF and PPARC/STFC. Currently approximately 1/3 of the planned VIKING survey has been observed, and a number of exciting science results are emerging, including discovery of three $z > 6.5$ quasars (Venemans et al 2012, in preparation), and near-IR identifications for over 10,000 Herschel-ATLAS sources (Fleuren et al 2012, MNRAS 423, 2407), and the prospect of discovering many hundreds of new gravitational lens systems (Gonzalez-Nuevo et al 2012, ApJ, 749, 65).

In the near future we expect other substantial advances from VIKING as wider wavelength coverage is available: matching visible-band data is now being observed by VST-KIDS, and mid-IR from the recent public release of all-sky WISE data; all these will combine to give 16-band coverage spanning visible, near, mid and far-IR bands, a fundamental sample for studying evolution of galaxies and AGNs at moderate redshifts.

Due to the very large size of the VIKING dataset (so far, approx 11000 GB of raw data files, and 1200 GB of reduced tile images and source catalogues), it is strongly unmanageable for individual science users to process or store the full dataset. Dedicated processing, archiving and access facilities are mandatory for this; especially important here is the SQL front-end engine at WFAU, which uses sophisticated indexing routines to provide fast and flexible access to any desired subset of the data, enabling complex multi-parameter queries to be run very efficiently.

Furthermore, current multi-band catalogues rely on independent source extractions per passband, subsequently band-merged. The next development, required by ESO for the next level data release, is to add *list-driven photometry*: i.e. a source list is generated using one “master” passband, such as KIDS $i-$band, and then list-driven matched-aperture photometry is produced in the VIKING near-IR bands for each of the resulting sources. This produces more reliable colours, and robust limits for non-detections, which is essential to optimise multi-colour selection and photometric redshifts.

On behalf of the VIKING team, I strongly support the renewal of the WFAU grant.

Yours sincerely,

Dr. Will Sutherland
PI, VIKING survey
(Senior Lecturer, Astronomy Unit,
QMUL)

Department of Physics
Imperial College London

Room 1115, Blackett Laboratory
Prince Consort Rd
London SW7 2AZ
Tel: +44 (0)207 5947554 Fax: +44 (0)207 5947541

s.j.warren@imperial.ac.uk
www.imperial.ac.uk/research/astro

19 June 2012

Prof. S. J. Warren
Professor of Astrophysics
Head of Astrophysics Group

Wide Field Astronomy Unit
Royal Observatory Edinburgh
Blackford Hill
Edinburgh
EH9 3HJ

Wide Field Astronomy Unit grant application

I am writing this letter, in my role as Survey Scientist of the UKIRT Infrared Deep Sky Survey (UKIDSS), to express my support for the grant renewal of the Wide Field Astronomy Unit (WFAU). The proposal sees WFAU working on UKIDSS data until Sept 2014. This letter explains the importance of this work.

The main UKIDSS observing campaign finished in May this year. Up to Sept 2013 further observations of the Ultra Deep Survey element will be undertaken. Additionally all the small holes in UKIDSS, where the data failed quality control, will be patched. Then the entire UKIDSS dataset will be prepared for the final data release. This legacy dataset will provide a valuable resource for several years, until overtaken by Euclid around 2020. For this reason it is vital that the most useful resource possible be provided in this final UKIDSS data release. To this end with WFAU the UKIDSS consortium have discussed the requirements for the final database design. This will include enhancements, such as coadding first and second epoch data to improve depth, that are now possible because the survey has been completed.

UKIDSS has been a highly successful programme, with over 300 papers so far, and currently accumulating publications at the rate of 100 per year. This success is founded on the work of the UKIRT operations team, the data processing by the Cambridge Astronomical Survey Unit, and the archiving by WFAU. From the evidence of the publication record so far we can anticipate that the final data release will be heavily used over several years, bringing the total publication list to well over the landmark of 500 papers.

Yours sincerely,

STEPHEN WARREN
Professor of Astrophysics
Head of Astrophysics Group