PIPELINE PROCESSING, DATA PRODUCTS & SCIENCE ARCHIVING FOR WFCAM DATA

Nigel Hambly & Bob Mann

Wide Field Astronomy Unit (WFAU), Institute for Astronomy, University of Edinburgh

Mike Irwin & Jim Lewis

Cambridge Astronomy Survey Unit (CASU), Institute of Astronomy, University of Cambridge

November 7, 2001

1 INTRODUCTION

The Cambridge Astronomy Survey Unit (CASU) and Edinburgh Wide Field Astronomy Unit (WFAU) have been involved in processing and archiving digital wide–field surveys for some time. We intend to use the experience gained from these surveys (eg. APM/SuperCOSMOS surveys, INT WFS and CIRSI survey) to devise pipeline processing schemes, define data products and produce a science archive for all WFCAM data (eg. data from consortium surveys – ie. UKIDSS, PATT and non–PATT programmes).

In addition to generic survey projects we have also developed pipeline processing tools to deal with user data from INGRID and the wide field mosaic optical imagers at the AAO and on the CFHT. We envisage that the tools available within these pipelines will form the basis for data processing all of the WFCAM data. Likewise our current science archiving and provision of user access to a multiplicity of different astronomical datasets will form the basis for the development of the WFCAM science archive, although we intend to base part of the system on the SX software (see later) developed by the SDSS (subject to official agreement).

The end–to–end WFCAM project is overseen by JAC at Hawaii, and in broad outline consists of:

- building and installing the camera and designing instrumentation software (including mountain–top quick–look pipeline) – responsibility ATC
- producing the UK–side data processing pipeline – responsibility CASU
- providing the end–user science archive – responsibility WFAU

This document is concerned with a top–level specification for pipeline processing, data products and science archiving. For the sake of brevity and simplicity, all technical details, discussion of issues and references is confined to the appendices. Note that it is envisioned that external projects, such as AstroGrid or VISTA, may influence the final design of, and produce enhancements, to the basic specification presented here.

2 PIPELINE PROCESSING REQUIREMENTS

2.1 Organisation

There is an operational and a practical requirement for three distinct data processing pipelines: a summit pipeline; a basic pipeline back in the UK; and an advanced pipeline also in the UK. All WFCAM data, including non-UKIDSS data, is to be processed through the summit and basic pipelines and the calibrated science data made securely available through the on-line WFCAM archive.

An on-line database describing the observations, and the summit-derived Data Quality Control (DQC) information will be maintained on Hawaii and mirrored in Cambridge, as is done currently with all UKIRT observations. These may require updating when final DQC information is available from the UK–run basic pipeline. This database will form the knowledge interface that helps define and plan the overall observing strategy.

2.2 Overall strategy

The ORAC control system will produce standard Multi Extension Files (MEFs) in FITS format for export to UK pipelines. Stored within will be sufficient ancilliary information to fully describe the data taken and to invoke the appropriate pipeline processing actions.

The summit pipeline will run on the 'native' NDF format used at UKIRT, but will otherwise provide a subset of the operations carried out in the basic pipeline.

Data will be taken according to a fixed menu of pre–defined observing protocols, in order to guarantee pipeline processing both at the summit and in the UK. These will include single shot, interleaving, dither sequences, tile sequences and so on – details TBD (see glossary in Appendix for definition of terms).

Our general philosophy is that all fundamental data products are FITS MEFs, including the generated catalogue binary tables, and that all derived information (DQC, photometric and astrometric calibration) and processing steps are also incorporated within the FITS headers.

The FITS headers thereby provide the basis for ingest into databases for archiving and databases for real time monitoring of survey progress and survey planning.

2.3 Summit Pipeline

The main purpose of the summit pipeline is to generate near real-time DQC information by processing data on-the-fly using library master calibration frames.

The summit pipeline will probably run under the control of the ORAC data aquisition system and will produce the export MEF FITS files and instigate a FITS header and data verification program.

The summit pipeline will have to maintain and update a "best guess" series of master calibration frames (dark, flats, skies ...) to provide basic 2D instrumental signature removal. The DQC feedback will probably require some form of catalogue generation software which can be used to produce measures for DQC monitoring, including: astrometric pointing accuracy - requires telescope to point to better than (TBD) arcsec rms; photometric throughput; image shape measures – trailing; sky brightness levels; sky noise levels ... (details TBD). This latter software could be a subset of the basic pipeline.

To minimise data volumes the data aquisition system must carry out basic operations such as reset-correction and stacking multiple sequential exposures (an 'integration') in situ before forwarding data to the rest of the system.

2.4 Basic pipeline

The basic pipeline shall operate on single pointings or "well-defined" observing sequences taken during a single night to produce instrument signature–free images and to derive object catalogues from them. The object catalogues will be used to provide the first pass astrometric and photometric calibration.

The two-dimensional image data products at this stage will be instrumental signature–corrected: single frames; lossless interleaved super frames; stacked dither (super)frames; tiled contiguous mosaics, combining four pointings with any of previous options; and confidence maps for all pipeline output products.

The catalogue data products will be lists of detected images with an agreed set of parameters summarising useful astronomical information and providing the necessary DQC information.

In more detail the requirements are:

- instrumental signature removal – bias, non-linearity, dark, flat, fringe
- sky background tracking and removal during stacking – possible need to also remove other 2D background variations from imperfect multi-quadrant operation of detectors
- define and produce a strategy for dealing with image retention from preceding exposures
- combine frames if part of an observed dither sequence or tile pattern
- consistent internal calibration to put observations on an approximately uniform system
- basic catalogue generation including astrometric, photometric, shape and DQC information
- final? astrometric calibration from the catalogue with an appropriate and agreed World Coordinate System (WCS) in all FITS headers
- basic photometric calibration from catalogue using suitable pre-selected standard areas covering entire field-of-view to monitor and control systematics
- each frame and catalogue supplied with provisional calibration information and overall morphological classification embedded in FITS files
- propagation of error arrays and use of confidence maps
- realistic errors on selected derived parameters
- nightly extinction measurements in relevant passbands (also possibly JCMT information on water vapour content since can alter apparent extinction/ effective bandpass more than airmass)
- pipeline software version control – version used recorded in FITS header

2.5 Survey progess Database

There is a requirement for automatic real time monitoring of survey progress with minimal user input via an online databse. This could either be from the normal SYBASE repository, a POSTGRESQL database similar to the experimental one used for monitoring the INT WFS or an interface to the proposed SX archive database discussed later (TBD).

The requirement is for automatic real time monitoring with minimal user input and the provision of simple visualisation of overall survey progress.

This should include interactive GUIs and CLI for querying survey DQC, state of processing, progress of survey, and should also be capable of feedback to observing strategy.

This database could also provide the driver for many of the more advanced processing options that require predefined (asynchronous) observations to have occurred in for example, multiple passbands or muliple visits to the same field for deeper stacking.

2.6 More advanced processing

Many things common to basic pipeline requirements/issues additional features that may be required include:

- stacking images taken at different times and producing contiguous tiled regions – these could be driven automatically from a DataBase driven list of survey fields and survey progress eg. is it worth (re)stacking this field ? [includes assessing various stacking techniques and producing object catalogues from stacked images]
- automated PSF generation over final product field
- optional PSF fitting to improve s:n of point sources

- adaptive kernel matching of PSFs for difference imaging
- detection and parameterisation of Low Surface Brightness objects
- list driven co-located photometry - cf. SDSS (eg. the ability to measure fluxes and limits to fluxes in a consistent way)
- mosaicing contiguous areas to produce seamless large area catalogues and images
- overlaps to refine photometric calibration and map systematics
- hiding gaps – optimal coverage to do this requires study of tiling patterns
- matching/merging frames in different passbands
- processing simulated data and completeness of catalogues
- federating with other catalogues
- allowing external users to upload own code to process selected regions
- morphological decomposition for brighter galaxies
- reprocessing of entire dataset after upgrades/bug fixes

2.7 Hardware/software requirements

In order to provide a backup for the main on-line science archive at ROE and also to facilitate several of the advanced processing/reprocessing stages needed, we propose to store on-line the science output and master calibration data of the basic pipeline in Cambridge. Costs of suitable scale RAID arrays (eg. initially 10 Tbytes capacity) are certain to fall sufficiently by 2004 to make this a cost effective and viable solution.

With on-line storage in place at two sites we do not foresee the need for an (expensive) near-line "tape" storage system providing that a high density transfer medium is used from Hawaii, whereby the $\approx$100 Gbytes of data expected per night is stored on one tape. If reprocessing from the raw data is required this would be feasible, even for on-the-shelf storage, with a one-to-one system like this.

We expect to have sufficient computing power in place as part of the Cambridge Data Processing Centre activities to handle the data flow from WFCAM. Our current pipeline processing activities will provide the basis for development of a UKIRT WFCAM pipeline data. We intend to appoint a dedicated UKIRT WFCAM data processing manager who will be responsible for overseeing the operation of the UKIRT pipeline and organising appropriate software effort as necessary.

3 SCIENCE ARCHIVE REQUIREMENTS

The science archive shall consist of a publicly accessible database of astrometrically and photometrically calibrated images and object catalogues along with housekeeping data. The science archive will contain *all* data from WFCAM (ie. the UKIDSS consortium surveys and data from PATT/non–PATT programmes). The science archive will thus consist of the following logically independent entities:

- several object catalogue survey databases for the UKIDSS surveys
- a database of images in UKIDSS–surveyed areas
- a database of all other WFCAM observations in the form of images and associated object catalogues
- a database of all housekeeping data associated with WFCAM observations

In addition to the independent archive of housekeeping data, such information pertinent to subsets of image/catalogue data will be stored with those data (in the form of header information). Housekeeping data will include generic Data Quality Control (DQC) measures including site information (conditions/weather), software versions, 'release' versions (for calibrations etc.)

Data products extracted from the science archive shall be in FITS format (images; small object catalogues attached to images; very large object catalogues) and ASCII format (general object catalogues and housekeeping data). Catalogues shall additionally be available in Tab–separated List (TSL) format for compatibility with ESO–compliant browsers and remote server communication (Davenhall 2000).

3.1 Functional requirements

The end–user shall access the archive via the internet and Graphical User Interfaces. Pixel data shall be supplied with astrometric WCS, flux calibrated pixels, and confidence maps which include bad pixel flags/masks; object catalogues shall be merged across avaliable passbands and shall include error estimates for subsets of the more important parameters.

The end–user shall be offered:

- pixel data in any available passband over small areas of order up to a few tens of arcmin (TBD) with associated matched object catalogue
- larger areas from survey data, mosaiced on–the–fly and blocked down as required/appropriate
- optimally stacked images where survey data are available
- seamless, merged multi–colour, multi–parameter catalogues with best available astrometric and photometric calibration from survey data
- ability to specify the object parameters in the user–extracted subset from survey data
- survey database federation with other source catalogues (primarily SDSS and Schmidt photographic sky surveys)
- ability to construct readily meaningful optical/IR colours for all sources, not just point sources, in the overlap with the SDSS
- ability to carry out *efficient* queries to locate objects conditional on ranges of the following parameters: spherical co-ordinates in equatorial/Galactic systems; radius within given sky position; JHK magnitudes and any linear combination thereof; object size; object classification (stellar/non–stellar/other) [1]
- ability to carry out queries on any other stored parameter
- for all queries, archive shall be able to return an estimate of the number of objects to be located and time required to implement the search (not only for user information, but also to help archive management, scheduling of queries, etc.)
- ability to carry out queries on recalibrated quantities or previous versions of calibrations
- ability to perform repeated queries ('batch' mode) on survey data
- ability to create simple finder chart (for use while observing for example)
- ability to access 'best', or all of a set of duplicate measurements where they exist in overlapping survey data (eg. generate light curves and/or proper motions from duplicate measurements of the same object)
- general access to housekeeping data – eg. what's currently available; how good is it; in short, user access to all housekeeping data specified in the previous section

[1] We envisage two generic types of searches, "efficient" and "advanced". The first is predicated on efficient indexing and storage and is intended to provide very fast access on the most commonly requested search parameters. The advanced search would allow more complex queries but at the expense of a reduction in response time.

3.2 Technical requirements

- data ingest from CASU: data will be supplied in the form of separate FITS multi–extension files (MEFs) for images and object catalogues with housekeeping/processing information stored in FITS headers
- UKIDSS survey object catalogues database will be implemented using SX; pixel databases will be implemented with in–house software
- SX GUI will be as for SDSS ('sdssQT'; currently written in TCL/TK but soon to be replaced by a Java version)
- Distributability: no requirement on WFAU to distribute all or part of the archive other than through the interface specified previously
- calibrated quantities to be specified via coefficient values and calibration scheme (astrometric/photometric)
- Security: archives shall be protected against corruption (ie. backed up and organised in such a way as to minimise the risk of accidental deletion/alteration)
- Security: proprietary data will have restricted access
- the science archive shall be constructed in such a way so as *not* to preclude AstroGrid–type functionality/enhancements
- the science archive shall be constructed in such a way so as to make data available to users as soon as is practicable – hence it shall be able to cope with incomplete observations and regular data ingest without disrupting online access to existing database

3.3 Hardware/software requirements

Hardware requirements in broad outline:

- Hard disk: the maximum size of the pixel archive is estimated as follows: a single WFCAM science frame is 4 devices $\times$ 2048 $\times$ 2048 pixels $\times 2 \times 2$ interlaced $\times 4$ bytes per pixel = 270 Mbyte. For a 7 yr lifetime, 75% UKIRT time devoted to WFCAM we have to deal with 2000 nights of science data. Assuming ~ 200 frames per night (an average of one every 3 mins over a 10hr night) then the final pixel archive is ~ 100 Tbyte in size. 10 Tbytes up–front will be more than adequate initially, then adding ~ 10 Tbyte per year is a possibility. All other data (derived from pixels or elsewhere) will be $< 10\%$ the size of the pixel archive.
- the backup or redundant storage for the above will be based in Cambridge as part of the basic pipeline
- CPUs: single–node server or Beowulf cluster
- network connectivity: efficient communication requires at least 100 Mbit/s capability[2]

Software requirement in broad outline:

- SX and associated database management system (currently Objectivity but this may well change over the next couple of years)
- database software for housekeeping information

[2]The Blackford Hill link to the outside world needs upgrading. Presently, the exisiting microwave link can handle 10 Mbit/s. In early 2002 a new fibre link will be established which will initially provide 100 Mbit/s upgradable to 1 Gbit/s. Within the ROE site, the core network currently runs at 1 Gbit/s and the existing archive server www–wfau has a 100 Mbit/s connection; these are reviewed and upgraded on a yearly basis.

- multi–node/user licences (as appropriate) for any commercial software
- parallelisation software for Beowulf, if required
- GUI software home–grown by inhouse programmer(s)
- remote server interface for network–based browsers

4 HUMAN RESOURCE ESTIMATES AND TIMESCALES

We estimate that development over the next two years will require approximately 3 FTE (most of this effort will be expended on designing and coding of the UK–side pipeline, photometric calibration and general database design). Thereafter, we estimate that maintaining the pipeline, data products and archive will require 6.5 FTEs. Again, the greater fraction of this will be expended in maintaining, developing and running the pipeline; implementing calibration and maintaining and developing the science archive. Within CASU/WFAU these resource requirements are under review as part of the Units' Rolling Grant review/renewal procedures. We are currently working on the assumption that first light for WFCAM will occur in the final quarter of 2003; the Units Rolling Grants are being reviewed in the first quarter of 2002 with expected grant renenwal covering the period 2002-2007.

A APPENDICES

A.1 Glossary of Terms

channel – data channel from each 2kx2k detector through to the backend of the pipeline

pointing – a telescope slew to position with new guide star acquisition. Note that a microstep sequence is done within one pointing

read – a read is the act of physically reading a detector. This generally means the act of digitising the data, and can be either destructive (ie. followed by a clear) or non–destructive

exposure – An exposure is a sequence of one or more reads, that are used to produce a single output image

integration – An integration is a sequence of one or more exposures. Whereas the exposure is the result of a sequence of reads, the integration is a summation (or similar) of a sequence of exposures

integration sequence – An integration sequence is a sequence of integrations as the name suggests. It is characterised by the types of operations that can take place during it; an integration sequence will contain oversampling operations etc.

microstep sequence – a sequence of 1, 2x2 or 3x3 integrations taken with sub–pixel shifts.
A microstep sequence is one type of integration sequence

observation – An observation is a sequence of integration sequences that makes up a usable scientific observation. An observation is the level at which calibration frames (flats, skies, darks etc.) are collected

reduced frame – result of reducing an integration

superframe – the result from combining (interleaving) the reduced frames from a microstep sequence. Superframe is 2kx2k, 4kx4k or 6kx6k depending on microstep mode

loops – multiple exposures on one place to increase integration time

dithers – sequence of exposures or loops macrostepped in "spiral" pattern for use in bad pixel rejection, deep stacking etc.....

tile – any or all of four separate exposures, loops, dithers or pointings taken so as to cover a contiguous area of sky involving all 4 detectors coadd – image resulting from combining loops taken sequentially

stack – image resulting from combining dithers including coadds if present

mosaic – complete contiguous image combined from tile exposures

confidence map – a normalised inverse variance weight map defining the "confidence" associated with the intensity value in each pixel, this also encodes for hot pixels, bad pixels, dead pixels and so on. The raw confidence maps for each frame output by the Data Acquisition System (DAS) will be derived from regular (frequency TBD) analysis of the calibration flats and darks. All output processed frames (stacked, tiled, mosaiced etc.) will also have an associated derived confidence map.

A.2 General Issues

All UK pipeline processed data will be archived online in the UK, with raw data probably being stored off line "on a shelf". There does not appear to be a compelling need for a near-line "tape" robotic storage system ?

Observing strategy for the interleaved mix of UKIDSS survey and PI observations is a further issue that needs addressing. This should take into account interactive feedback from derived DQC information and optimising data taking sequence to address sky variablity issues and so on. We suggest that the JAC + UKIDSS survey project scientist should be responsible for overseeing the development of the toolkit required to do this.

Test observations to answers question regarding image persistence, possible intra-pixel sensitivity problems, non-linearity monitoring and so on ... who is organising them and guaranteeing that they are addressed ?

Setting up photometric standard fields covering entire FOV – see related photometric calibration document.

It is crucial that accurate, agreed and complete FITS headers are attached to all files for automatic pipeline processing (details TBD but should also include site monitoring information). A FITS header (and data) verification program should be run on all summit-produced FITS files before export to the UK. Any faults should be fixed in situ before shipping.

Are there science requirements for real–time data processing on the summit ? eg. transient event detection such as moving objects, variables ..., and related to this, are data products other than DQC information required at summit ?

Constraints on interleaving and dither patterns from non-linear astrometric field distortions to avoid pixel interpolation schemes – tiling will undoubtedly require non-linear resampling in all cases. The envisaged astrometric field distortion for WFCAM implies that for "dithers" up to $\approx$20 arcsec, non-linear resampling will not be needed for stacking data within one device. Satisfying this constraint is desirable for various reasons.

Defining a standardised set of parameters for each detected object describing the image location, the flux, the shape and error estimates for these parameters (details TBD but see Appendix C for starting point cf. also with SuperCosmos, APM, SDSS, 2MASS)

Undersampling and intra-pixel sensitivity variations problems

How to deal with bright stars, ghostimages, trails ? One possiblity it to use input predictions for positions of bright stars, galaxies, globular clusters, solar system objects and so on.

Do we want image atlases for extended sources ? - colour pictures ? directed public outreach products?

Community input on data processing products and requirements – UKIDSS is represented directly is there a need for wider consultation ?

Documentation and release of software ? Should we release the pipeline software ?

Official UK set of survey data products - can be multiple copies ? Continuous -v- staged release of survey data products ? Publication of survey and responsiblity for that ?

A.3 Outline of Basic Pipeline

The steps in the basic pipeline are as follows:

1. Input reset-corrected single frames 1 MEF per pointing
2. Non-linearity correction if needed
3. Update master dark, flatfield frames and bad pixel masks
4. Dark-correct, flatfield and gain-correct on-sky data
5. Update current running sky estimate if appropriate
6. Sky subtract all target frames
7. Remove further instrumental artefacts if present
8. Derive internal astrometric transformation if necessary
9. Form interlaced superframe and pixel variance map
10. Generate object catalogues
    - Astrometric and photometric calibration
    - Morphological classification
    - Data quality control measures
    - Check by matching with 2MASS catalogue

A.4 Data Products from Basic Pipeline

The data products from the basic pipeline will be:

- Observing logs built from FITS headers
- SYBASE user interface to access data
- Library dark frames, flatfield frames, bad pixel masks, sky/defringing frames and non-linearity corrections
- Master colour equations for all filters
- Astrometric calibration – WCS in FITS headers
- Photometric calibration – zero points and extinction
- Data quality control diagnostic information
- Calibrated and instrumental signature-removed 2D image maps
- Morphologically classified and calibrated object catalogues
- Record of all processing in FITS headers
- Selected Matched catalogues for colour-magnitude, variability and proper motion analysis

B Data Quality Monitoring

WFCAM data will come with the following DQC measures that will greatly enhance the usefulness of data products for the end user if available:

- Night–time observing conditions/weather
- Instrument status values: detectors/filters/controllers OK

Individual frames will have the above information written into their headers; quick–look summit pipeline will then derive further DQC measures and insert into headers:

- sky brightness – saturated? Too bright? No signal?
- sky noise – detectors OK? Pattern noise?
- average stellar ellipticity – trailed images?
- average seeing measure
- stellar aperture correction – weird PSF?
- no. of spurious images – corrupted data?
- astrometric errors – wrong RA, Dec?
- throughput from rough photometric calibration

All processing done by the UK–side pipeline will record progress in image FITS headers and derive further DQC measures.

C Image Descriptors

APM/SuperCOSMOS analysis produces 32 4-byte parameters per detected object. As an example the INT Wide Field Camera survey pipeline produces the following set (which has evolved from the original APM set):

No.	Name	Description
1	Seq. no.	Running number for ease of reference, in strict order of image detections
2	Isophotal flux	Standard definition of summed flux within detection isophote, apart from detection filter is used to define pixel connectivity and hence which pixels to include. This helps to reduce edge effects for all isophotally derived parameters.
3	Total flux	An attempt at generating automatically a total flux using a curve–of–growth technique. The closest description as to how it works is given in Hall & Mackay (1984). This is better than the isophotal flux in terms of systematic error, but still not perfect and definitely worse for random error.
4	Core flux	Best used if a single number is required to represent the flux for ALL objects. Basically aperture integration with radius rcore (in the FITS header) but modified to simultaneously fit 'cores' in case of overlapping images.
5	X co-ord	Intensity–weighted centre–of–gravity in X
6	Y co-ord	Intensity–weighted centre–of–gravity in Y
7	Gaussian sigma	These are derived from the intensity–weighted second moments. The
8	Ellipticity	equivalence between them and a generalised elliptical Gaussian distribution
9	Position angle	is used to derive Gaussian sigma $= (\sigma_a^2 + \sigma_b^2)^{1/2}$
		Ellipticity $= 1.0 - \sigma_a/\sigma_b$

		Position angle = angle of ellipse major axis wrt x axis
10	Peak height	in counts relative to local value of sky
11	Areal profile 1	Number of pixels above a series of threshold levels relative to local sky.
12	Areal profile 2	Levels are set at T, 2T, 4T, 8T ... 128T where T is the threshold. These
13	Areal profile 3	can be thought of as a sort of poor man's radial profile. Note that for
14	Areal profile 4	deblended, ie. overlapping images, only the first areal profile is computed
15	Areal profile 5	and the rest are set to −1 flagging the difficulty of computing accurate
16	Areal profile 6	profiles.
17	Areal profile 7	
18	Areal profile 8	
19	Core 1 flux	A series of different radii core/aperture measures similar to parameter 4.
20	Core 2 flux	Together with parameter 10 these give a simple curve–of–growth analysis from
21	Core 3 flux	peak pixel, $1/2\times$ rcore, $\sqrt{2}\times$ rcore, $2\times$ rcore, $2\sqrt{2}\times$ rcore
22	Core 4 flux	
23	RA	RA and Dec explicitly put in columns for overlay programs that cannot, in
24	Dec	general, understand astrometric solution coefficients. Derived exactly from
		WCS in header and XY in parameters 5 & 6
25	Classification	Flag indicating what image is most probably: −1 stellar, +1 non–stellar, 0 noise
26	Statistic	An equivalent N(0,1) measure of how stellar–like an image is, used in
		deriving parameter 25 in a 'necessary but not sufficient' sense
27	Core 5 flux	$4\times$ rcore, extra to ensure $\sim 99\%$ of PSF flux
28	Sky level	Interpolated sky level from background tracker
29	Child/parent	Flag for parent or part of deblended deconstruct
30	Blank	
31	Blank	
32	Blank	

This will be enhanced to include error estimates for a TBD subset of the parameters.

As an alternative the SDSS archive has a much larger set of parameters per object (see for example `http://www.sdss.jhu.edu/`) – many are error estimates on parameters analogous to the above. Others include different flux measures – eg. Petrosian magnitudes, and list–driven flux measures expressed in asinh magnitudes whereby flux estimates at the same position in all other wavebands are obtained given a significant detection in any one passband. It is proposed that the parameters derived froma WFCAM image in a single passband will be similar to the above INT–WFC set. Input from the WFCAM user community (especially UKIDSS) would be welcomed to help finalise the details of the parameter set, but it is clear that any significant enhancement lies outside the scope of the WFCAM Science Archive as achievable with existing resources and would require additional resources, eg. from AstroGrid or specific science projects.

Over the last few years, provision for user access to large databases has grown from simple position and proximity–based querying capability to highly flexible systems employing 'structured query language' (SQL) syntax. Examples of the former are the UK photographic archives at Cambridge (APM) and Edinburgh (SuperCOSMOS):

`http://www.ast.cam.ac.uk/~mike/apmcat`
`http://www-wfau.roe.ac.uk/sss`

while an example of the latter is the 2MASS archive:

`http://www.ipac.caltech.edu/2mass`

The most recent large–scale archive, that of the SDSS, employs an extended SQL (SXQL) syntax in a flexible database system (SX) which sits on top of the commercial object–oriented database management system Objectivity; details are available at `http://www.sdss.jhu.edu/` and see also Thakar, Kunszt & Szalay (2001) and references therein. The object–oriented approach brings many advantages, not the least of which is enabling whole database searches for specific queries to be executed quickly provided the database is indexed (or 'tagged') in an appropriate way. A simple example of an SXQL procedure is:

```
// Select all stars in a given RA range

SELECT
RA(),DEC(),
g,r,i,z
FROM sxStar
WHERE (
RA() BETWEEN 15 AND 16 && g < 20
)
```

where the the select–from–where construct allows the user to choose the parameters to be output, the 'class' of object to be searched, and the search criteria to be applied to all images in the searched class.

A mirror site for the SDSS EDR is being established at Edinburgh. Given the manpower effort expended on the state–of–the–art SDSS system and its obvious flexibility, it is proposed that it be employed as the basis for the WFCAM Science Archive. Experiments are underway at Edinburgh with SX/Objectivity; these will include implementation on a Beowulf cluster for enhanced access speed. The general feeling is that SX is likely to become the community–wide standard since much development has already been done on it.

References

Davenhall A.C., 2000, Starlink System Note 75.1: Writing catalogue and image servers for GAIA and CURSA

Hall P., Mackay C.D., 1984, MNRAS, 210, 979

Thakar A.R., Kunszt P.Z., Szalay A.S., 2001, In: Mining The Sky, eds. A.J. Banday, S. Zaroubi & M. Bartelmann