

VDFS: The VISTA Science Archive

Nicholas Cross, Rob Blake, Ross Collins, Mark Holliman, Mike Read, Eckhard Sutorius, Nigel Hambly, Andy Lawrence, Bob Mann, Keith Noddle

*Wide Field Astronomy Unit, Institute for Astronomy,
University of Edinburgh, UK*

What is the VSA?

- Relational Database Management System
 - Linked tables containing different types of data
 - Design emulates data structure.
 - meta-data from images
 - Catalogue data – detections, merged sources, variability statistics
 - Self-describing: information about each programme and processing in curation tables
 - Microsoft SQL Server: reliable product used by SDSS, WSA, 2 of the most successful astronomical archives.
- Main DB and documented static releases
- Multiple interfaces for different scientific usage

Purposes of Science Archive

- Interface for survey teams and community to explore survey products to do **science**.
- Interface for survey teams to check data for **quality control**.
- **Repository** of VISTA data from reduced images to complex, catalogue products.
- Requires both:
 - a **dynamic main-DB** which is updated with new data, better calibration, reprocessing, quality control, higher order products.
 - Static, well documented **release-DBs** that can be referred to in publications.

The VSA: <http://surveys.roe.ac.uk/vsa>

The screenshot shows a Firefox browser window displaying the VISTA Science Archive website. The browser's address bar shows the URL horus.roe.ac.uk/vsa/. The website's header includes navigation links: Home | Overview | Browser | Access | Login | Cookbook, followed by the VSA logo. The main content area features the VSA logo and the title "VSA - VISTA Science Archive". Below the title, there is a paragraph explaining the archive's purpose: "The VISTA Science Archive (VSA) holds the image and catalogue data products generated by VIRCAM on the Visible and Infrared Survey Telescope for Astronomy (VISTA). The primary contents of the archive originate from the VISTA Public Surveys. Survey science-ready catalogue data will be released in phases, while standard flat-file data products (both images and derived single passband catalogues) become available continually after routine observation and processing operations. Information on the various archive releases can be found on the [surveys page](#)". A second paragraph states: "The history of archive releases, updates and bug fixes is recorded under the [release history](#) page. Users wishing to receive email announcements of such entries should subscribe to the VSA_Announcelist (contact vsa-support@roe.ac.uk)."

On the left side of the website, there is a vertical navigation menu with the following items: VSA Home, Start Here, Data Overview, Known Issues, the Surveys, Schema browser, Data access, Login, Archive Listing, GetImage, MultiGetImage, Region, Freeform SQL, CrossID, SQL Cookbook, Q&A, and Glossary.

The main content area features a large image of the Milky Way galaxy with several survey fields overlaid. The fields are labeled: VIKING (green), UltraVISTA (red), VHS (blue), VVV (cyan), VIDEO (white), and VMC (yellow). The VVV field is highlighted with a cyan box.

The browser's taskbar at the bottom shows the system tray with the date and time: 17/03/2012, 01:34. The system tray also includes icons for network, volume, and power.

WFAU tasks

- **Ingest** nightly processed image and catalogue data from CASU – **See MJI talk**
- **Provenance** - link related images
- **Quality Control**: Automated + input from teams.
- Link **tile and pawprint** data
- Process data for **semester** - done **per programme** :
 - Produce and ingest **deep** stacks/tiles/mosaics + catalogues
 - **Merge** pass-band catalogues to create source tables
 - Create **neighbour** tables to link external catalogues
 - Link multi-epoch data and calculate **variability** statistics
- **Release** a documented, static data product to users
- Create useful **interface tools** for users to query specific data, view and analyse it

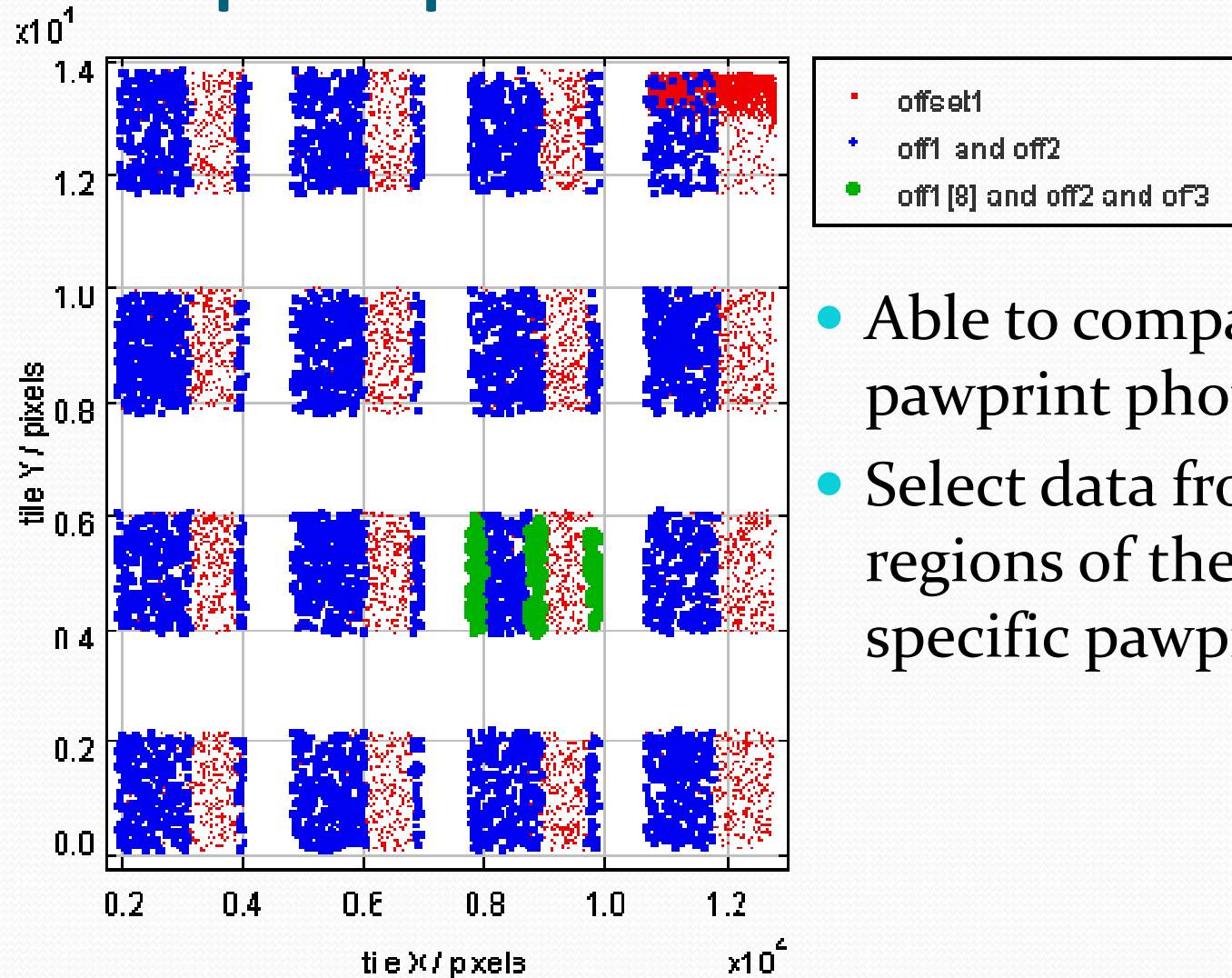
Automated Pipeline

- Post QC tasks run in automated pipeline
 - Uses DB to determine what needs to be done
 - How many pointings, how many filters, how many epochs?
 - What has already been completed?
 - Have processes been done in correct order?
 - Consistency between expected products and actual
- Reduces workload on operations (when it is working)
- Essential for processing many PI programmes with same range of products as main surveys.

VISTA complications

- Technical:
 - Pawprints + Tiles: two layers of products, detections from both kept
 - **10X** increase in catalogue data
 - VVV so time consuming that a separate server is needed, but some tables and data common – synchronisation
- Political:
 - WFCAM: WFAU deal with UKIDSS, CASU, UKIRT (v. occasionally)
 - VISTA: WFAU deal with separate survey teams, ESO, CASU

Tile-pawprint detections linked



- Able to compare tile and pawprint photometry.
- Select data from specific regions of the tile or specific pawprints.

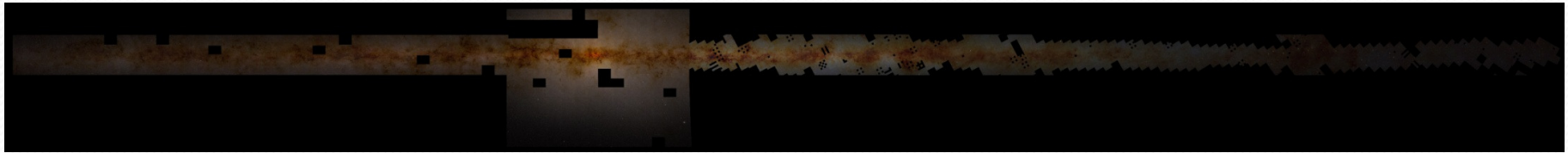
Data Rates WFCAM → VISTA

- Images:
 - WFCAM 1720 raw image frames a day (4 2kx2k)
 - VISTA 580 raw image frames a day (16 2kx2k)
 - Only 30% increase in raw image volume.
- Catalogues:
 - More area (2.9x), increased sensitivity (~2x), tile + pawprints (~3x). Expect 15-20x as many detections.
 - **14X** WSA: 3.2M detections per day, VSA: 44M
 - **Catalogues are important factor for relational DB**
 - VVV: VSA currently has >21billion rows, **expect ~10¹¹**

VVV takes archive to new level

- Up to P87:
 - VMC: 500 million detections, 18 million objects
 - VHS: 1.9 billion detections, 270 million objects
 - VVV: 21 billion detections, 500 million objects
- **Data rate of VVV is 10x other surveys.** Combination of shallow and dense fields
- ~1 billion stars, 80 – 100 epochs.

Galactic Plane and Bulge from VISTA and WFCAM



VVV + GPS mosaic ~ 1 billion stars

<http://djer.roe.ac.uk/vsa/vvv/iipmooviewer-2.0-beta/lb.html>

March 29th 2012, #1 most read article on BBC

VVV processing

- VVV takes months to process 1 year of data.
- Variability statistics is slowest stage.
 - Many speed ups already
 - separate servers for main and release DBs
 - Split detection tables into months
 - Parallelisation of CPU intensive processes that do not call DB
 - I/O is main bottleneck now. Solutions:
 - Improved optimisation of curation queries
 - Column orientated databases
 - Solid state disks for DB.
 - Reprocessing of data a major headache

Using the Archive

- Support
- Documentation / Publications
- Quality Control
- Different Interfaces

Expert support:

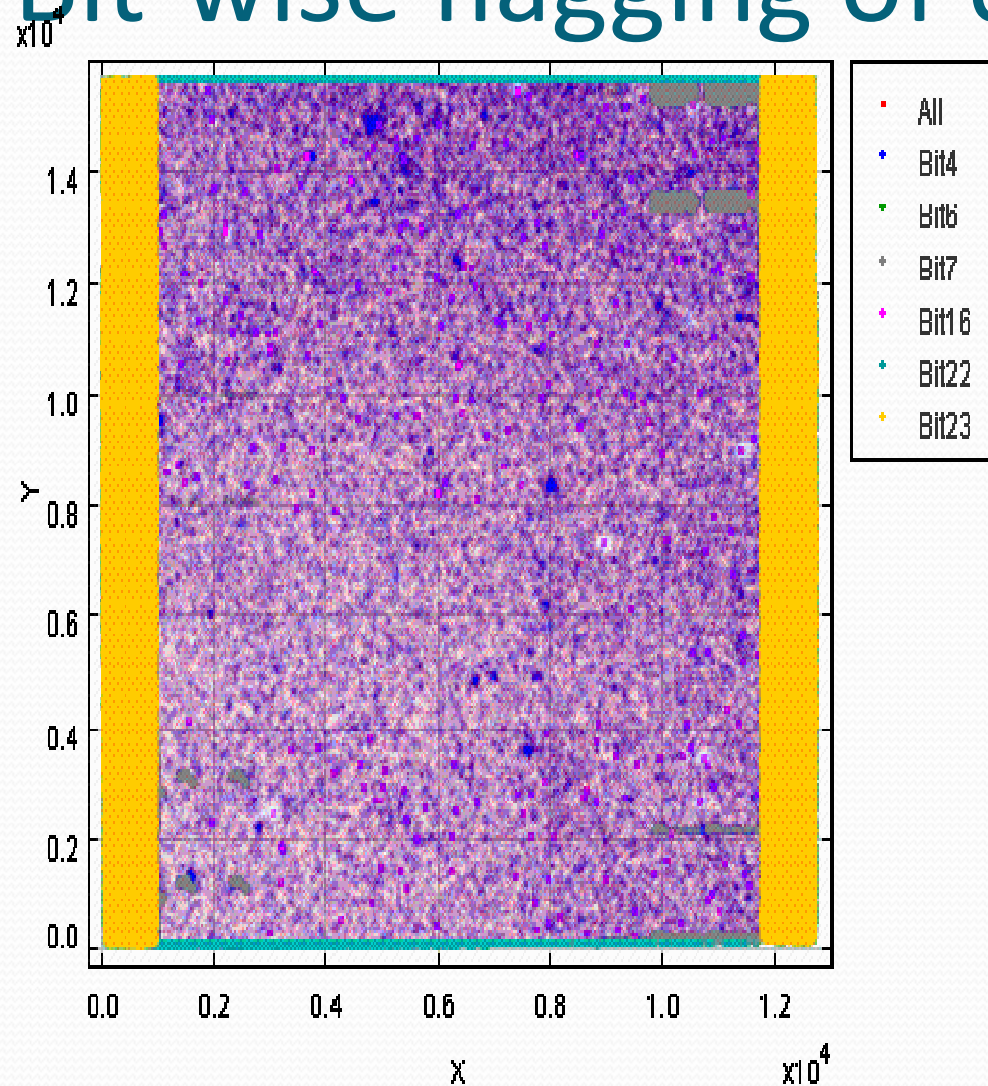
vsa-support@roe.ac.uk

- VSA: 10-20 helpdesk queries per month
 - Similar in WSA, more mature archive.
- Types of queries posed
 - Detailed knowledge of data (and data quality)
 - Detailed knowledge of SQL
- WFAU have a mixture of technical and scientific knowledge.

Provenance tracking is crucial

- Rare object search is major analysis mode
 - Selecting on wide range of attributes
- Reproducibility
- Is this an unusual object or a junk data value?
 - Quality control at image and detection level is vital.
- Track back from each data value to parent data and processing chain
 - Complex data structure – absent in flat-file catalogue
 - At file level, e.g. tiles – pawprints, deeps – OBs, stack – raw.
 - Detection level, multi-band – single band, tile detection matched to pawprint detections.

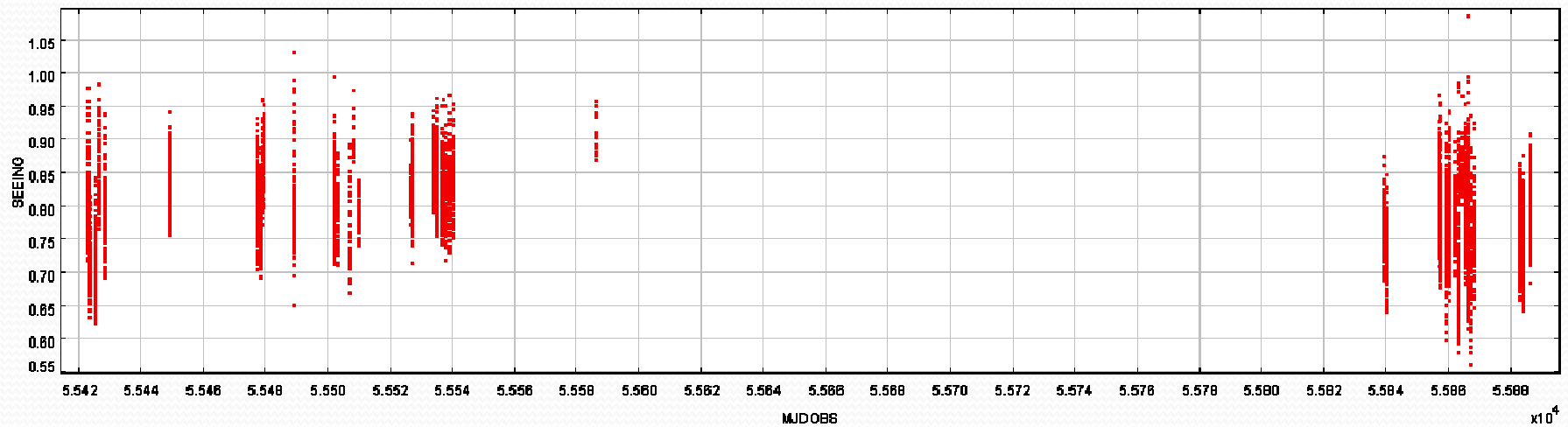
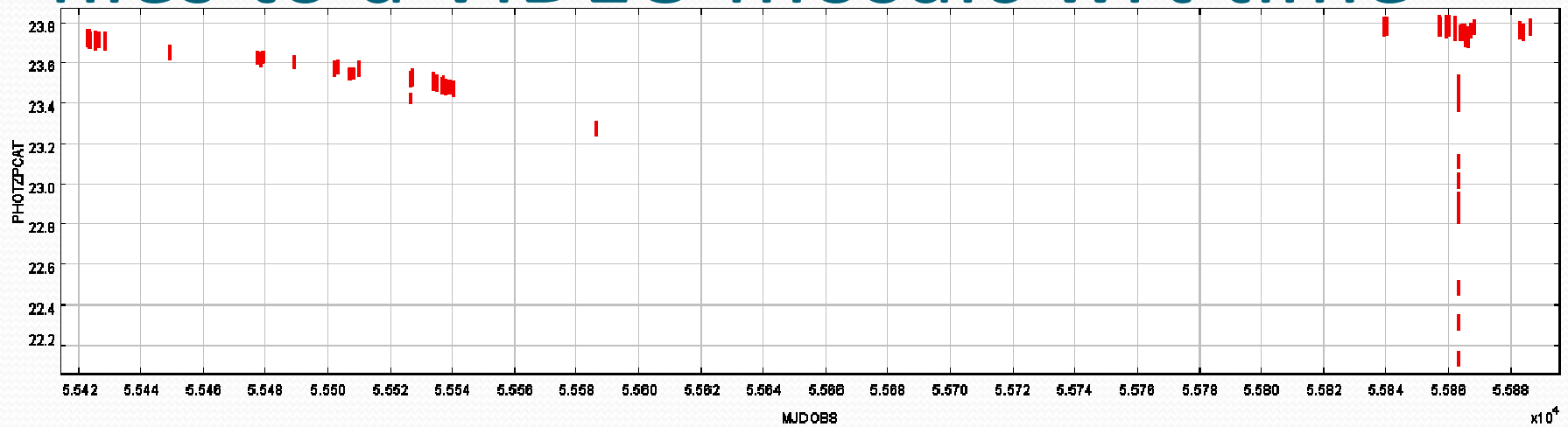
Bit-wise flagging of detection data



VMC deep tile.

- Under-exposed strip
- Edges
- Low confidence
- Deblends
- Saturated
- Bad pixels
- Detector 16
- **Bright stars to come**

Provenance: ZP and seeing of input files to a VIDEO mosaic wrt time



Using the VSA

VISTA SCIENCE ARCHIVE - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://horus.roe.ac.uk/vsa/

munich observatory university

Most Visited Getting Started Latest Headlines Personal banking: Bank ...

Search Amazon BBC News BBC Sports YouTube Facebook Weather Games Financial Times

AVG Aktuelle Search Page Status News IDENTITY GUARD E-mail Weather

VISTA SCIENCE ARCHIVE

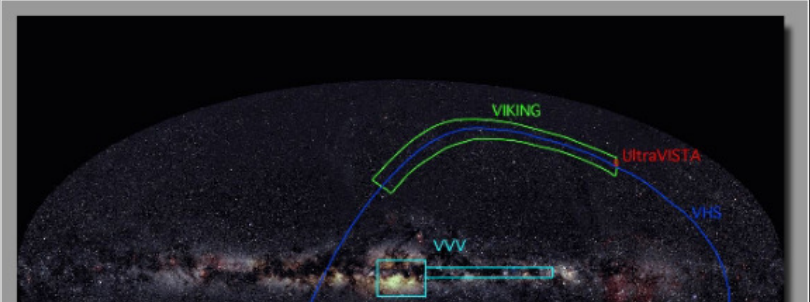
Home | Overview | Browser | Access | Login | Cookbook VSA

VSA - VISTA Science Archive

The VISTA Science Archive (VSA) holds the image and catalogue data products generated by VIRCAM on the Visible and Infrared Survey Telescope for Astronomy (VISTA). The primary contents of the archive will originate from the VISTA Public Surveys. Survey science-ready catalogue data will be released in phases, while standard flat-file data products (both images and derived single passband catalogues) become available continually after routine observation and processing operations. Information on the various archive releases can be found on the [surveys page](#)

We apologise for the lack of service between 16/09/2010 and 21/09/2010. There was a power outage at the Royal Observatory during this period due to a meltdown of a transformer at the substation.

The history of archive releases, updates and bug fixes is recorded under the [release history](#) page. Users wishing to receive email announcements of such entries should subscribe to the VSA_Announcelist (contact vsa-support@roe.ac.uk).



Done

EN 22:37 19/01/2011

VSA – schema browser

The screenshot shows a web browser window displaying the VSA Browser interface. The browser's address bar shows the URL `http://horus.roe.ac.uk/vsa/www/vsa_browser.html`. The page header includes the VSA logo (VISTA Science Archive) and the WFAU logo. Navigation links include Home, Overview, Browser, Access, Cookbook, Links, and Credits. The main content area displays the schema for the `vhsSource` table. A sidebar on the left lists various database objects and tables, with `vhsSource` selected. The table description states: "Contains merged sources from detections in vhsDetection." and provides a prescription for merging: "Prescription for the merging (from RequiredFilters) is: Y, J, H, Ks". It also lists required constraints: "Primary key is (sourceID)" and "(frameSetID) references vhsMergeLog(frameSetID)". A table below lists the columns: sourceID (bigint, 8), cuEventID (int, 4), and frameSetID (bigint, 8), along with their descriptions and unified content descriptors.

TABLE vhsSource
Contains merged sources from detections in vhsDetection.

Prescription for the merging (from RequiredFilters) is: Y, J, H, Ks

Required constraints:

- **Primary key is (sourceID)**
- (frameSetID) references vhsMergeLog(frameSetID)

Name	Type	Length	Unit	Description	Default Value	Unified Content Descriptor
sourceID	bigint	8		UID (unique over entire VSA via programme ID prefix) of this merged detection as assigned by merge algorithm		ID_MAIN
cuEventID	int	4		UID of curation event giving rise to this record		REFER_CODE
frameSetID	bigint	8		UID of the set of frames that this merged source comes from		REFER_CODE

7. Using multi-epoch data: light curves etc

7.1 Selecting variable stars

```
SELECT v.sourceID
FROM vwSource AS s,vwVariability AS v
WHERE s.sourceID=v.sourceID AND s.mergedClass=-1 AND v.variableClass=1
```

This will select the stars which have been classified as variable using the current algorithm in the Schema Browser glossary.

Other selections using the statistics calculated in the Variability table can be used, e.g. selecting bright objects with high rms in J and Ks and large absolute skew in each band.

```
SELECT v.sourceID
FROM videoSource AS s,videoVariability AS v,videoVarFrameSetInfo AS i
WHERE s.sourceID=v.sourceID AND s.mergedClass=-1 AND v.frameSetID=i.frameSetID AND
ksMeanMag<(i.ksexpML-3.) AND v.ksMeanMag>0. AND v.jMeanMag<(i.jexpML-3.) AND v.jMeanMag>0. AND
v.jMagRms>(5.*v.jExpRms) AND v.ksMagRms>(5.*v.ksExpRms) AND abs(v.jskewness)>10. AND
abs(v.ksskewness)>10.
```

7.2 Producing light curves

```
SELECT m.mjdObs,d.aperMag3,d.aperMag3Err,d.ppErrBits,d.seqNum,x.flag
FROM videoSourceXDetectionBestMatch AS x,videoDetection AS d,Multiframe AS m
WHERE x.sourceID=446677514563 AND x.multiframeID=d.multiframeID AND x.extNum=d.extNum AND
x.seqNum=d.seqNum AND x.multiframeID=m.multiframeID AND d.filterID=5 ORDER BY m.mjdObs
```

This gives the necessary data for the ks-band light curve for sourceID=446677514563 in VIDEO. Objects with

Archive Listing

The screenshot shows a Mozilla Firefox browser window with the following details:

- Address Bar:** http://horus.roe.ac.uk/wsa/cgi-bin/display.cgi?file=/disk35/vsa/products/stacks/20100910_v4/e20100910_0013000027_0
- Search Bar:** Includes a search engine (Ask) and various shortcuts like Amazon, BBC News, BBC Sports, YouTube, Facebook, Weather, Games, and Financial Times.
- Taskbar:** Shows several open tabs: "Changeset 777...", "OGLE", "OGLE-III Catalo...", "VMC Survey", "Stephen Rough...", "http://...ID=385", "VSA ImageList", and "SSA Browser: D...".
- Page Content:** Displays a grid of 12 image detector frames, labeled "image/detector: 1" through "image/detector: 12". Above the grid, there are download links for compressed (40.41 Mb) and uncompressed (30.59 Mb) FIT files, and a note about image compression.
- System Tray:** Shows the time as 22:30 on 17/01/2011, along with system icons for network, volume, and power.

18th January 2013

VISTA, A Celebration!

22

An Example: Selecting point source

4.5
4.0



Firefox | VSA Database - SQL Query R... | Microsoft Outlook Web Access | https://webmai...saic.movie.avi | VISTA SCIENCE ARCHIVE

AS NAM

3BC Sports

Options

Upload SQL query from file into this for

or enter SQL statement:

```
SELECT
/* Select time, filter,
error and flags */
d.mjd,d.filterID,d.apex
d.ppErrBits
/* From BestMatch table
observations of the ss
videoDetection for each
FROM videoDetection as
videoSourceXDetectionBe
/* First join tables */
WHERE b.multiframeID=d
b.extNum=d.extNum AND b
/* then select only de
equal to object in pre
which has a Ks-band RMS
AND d.seqNum>0 AND b.sc
SELECT s.sourceID
FROM videoVariability v
/* join tables */
WHERE v.sourceID=s.sou
/* point source variabl
s.mergedClass IN (-1,-2
v.variableClass=1 AND
/* delta mag in > 0.1 i
at least 5 good obs in
(((zMaxMag-zMinMag)>0.
AND znGoodObs=>5) OR ((
AND yMinMag>0. AND ynGo
(((jMaxMag-jMinMag)>0.1
jnGoodObs=>5) OR ((hMa
hMinMag>0. AND hnGoodOb
((kMaxMag-kMinMag)>0
kenGoodObs=>5))
/* zRange|>4 mag */
AND ((zMaxMag-zMinMag)>
/* order by time */
ORDER BY d.mjd
```

Submit ensure one of the file formats is selected below if you want to save your results.

WFAU
IFA ROE

EN 22:52
28/03/2012

Enhanced Queries

The screenshot shows a Mozilla Firefox browser window displaying the VISTA Science Archive website. The address bar shows the URL: `http://horus.roe.ac.uk:8080/vdfs/VSQL_form.jsp?disp=adv`. The page title is "VISTA SCIENCE ARCHIVE - Mozilla Firefox". The browser's menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. The address bar also shows "Toshiba Satellite Pro". The browser's toolbar includes a search bar, a search button, and various icons for Amazon, BBC News, BBC Sports, YouTube, Facebook, Weather, Games, Financial Times, and Options. The browser's tabs include "Re: issues with prin...", "BT FON", "OGLE", "OGLE-III Catalog o...", "[3.0] Awk Example...", "VISTA SCIENCE AR...", and "VISTA SCIENCE A...".

The main content area of the browser shows the VISTA Science Archive interface. The left sidebar contains a navigation menu with the following items: Data Overview, Known Issues, the Surveys, Schema browser, Data access, Login, Archive Listing, GetImage, MultiGetImage, Region, Freeform SQL, CrossID, SQL Cookbook, Q&A, Glossary, Release History, Gallery, Publications, Monitor, and Downtime. The main content area has a "Database release to use:" dropdown menu set to "VMCv20100930" and a link to "notes on VSA (QC) database". Below this is a form for uploading a SQL query from a file, with a "Browse..." button and an "Upload" button. The "or enter SQL statement:" section contains a text area with the following SQL query:

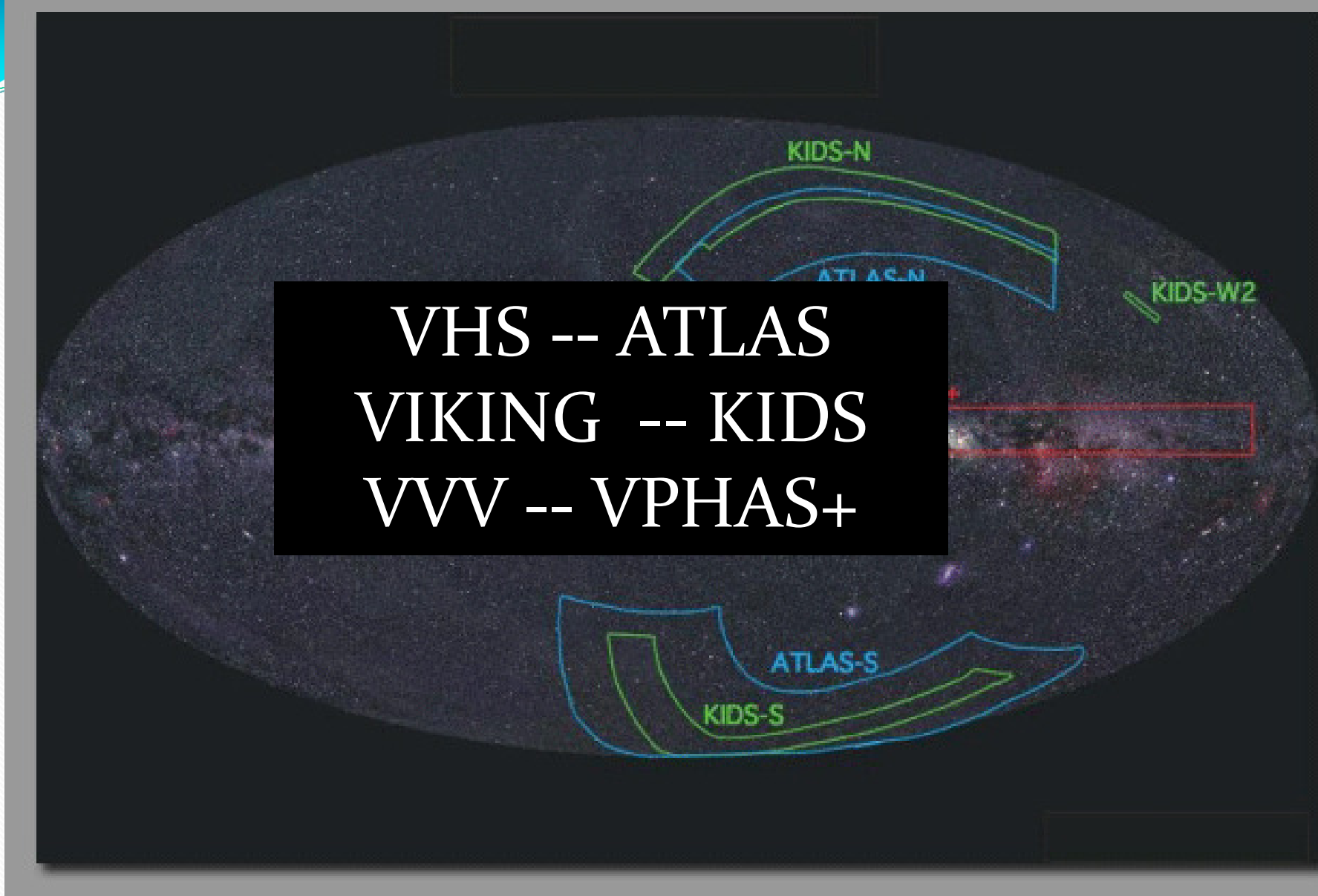
```
select
b.sourceID, l.meanMjdObs, s.ksaperMag3, s.ksaperMag3Err, s.kspErrBits, b.flag
from #userTable as u, vmcSourceXSynopticSourceBestMatch as b, vmcSynopticSource
as s, vmcSynopticMergeLog as l
where u.sourceID=b.sourceID and b.synFrameSetID=s.synFrameSetID and
b.synSeqNum=s.synSeqNum and b.synFrameSetID=1.synFrameSetID and s.synSeqNum>0
and s.ksseqNum>0 order by b.sourceID, l.meanMjdObs
```

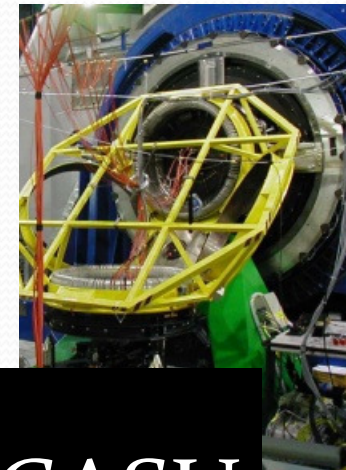
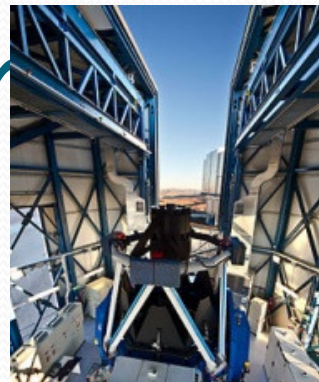
Below the text area is another "Browse..." button and a "Format of upload file:" dropdown menu set to "fits". At the bottom of the browser window, the Windows taskbar is visible, showing the Start button, several application icons, and the system tray with the date and time "16:33 16/01/2011".

Overlaid on the right side of the screenshot is red text that reads: "Use crossID to match with main survey Source table. Take data from outside. Save matched table as title VSA match own data as FITS/VO table #original, as a standard queries".

Surveys are not independent entities

- Importance of multi-wavelength astronomy
 - GAMA, COSMOS fields, deciphering galaxy evolution, with complex star formation histories with dust.
 - Rare objects (e.g. high-z QSOs, BDs, odd transients)
- Common survey fields
 - VST-ATLAS and VHS
 - KIDS and VIKING (GAMA extends this in some areas)
 - VVV, GPS, IPHAS, VPHAS+
- Need for data integration
 - Cross-neighbours tables, publishing to VO
 - Matched aperture photometry





EARLIER PROCESSING AT CASU



Combining Data from Surveys

- Neighbour Tables (Main Existing method)
 - Joins to main external surveys: 2MASS, SDSS, SSA, GALEX, FIRST, WISE,
- Matched aperture photometry.
 - Pipeline almost ready. First use on P90 data.
- Using VO interfaces.
 - Output in VOTables, launcher for TOPCAT
 - SIAP services, footprints for Aladdin.
 - New MyDB style applications on the way.

WFAU VDFS Publications

Hambly et al. 2008, MNRAS, 384, 637 (WSA)

Cross et al. 2009, MNRAS, 399, 1730 (Multi-Epoch processing)

Cross et al. 2012, A&A, 548A, 119 (VSA)

Summary

- VISTA an order of magnitude increase in catalogue data volume of WFCAM
 - VSA met the challenge and producing regular releases to the survey teams and wider community.
 - VVV very difficult, but challenge helps to keep WFAU as one of the leading data centres in the world.
- VSA is the most efficient way for many types of science:
 - to find rare objects (transients in the VVV, very cool brown dwarfs, $z > 7$ QSOs).
 - Working with data across wide areas, multi-wavelengths
- Edinburgh Data Centre with VSA at the centre, linking with WSA, OSA, SSA and external data.
- VISTA will be very successful, many papers already published, many to come: see PI talks.