



# ***Curation of Science Archives at WFAU***

***Eckhard Sutorius***



***Wide Field Astronomy Unit, IfA, Edinburgh***

***R. Blake, R. Collins, N. Cross, N. Hambly,  
M. Holliman, M. Read***

# Overview

- ★ Data Transfer and Ingest
  - ➔ Flat File Access
- ★ Recalibration
- ★ Post-nightly pipeline data products:
  - Deep Stacks
  - Deep Tiles
  - Difference Images
- ★ Provenance
- ★ Quality Control & QC Flags
- ★ Source Merging & Resampling
- ★ Neighbour Table Creation
  - ➔ Release Database

# Data Transfer & Ingest

- ★ Data transfer from  to  via UKLight  
[2.5-5h per observing night; run monthly]
- ★ Compressed Images (JPEGs) for each detector  
[~2.5-5h per observing night]
- ★ Extraction, process & ingest of multiframe metadata  
[~1-2h per observing night]
  
- Flat File Access (images, confidence maps, single band catalogues)!  
[1-2 weeks after CASU releases data]
  
- ★ Extraction, process & ingest of catalogue data  
[2-4h per observing night]
  
- Parallelised processing, run-time mainly determined by ingest.



# Monitor Page

<http://surveys.roe.ac.uk/vsa/index.html>

- VSA Home
- Start Here
- Data Overview
- Known Issues
- the Surveys
- Schema browser
- Data access
- Login
- Archive Listing
- GetImage
- MultiGetImage
- Freeform SQL
- SQL Cookbook
- Q&A
- Glossary
- Release History
- Gallery
- Publications
- Monitor**
- Downtime
- Links

go to [Monitoring home page](#)

**Semester 2009SV Monitoring**  
generated: 2010-01-26 12:11:58.87 UT

**Legend**

- This column shows whether the data are ready for inspection, ie. data are transferred, JPEGs exist, and image metadata are ingested.
- No tickmark means no metadata are ingested yet, so flat file access is not available for this date.
- No JPEGs have been generated yet, but image metadata are ingested so flat file access is available for this date.
- Not all JPEGs have been generated/ingested, but image metadata are ingested so flat file access is available for this date.
- All JPEGs are calculated and ingested, image metadata are ingested.

No VIRCAM data taken

- CU1** Number of files transferred from CASU. Subdivided by types: science frames, catalogue files, all files (incl. calibration frames).
- CU2** Number of calculated JPEGs. Subdivided by number of JPEGs and number of FITS files. Normally each FITS file has 16 JPEGs associated, one per extension.
- CU3** Number of FITS files that have image metadata ingested. Subdivided by types: pixel data files, catalogue files, all files.
- CU4** Number of catalogue data objects that have been ingested: **only Raw ingested, Photometry ingests dimer from Raw**. Subdivided by survey: VHS, VVV, VMC, VIKING, VIDEO, SVORION, SVNGC253, ULTRAVISTA, CAL & COMM, all catalogue data objects.

**October 2009 (2009SV)**

Date		CU1 (transfer from CASU) (version/#sci/#cat/#all)			CU2 (JPEGs calculated) (version/#jpgs/#files)		CU3 (Image metadata ingested) (version/#pix/#cat/#all)			CU4 (Detections ingested) (version/#ultravista/#vhs/#video/#viking/#vmc/#vvv/#svorion/#svngc253/#cal&comm/#all)													
15		0	157	24	211	0	2992	187	0	187	24	211	0	0	0	0	0	0	0	555668	155790	711458	
16		0	414	96	582	0	7456	466	0	466	96	582	0	0	0	0	0	0	3236256	1042541	124226	4403023	
17		0	524	135	717	0	9312	582	0	582	135	717	0	0	0	0	0	0	5255583	522018	97971	5875572	
18		0	499	110	610	0	8000	500	0	500	110	610	0	0	0	0	0	0	3950205	1218045	0	5188250	
19		0	537	99	683	0	9344	584	0	584	99	683	0	0	0	0	0	2693086	2343633	845676	2254028	8137203	
20		0	599	143	802	0	10544	659	0	659	143	802	0	0	0	0	0	4220399	3805042	1063220	2222613	11311274	
21		0	539	125	724	0	9564	599	0	598 (-1)	124 (-1)	722	0	0	0	0	0	4760859	3851596	665662	2117091	11395208	
22		0	618	141	820	0	10864	679	0	676 (-3)	138 (-3)	814	0	0	0	0	0	1264721	3277377	840578	1017536	6400212	
23		0	806	185	1056	0	13936	871	0	871	185	1056	0	0	0	0	0	21006946	4595945	978356	1910138	28491385	
24		0	476	104	627	0	8366	523	0	523	104	627	0	0	0	0	0	2370253	3262857	1330340	2170713	9134163	
25		0	582	138	770	0	10112	632	0	632	138	770	0	0	0	0	0	2357007	3885101	1136937	2103522	9482567	
26		0	586	144	786	0	10272	642	0	642	144	786	0	0	0	0	0	1991660	4721994	952135	1486175	9151964	
27		0	598	133	795	0	10592	662	0	661 (-1)	132 (-1)	793	0	0	0	0	0	2115572	3715859	1704773	1570719	9106923	
28																							
29		0	582	129	771	0	10272	642	0	642	129	771	0	0	0	0	0	1057128	4191820	793824	89831	6132603	
30																							
31																							

**November 2009 (2009SV)**

Shows progress of ingest procedure. If one of the tickmarks is present, flat file access is available, but all JPEGs may not be created yet.



# *Post-nightly pipeline Data Products*

- ★ Recalibration (if necessary)
- ★ Difference images (eg. in WSA: GPS)
- ★ Deep stacks (eg. in WSA: DXS)
- ★ Deep tiles (eg. Mosaics in WSA: UDS)
- ★ Catalogues for these image products
- ★ Quality control by checking correct magnitude limits
- ★ Synoptic data products
- ★ ...

# *Provenance & Quality Control*

- ★ *Provenance*: Information about file provenance
- ★ *Quality Control of detections*: done by archive scientists, some files marked as deprecated, e.g.:
  - 0 = good data
  - 60-70 = eyeball check deprecation
  - >127 = frame superceded by reprocessing
- ★ Update of *Quality Error Bit Flags* (ppErrBits): deblended sources, bad pixel, boundary sources, saturated sources, cross-talk artefacts, etc.

# *Source Merging & Reseaming*

On detections that passed Quality Control:

- \* *Source merging*: creation of multi-colour source lists from individual passband detections.
- \* *Reseaming*: flagging of duplicate objects as primaries or secondaries (relies on quality error bit flag).

# *Neighbour Tables*

- \* Neighbour tables between sources and detections to quickly generate light curves and look for variable objects
- \* Neighbour tables with external catalogues to specifications:
  - SDSS, DENIS, MGC, NVSS, 2XMM, 2MASS, ROSAT, IRAS, FIRST, GLIMPSE, SSA, UKIDSS & VISTA surveys...



# *Release*

- ★ Creation of the Release Database
  - ➔ Created in individual file-groups for faster access
- ★ Copying to Public Catalogue Server
  - ➔ Update of Science Archive Browser pages to reflect changes, eg.
    - ➔ [http://surveys.roe.ac.uk/vsa/www/vsa\\_browser.html](http://surveys.roe.ac.uk/vsa/www/vsa_browser.html)
    - ➔ [http://surveys.roe.ac.uk/wsa/www/wsa\\_browser.html](http://surveys.roe.ac.uk/wsa/www/wsa_browser.html)



# Timeline

## UKIDSS DR3

<b>Transfer</b>	7h per observation night	50d (~7d after last pipeline proc.)
<b>JPEGs</b>	2h calc. + 0.3h ingest per observation night	~7d calc. (parallel) + 2d ingest
<b>Metadata</b>	0.2h calc. + 0.2h ingest per observation night	~1d calc. (parallel) + 1.5d ingest
<hr/>		
<b>Catalogue data</b>	1h calc. + 1h ingest per observation night	4d calc. (parallel) + 7d ingest
<b>Recalibration</b>		4d
<b>GPS diff. images</b>		1.5d
<b>UDS stacks</b>		1d
<b>DXS mosaics</b>		3d
<b>Provenance</b>		1d
<b>Quality Control</b>	depending on PI input	28d
<b>Error bit flags</b>		2d
<b>Source merging</b>	UDS+DXS: 0.5h;GCS: 1d; LAS: 2d; [GPS: 3d]	3d
<b>Resampling</b>	UDS+DXS: 2h;GCS:2d; LAS: 0.5d; [GPS: ~14d]	2.5d
<b>Neighbour tables</b>	UDS+DXS: 1h;GCS:1h; LAS: 4d	4d
<b>Release database</b>	15h creation; 7h copying	1d
<hr/>		
<b>Total Computational Time</b>		~90d




# *Science Archive Releases*

- \* Quality Control of detections is mainly the PI's responsibility
  - Quick but thorough QC leads to earlier releases.
- \* ESO's release policy
  - Major data releases with uniform photometric and astrometric calibrations at least once per year.
  - First delivery expected no later than **18** months after beginning of the observations.


# Science Archive Browser

The portal to data products


[http://surveys.roe.ac.uk/\[vw\(o\)\]sa/index.html](http://surveys.roe.ac.uk/[vw(o)]sa/index.html)



- [VSA Home](#)
- [Start Here](#)
- [Data Overview](#)
- [Known Issues](#)
- [the Surveys](#)
- [Schema browser](#)
- [Data access](#)
- [Login](#)
- [Archive Listing](#)
- [GetImage](#)
- [MultiGetImage](#)
- [Freeform SQL](#)
- [SQL Cookbook](#)
- [Q&A](#)
- [Glossary](#)
- [Release History](#)
- [Gallery](#)
- [Publications](#)
- [Monitor](#)
- [Downtime](#)
- [Links](#)



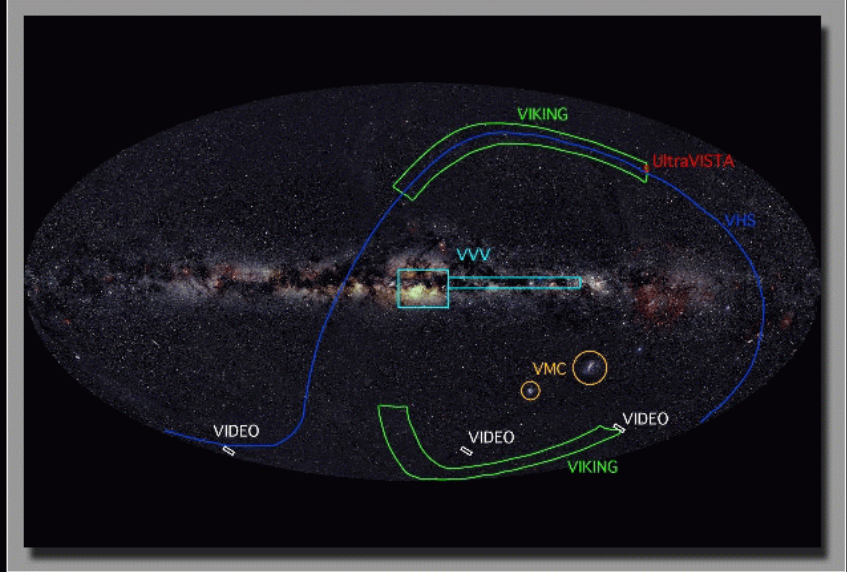
IFA ROE

[Home](#) | [Overview](#) | [Browser](#) | [Access](#) | [Login](#) | [Cookbook](#) 

## 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 be science-ready catalogue data which 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 archive releases can be found on the [surveys page](#)

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 (c



*Picture: Sky coverage of VISTA surveys, overlaid on a 2MASS image of the whole sky.*  
Credit: VISTA

[Home](#) | [Overview](#) | [Browser](#) | [Access](#) | [Login](#) | [Cookbook](#)  
[Listing](#) | [FreeSQL](#)  
[Links](#) | [Credits](#)

# Science Archive Browser: Access

Fill in your login details and your database is ready for querying!

Home | Overview | Browser | Access | Login | Cookbook

## VSA - Data Access

Users can access the data held in the VSA through web-based forms. These forms parse the user's input parameters and submit **SQL** (Structured Query Language) queries to the database.

- **Archive listing** - retrieve listings of the multiframes held in the VSA. Links are returned allowing the user to view the library jpegs and download the FITS files.
- **Freeform query** - submit an SQL query directly

**Catalogue/Table queries**  
Query results i.e. rows selected from the database, are displayed in an HTML table and/or written to a file (ASCII, FITS or VOTable format) that can be downloaded.

Users should note the following general points:

- HTML table output is only intended as a summary and the number of rows displayed is limited. A note at the end of the table informs the user if this was exceeded.
- The number of result rows written to files is also limited and depends on how many parameters have been requested i.e.  $maximum\ rows\ written\ to\ file = \min(15000 / no.\ parameters) \times 1000$

So if only three parameters have been requested than the file can contain up to 5 million rows. Again users are warned if the limit was exceeded and an indication of how many extra rows were requested.

- Users who exceed the file row limit should submit their query in parts e.g. by querying sections in RA or Dec or magnitude slices.
- If an email address is supplied queries are allowed to run for 30 seconds before they are placed in the background and the browser window is released. On completion an email is sent informing the user.
- The tables in the database do not contain any NULL values. Where values are unavailable for a given object parameter default values have been inserted. Users should be aware of this when querying.
- See the **schema browser** for details of a given parameter's default value.
- NULL values can be returned if users JOIN tables. If this occurs the values are written out as zeroes.

Home | Overview | Browser | Access | Login | Cookbook  
Listing | FreeSQL  
Links | Credits

WFAU, Institute for Astronomy,  
Royal Observatory, Blackford Hill  
Edinburgh, EH9 3HJ, UK  
Tel +44 131 668 8356 (office)  
or +44 131 668 8100 (switchboard)

vsa-support@roe.ac.uk  
6/11/2009



# Schema Browser

## Schemas, Tables, Attributes and more...



VSA Browser

Home | Overview | Browser | Access | Cookbook | Links | Credits



VSASV

→ other Browser versions

- [-] VSA
- [-] VSA UltraVISTA
- [-] VSA VHS
- [-] VSA VIDEO
- [-] VSA VIKING
- [-] VSA CALIBRATION
- [-] VSA SV NGC253
- [-] VSA SV ORION
  - [-] Database Objects
  - [-] Tables
    - svOrionDetection
    - svOrionMergeLog
    - svOrionSource
    - svOrionSourceNeighbours
    - svOrionSourceXbwomass\_psc
    - ArchiveCurationHistory
    - AstrCalVers
    - CurationTask
    - CurrentAstrometry
    - ExternalSurvey
    - ExternalSurveyTable
    - Filter
    - Multiframe**
    - MultiframeDetector
    - MultiframeDetectorEsoKeys
    - MultiframeEsoKeys
    - PhotCalVers
    - PreviousMFDZP
    - Programme
    - ProgrammeCurationHistory
    - ProgrammeFrame
    - ProgrammeTable
    - Provenance
    - Release
    - RequiredDiffImage
    - RequiredFilters
    - RequiredListDrivenProduct
    - RequiredMosaic
    - RequiredNeighbours
    - RequiredStack
    - Survey



TABLE **Multiframe**

Contains details of all multiframe stored in the archive.

Required constraints:

- **Primary key** is (multiframeID)
- (filterID) references Filter(filterID)

Name	Type	Length	Unit	Description	Default Value	Unified Content Descriptor
multiframeID	bigint	8		UID of the multiframe (assigned sequentially by the archive ingest process)		ID_FRAME
vistaRunNo	int	4		Original VISTA run number (from filename)		REFER_CODE
creationDate	datetime	8	MM-DD-YYYY	File creation date (YYYY-MM-DD HH:MM:SS) (image primary HDU keyword: DATE)	12-31-9999	TIME_DATE
frameType	varchar	64		The type of multiframe (e.g. stack tile mosaic etc.)	normal	meta.code.class
cuEventID	int	4		UID of curation event giving rise to this record		REFER_CODE
julianDayNum	int	4	Julian days	The Julian Day number of the VISTA night		TIME_DATE
fileTimeStamp	bigint	8		Time stamp digits (from the original CASU directory name and file time stamp) for enforcing uniqueness		??
filterName	varchar	8		VISTA combined filter name (image primary HDU keyword: HIERARCH ESO INS FILT1 NAME)		??
filterID	tinyint	1		UID of combined filter (assigned in VSA: 1=Z,2=Y,3=J,4=H,5=Ks)		INST_FILTER_CODE
project	varchar	64		Time-allocation code	NONE	REFER_CODE
telescope	varchar	16		ESO telescope name (image primary HDU keyword: TELESCOP)	NONE	
instrument	varchar	8		Instrument name (image primary HDU keyword: INSTRUME)	NONE	
arcfile	varchar	64		Archive File Name (image primary HDU keyword: ARCFILE)	NONE	
utDate	datetime	8	MM-DD-YYYY	Observation date (MM-DD-YYYY)	12-31-9999	TIME_DATE
dateObs	datetime	8		Observing date (image primary HDU keyword: DATE_OBS)	12-31-9999	TIME_DATE

# *Discussion Points*

- ★ Calibration done by CASU?
- ★ Data & code from CASU as for VDFS products?
- ★ Data volume (pixel/catalogues)?
- ★ Proprietary rights on Flat File Access:
  - 12 month to survey group?
- ★ Proprietary rights on Catalogue Database Releases:
  - Immediate world wide access?
- ★ Transfer of flat file products into ESO's SAF?