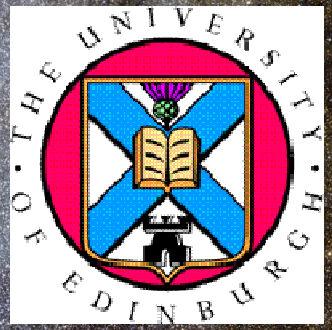
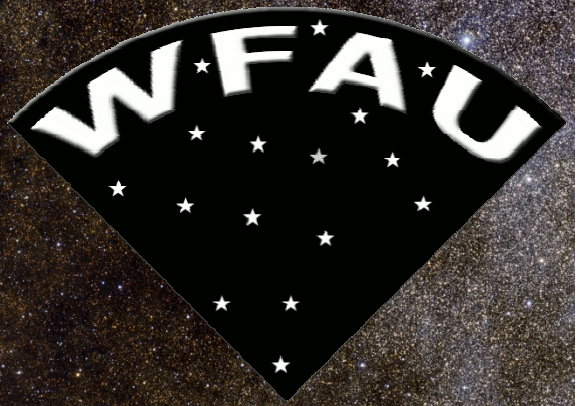




VDFS: The VISTA Science Archive

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

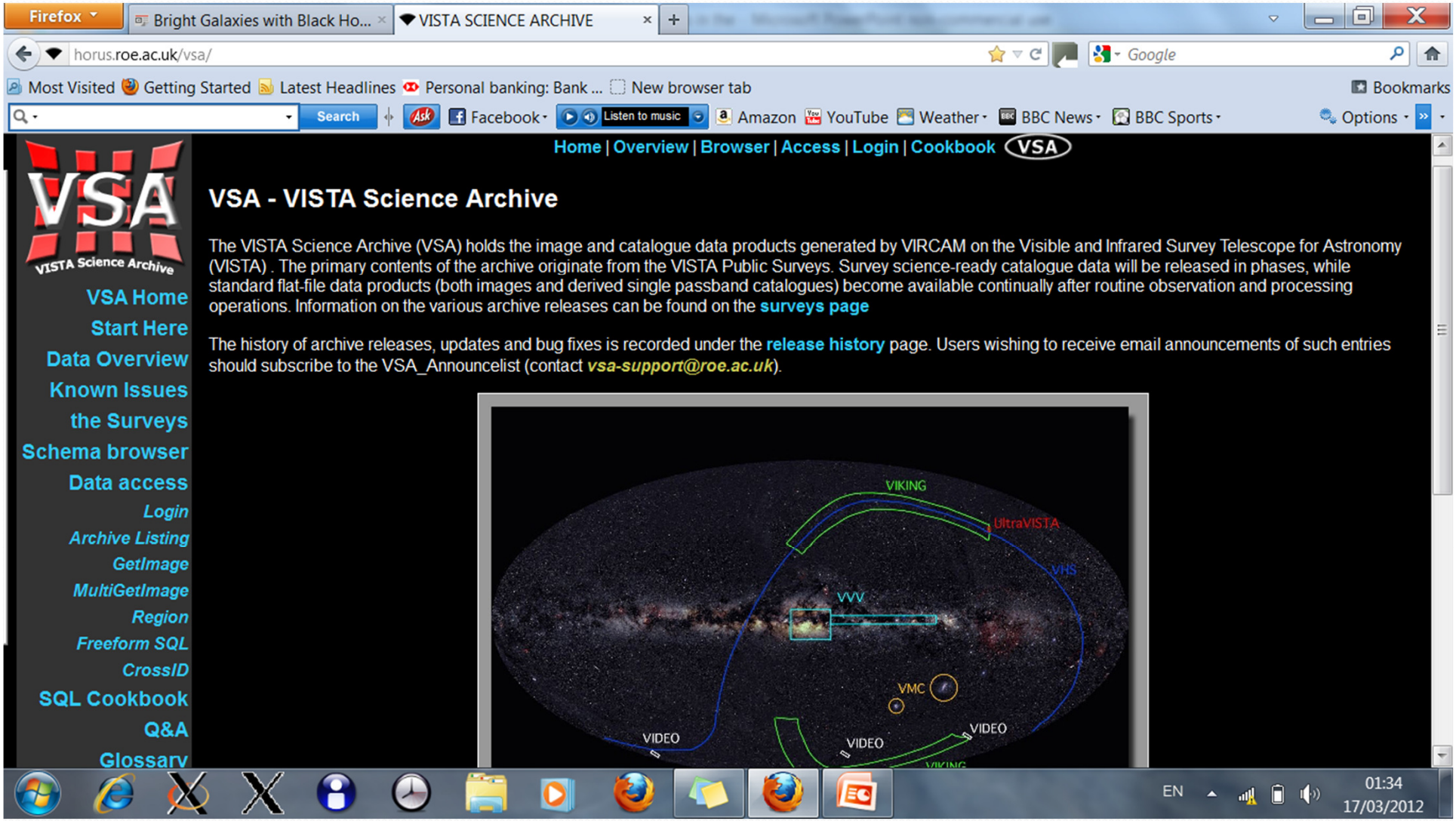
What is the VSA?

- 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



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.

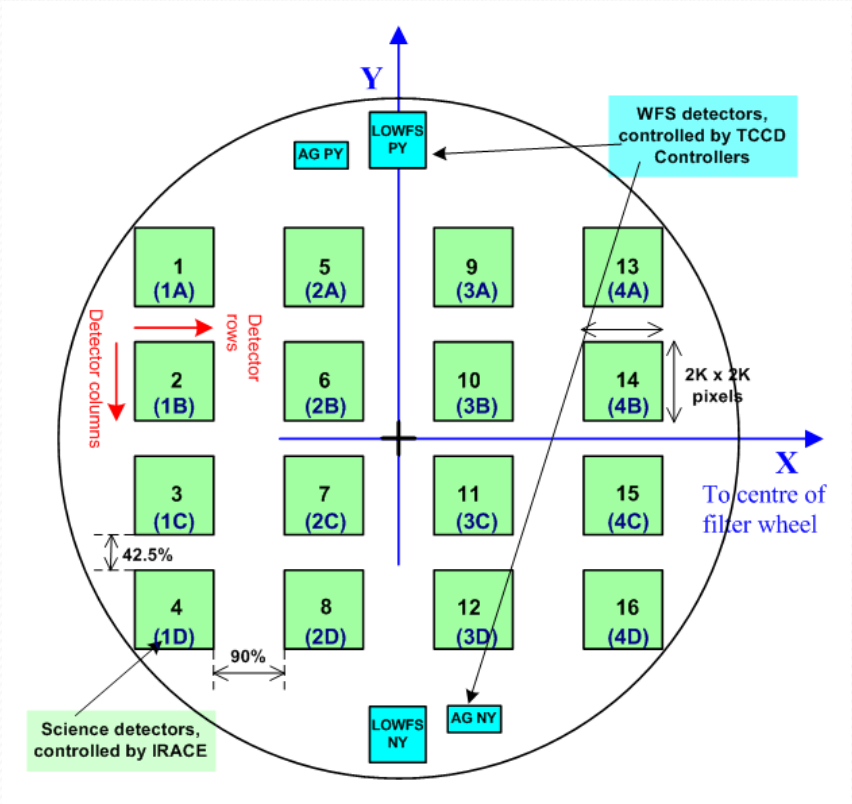
VISTA complications

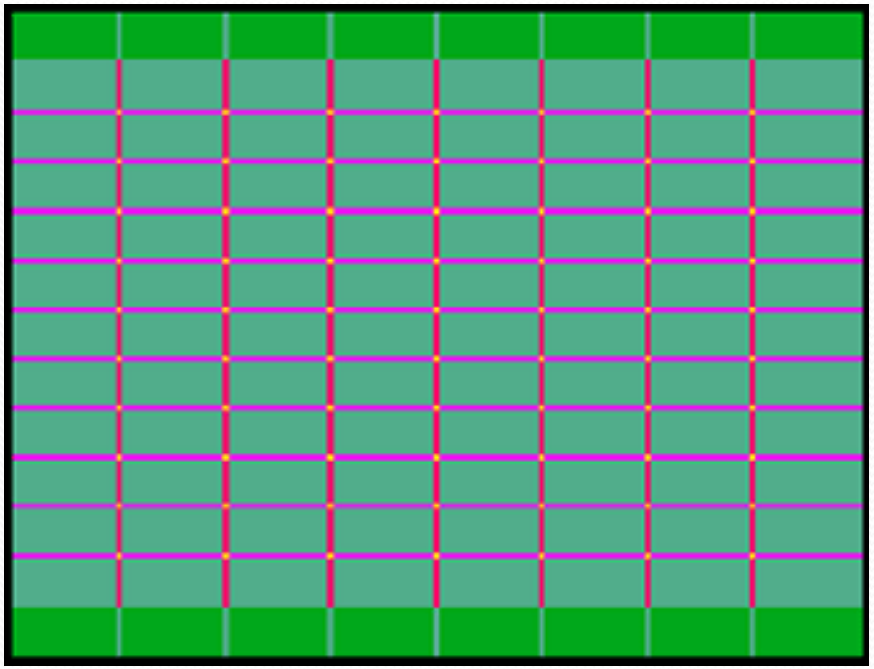
- Technical:
  - Pawprints + Tiles: two layers of products, detections from both kept
  - 10x increase in catalogue data
  - VVV so time consuming that a separate server is needed, but some tables and data common – synchronisation
- Political:
  - WFCAM: WFAU deal with UKIDSS, CASU, UKIRT (v. occasionally)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

VISTA Tiles

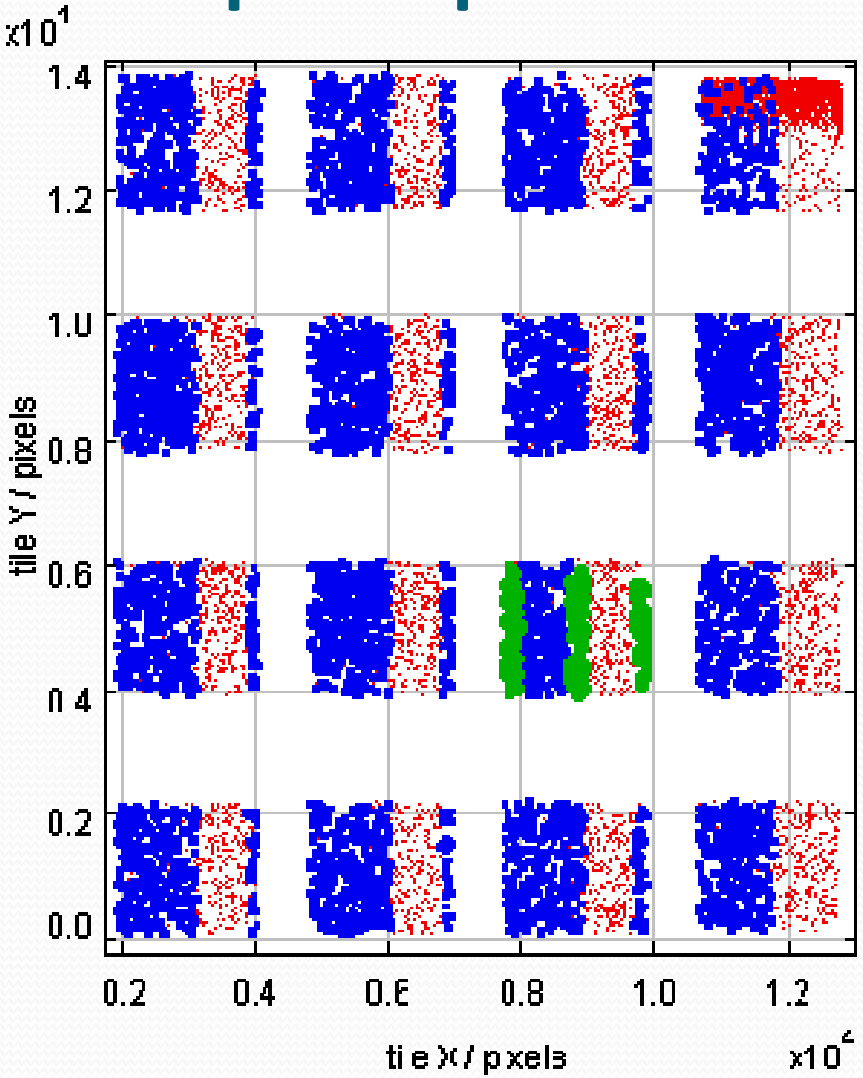
Pawprint: 16 detectors spaced 90% apart in X direction and 42% in Y

6 pawprints can make a tile: 2x depth on average, but “ears” have single depth.





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

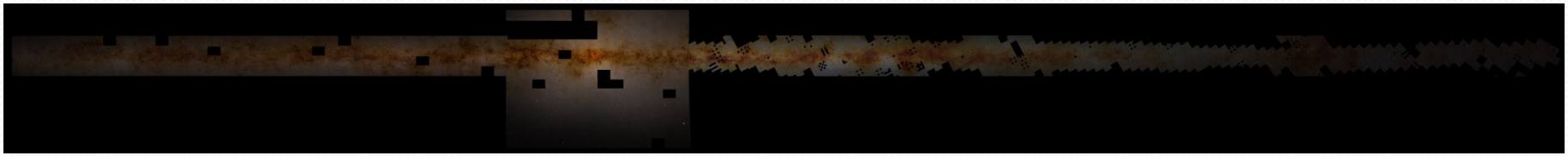
- Images:
  - WFCAM 1720 raw image frames a day (4 2kx2k)
  - VISTA 580 raw image frames a day (16 2kx2k)
  - Only 30% increase in raw image volume.
- Catalogues:
  - More area (2.9x), increased sensitivity (~2x), tile + pawprints (~3x). Expect 15-20x as many detections.

- **14X** WSA: 3.2M detections per day, VSA: 44M
- Catalogues are important factor for relational DB
- VVV: VSA currently has >21billion rows, expect **~10^{11}**

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:

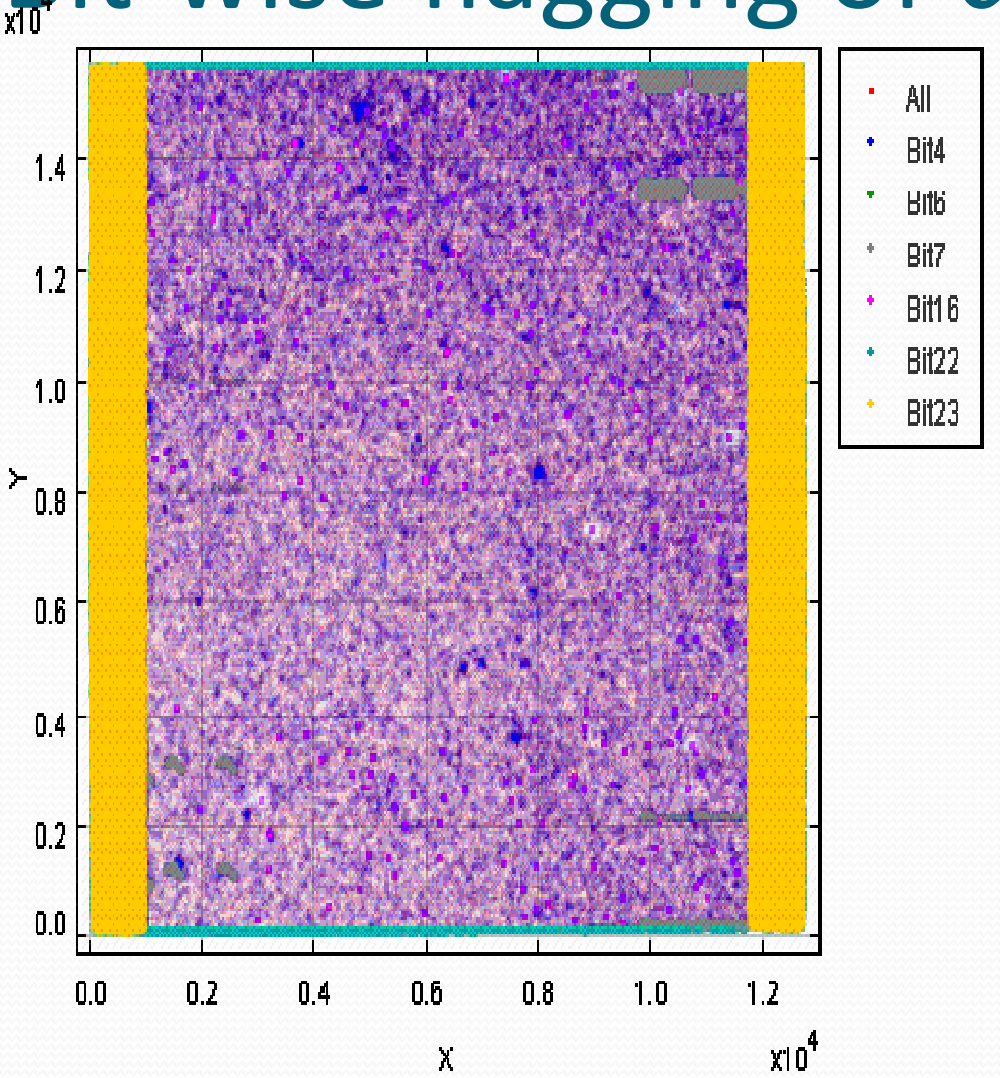
vsa-support@roe.ac.uk

- 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.

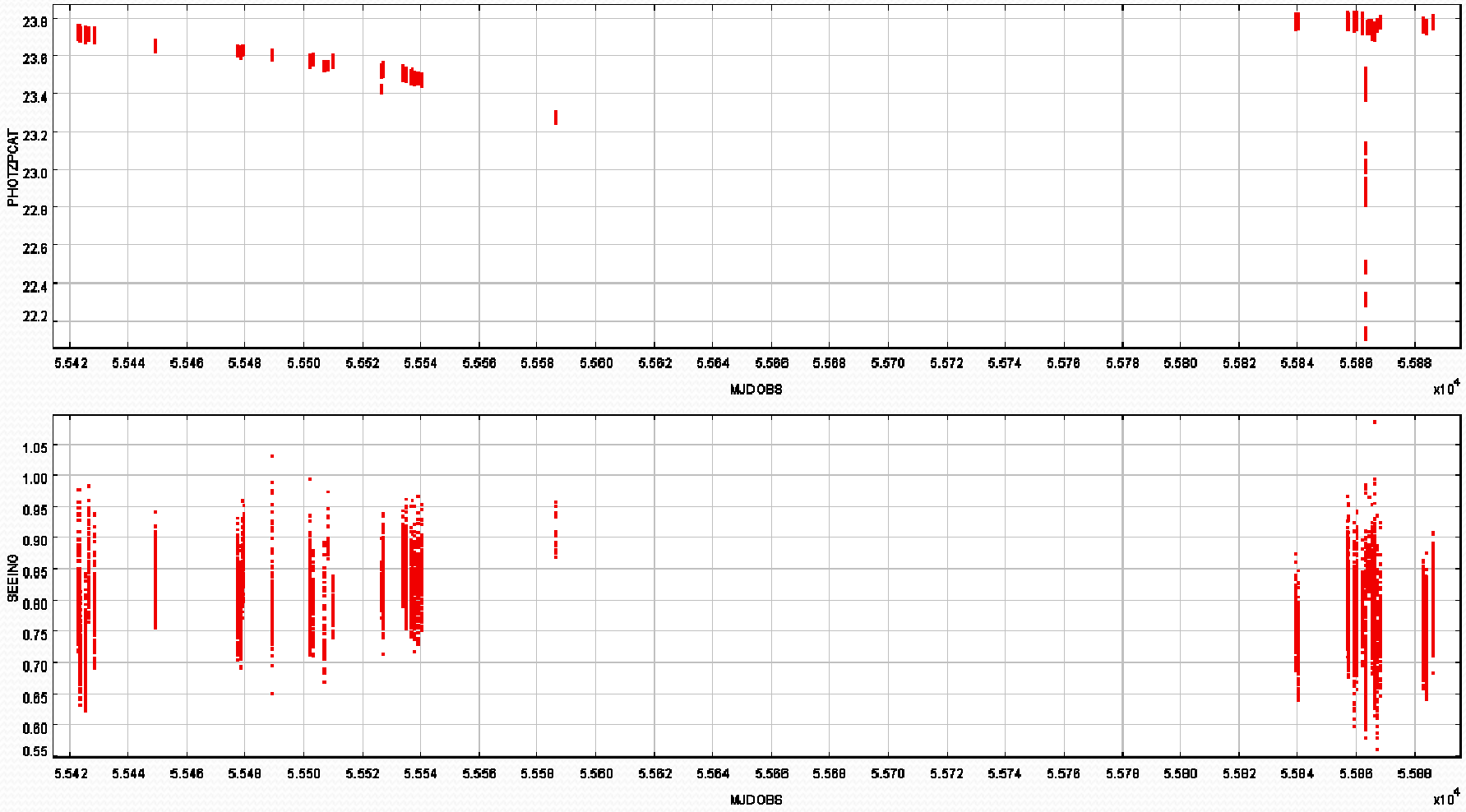
Bit-wise flagging of detection data



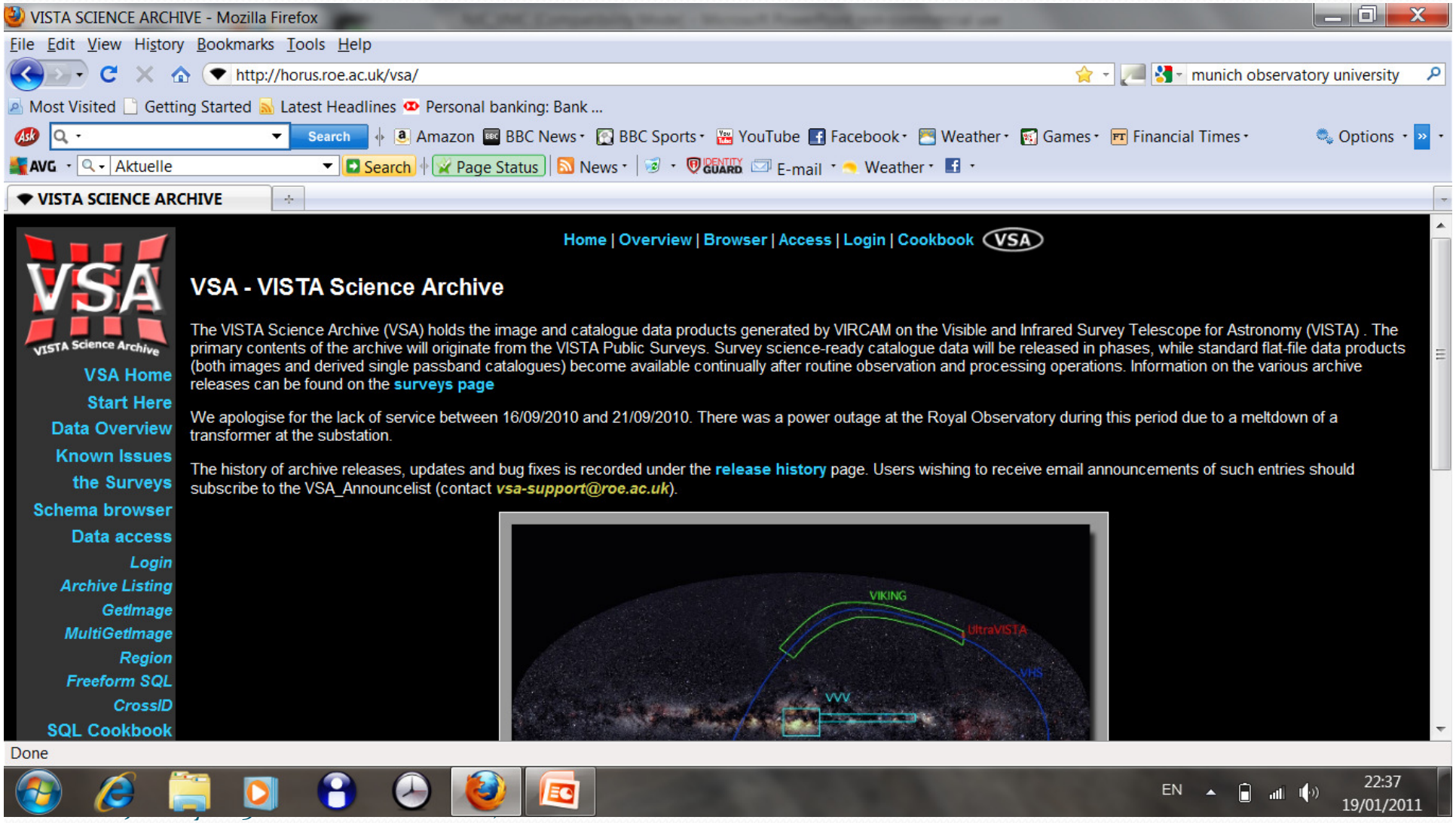
VMC deep tile.

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

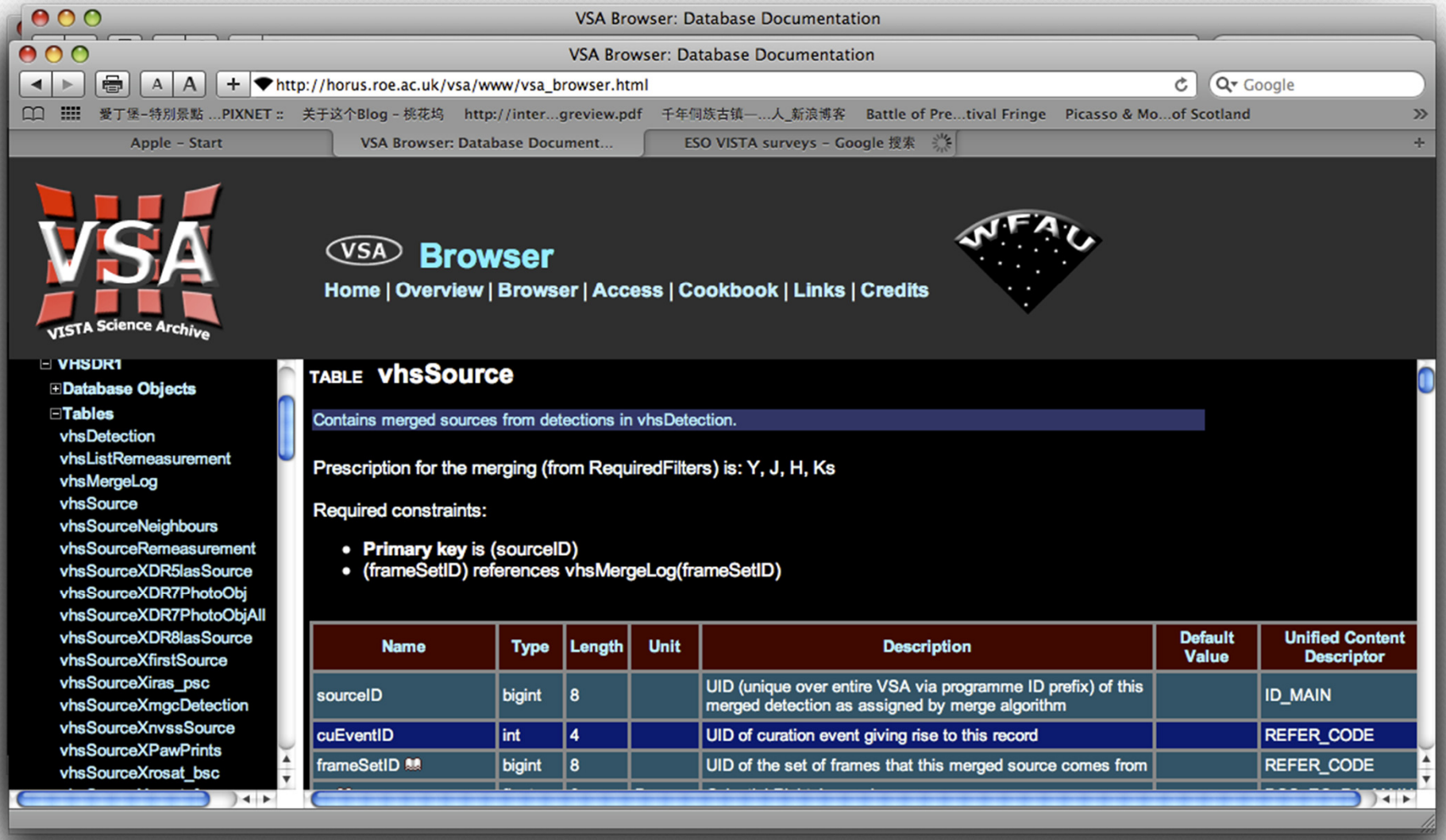
Provenance: ZP and seeing of input files to a VIDEO mosaic wrt time

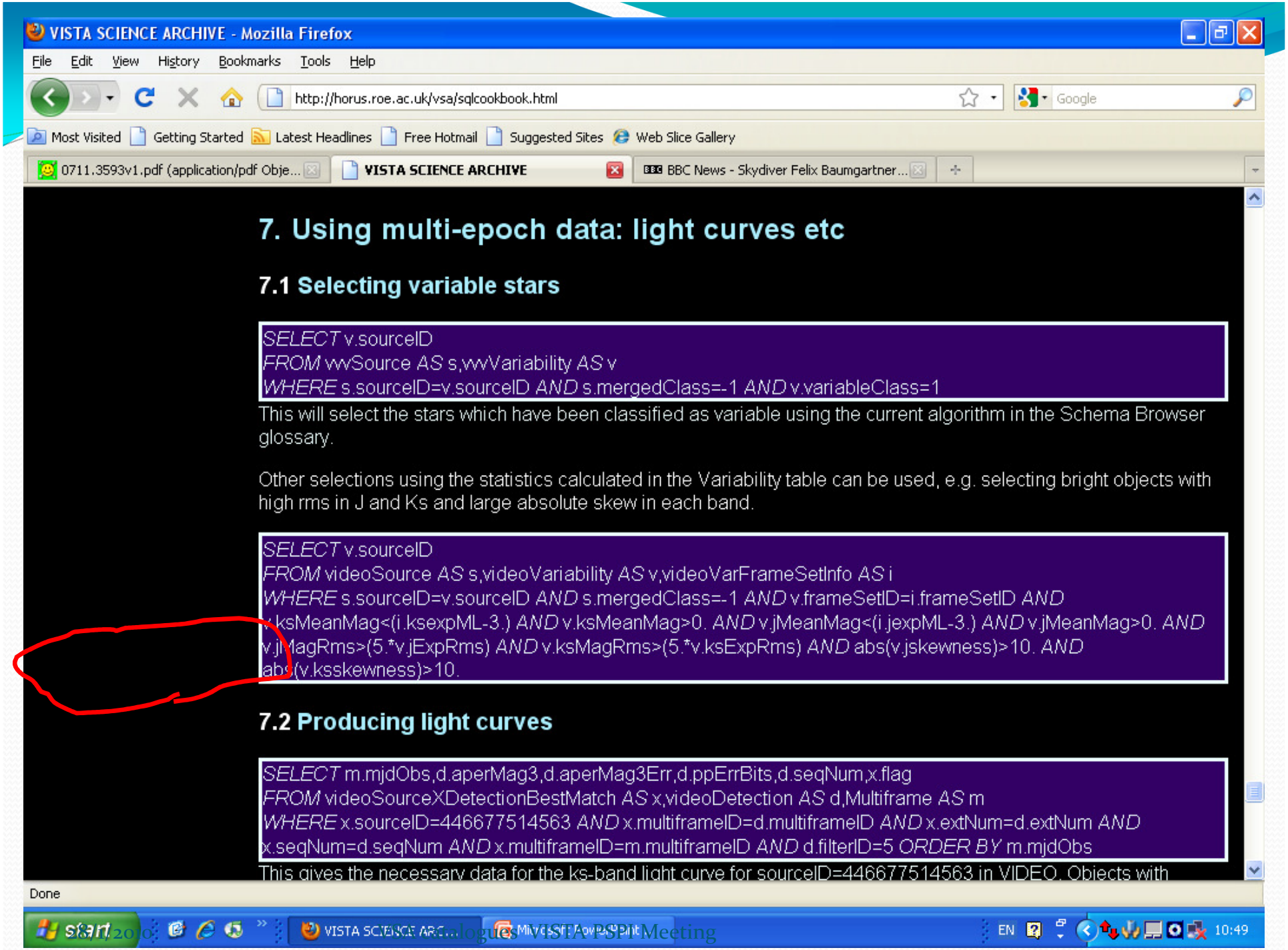


Using the VSA

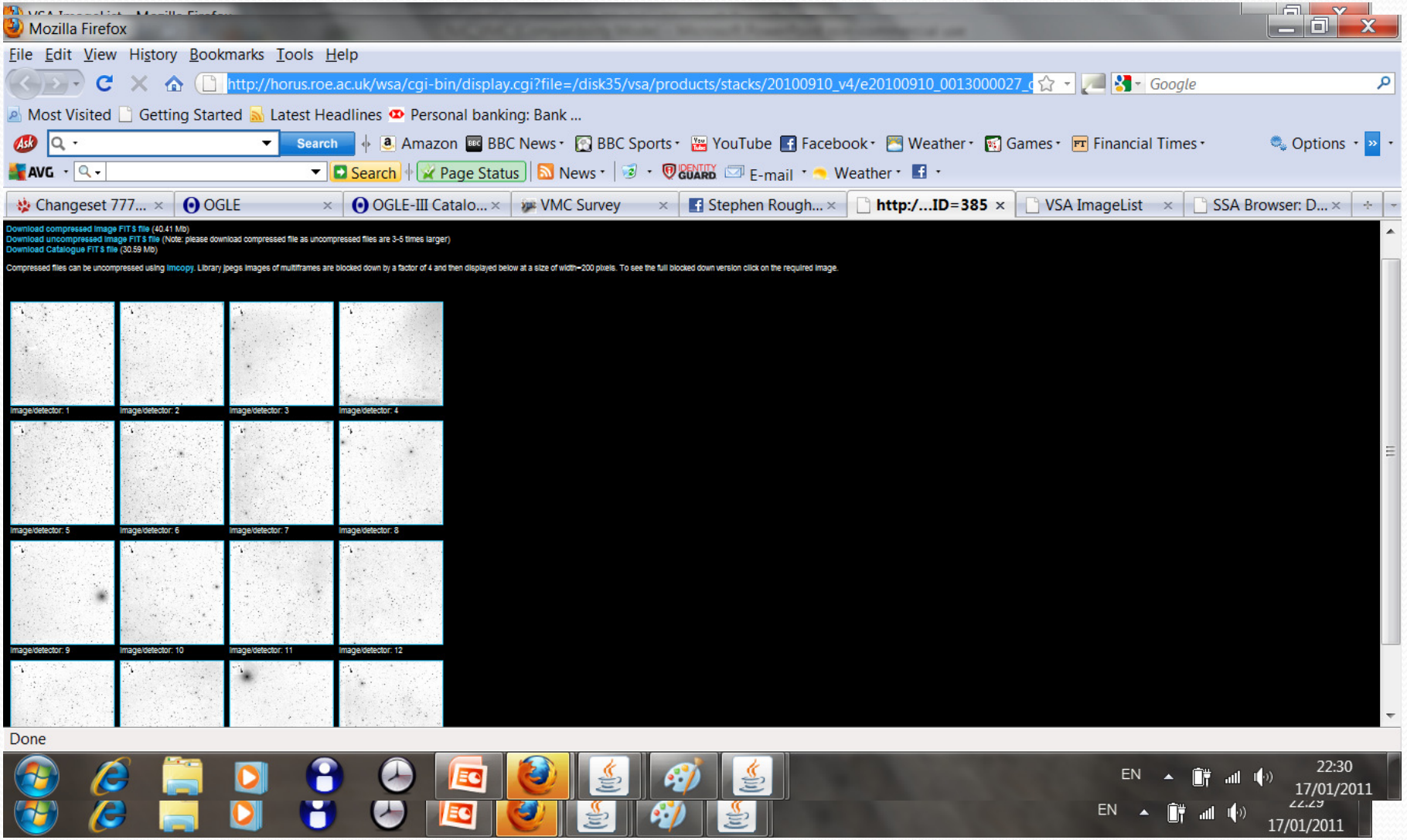


VSA – schema browser

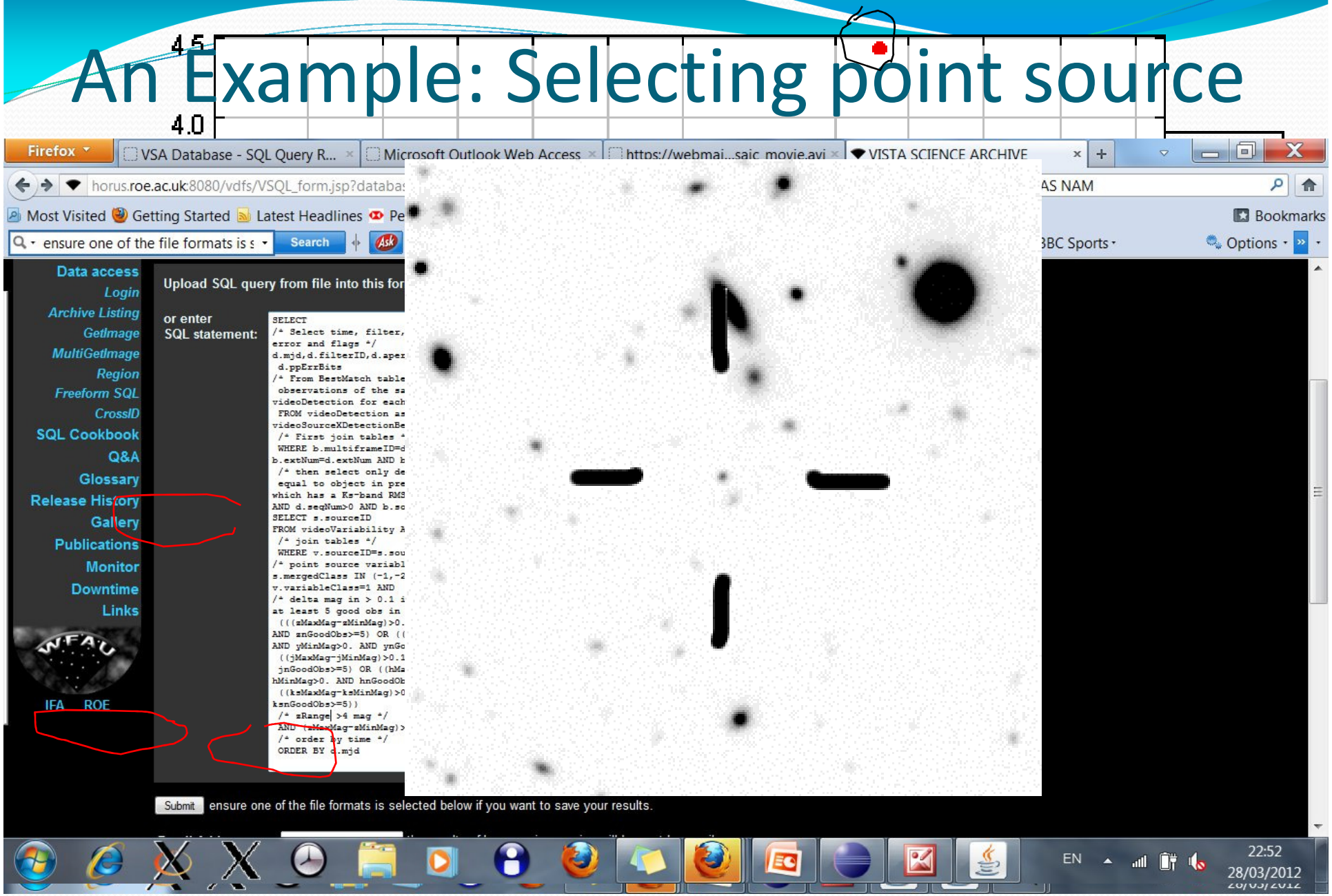



VISTA SCIENCE ARCHIVE - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://horus.roe.ac.uk/vsa/sqlcookbook.html
Most Visited Getting Started Latest Headlines Free Hotmail Suggested Sites Web Slice Gallery
0711.3593v1.pdf (application/pdf Obje...
VISTA SCIENCE ARCHIVE
BBC News - Skydiver Felix Baumgartner...
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
v.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
Done
start
VISTA SCIENCE ARC...
Microsoft PowerPoint
EN
10:49

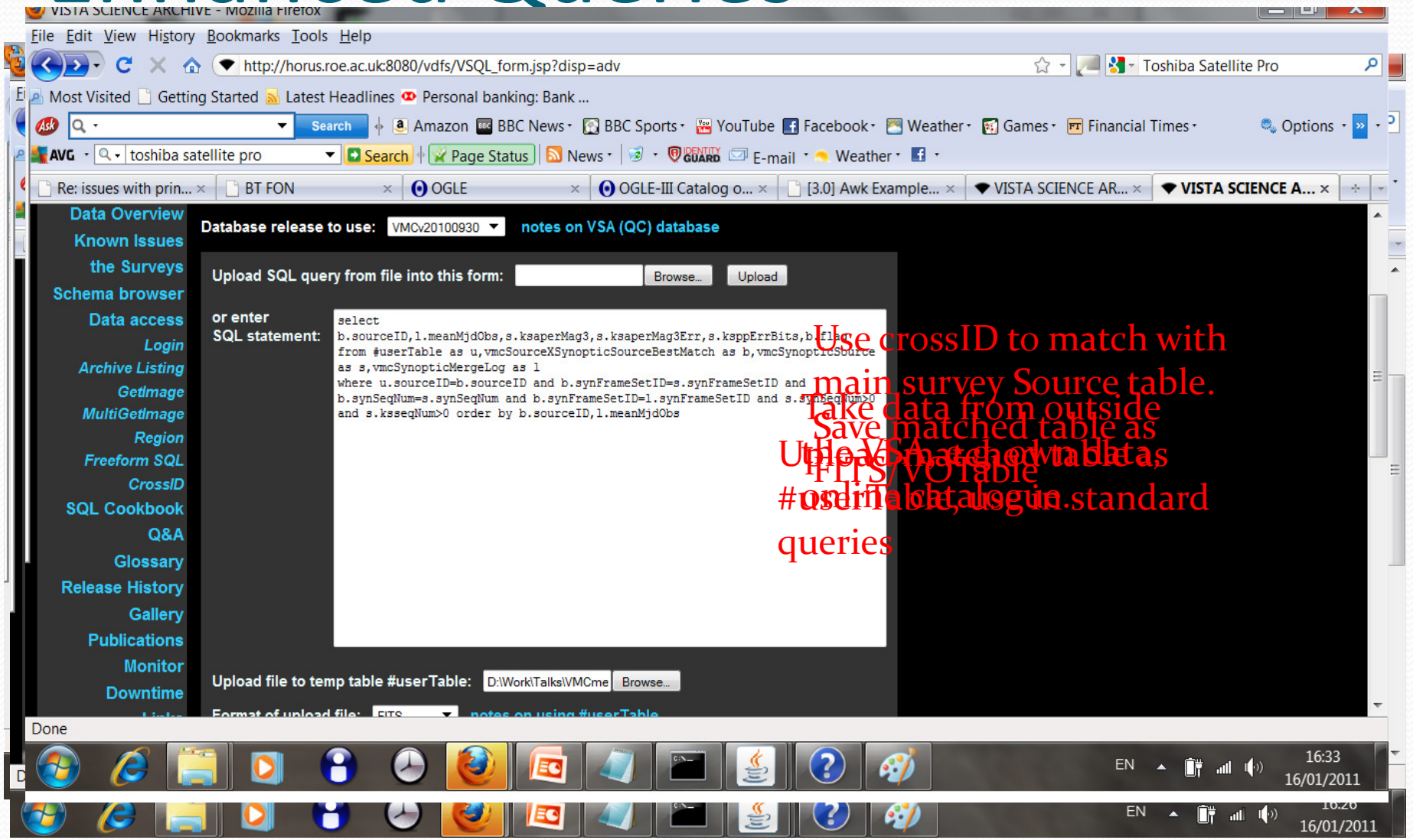
Archive Listing



An Example: Selecting point source



Enhanced Queries



Use crossID to match with main survey Source table.

Take data from outside

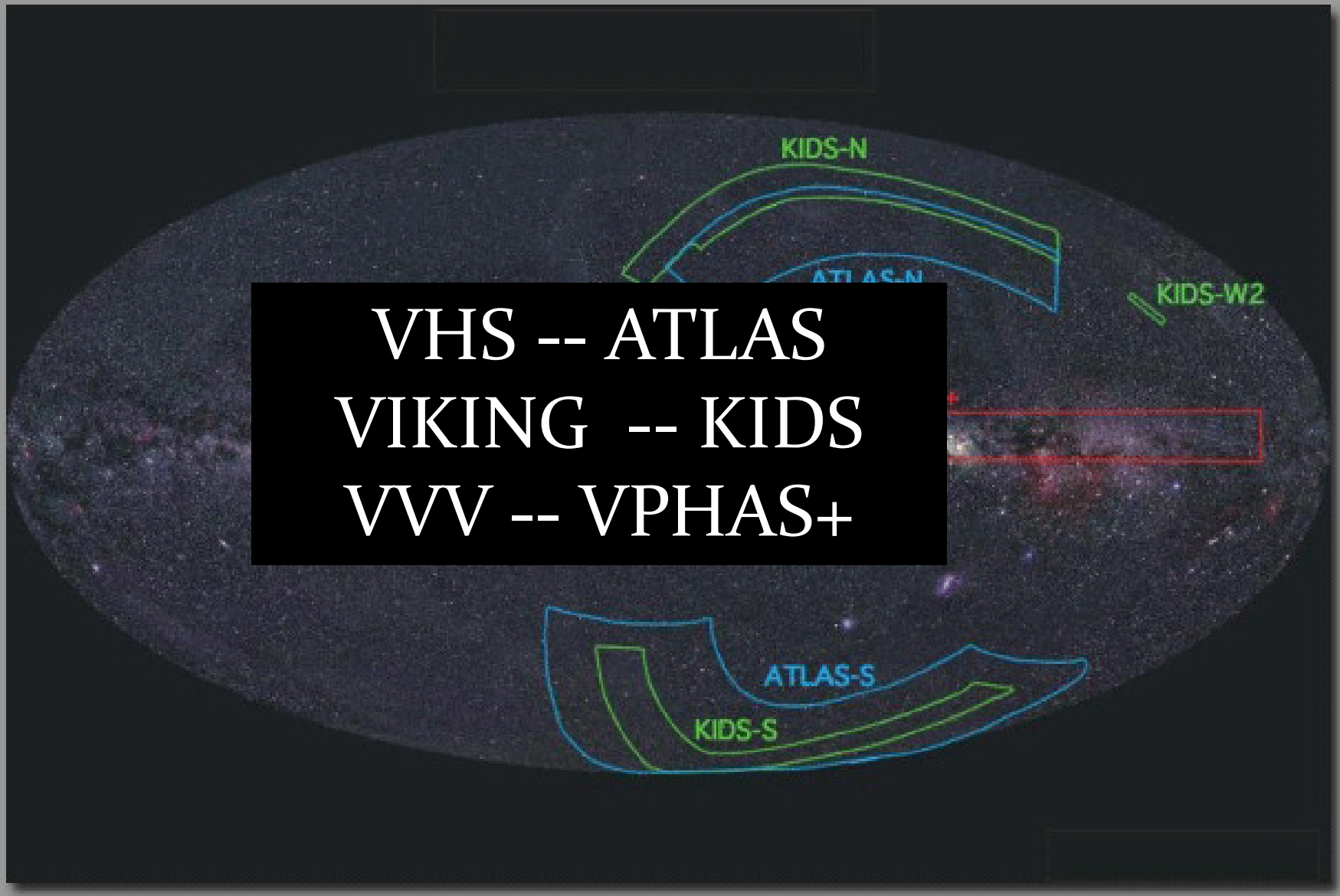
Save matched table as #userTable, log in

Use VSA to download data as FITS/VOTable

Use #userTable in 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


KIDS-N
KIDS-W2
VHS -- ATLAS
VIKING -- KIDS
VVV -- VPHAS+
ATLAS-S
KIDS-S

EARLIER PROCESSING AT CASU

WFAU

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.