Big Surveys, Big Databases

how to avoid drowning in the data deluge

ROYAL OBSERVATORY EDINBURGH

- big survey bottlenecks
- UKIDSS
- demo
- cultural change

Big Survey Bottlenecks

big survey science

- key science goals need big surveys
  - statistics : eg DM, DE mapping
  - large structures : eg Galactic Archaeology
  - rare objects : eg z=10 QSOs, NEOs, free floating planets
- and/or data intensive computing
  - N**2 calcns
  - monitoring; fast alerts (LSST, SKA,GRBs)
  - operations : MCAO, correlators

What is a survey ?

- Includes any big data collection
  - images; catalogues; spectra; event files; fringes; light curves etc
- Two step process
  - collect data, summarise, archive
  - do science with the archive
- Why do it this way ?
  - some science needs that much data...
  - surprises
  - many experiments with same data
  - multi-lamda resources

scary data ?

- 2008 : 20 TB/yr (UKIDSS)
- 2009 : 100TB/yr (VISTA)
- 2015 : 5PB/yr (LSST)
- 2020 : 100PB/yr (SKA)

- need to process, document, and store at professional data centres

bottlenecks

- end user b/w and disk-cpu b/w
  do not scale with Moore's law
- downloading 1TB : all week
- searching 1TB : ditto

download the results not the data

"download and hack" doesn't work

==> analyse in situ

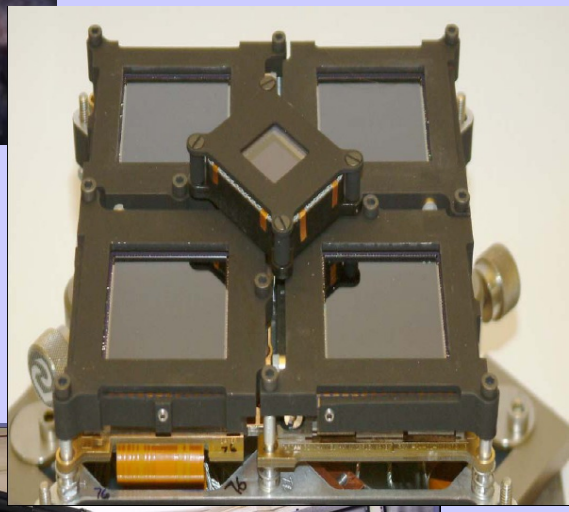
==> data centres must hardware and tools

service economy

UKIDSS

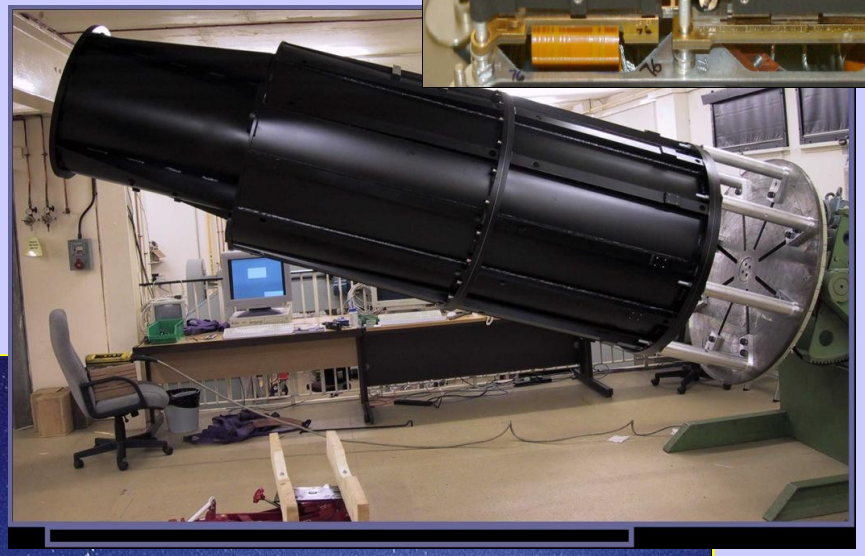
UKIDSS

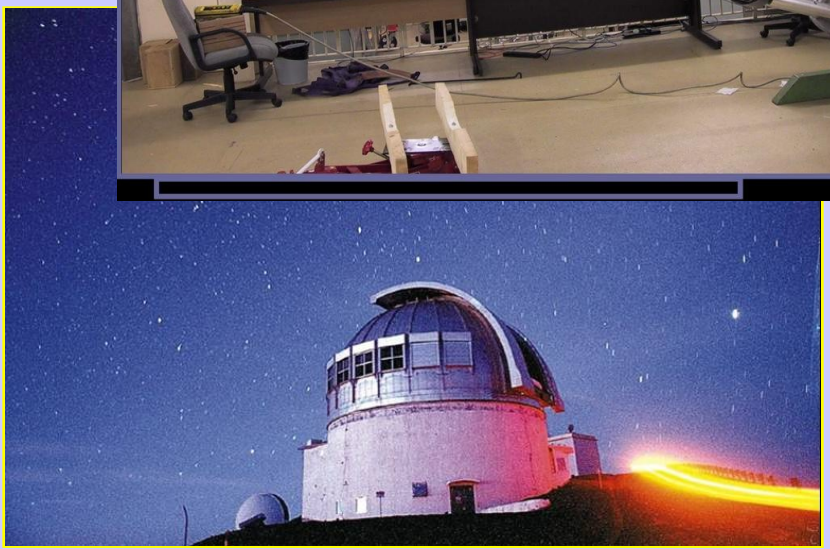
Lawrence *et al* 2007



- ESO public survey
- uses UKIRT Wide Field Camera (WFCAM)
- 1200 nights over 8yrs
- UKIDSS = 20 X 2MASS volume
- near-ir SDSS
- began 2005 May 13
- processed by CASU/WFAU
- data available at

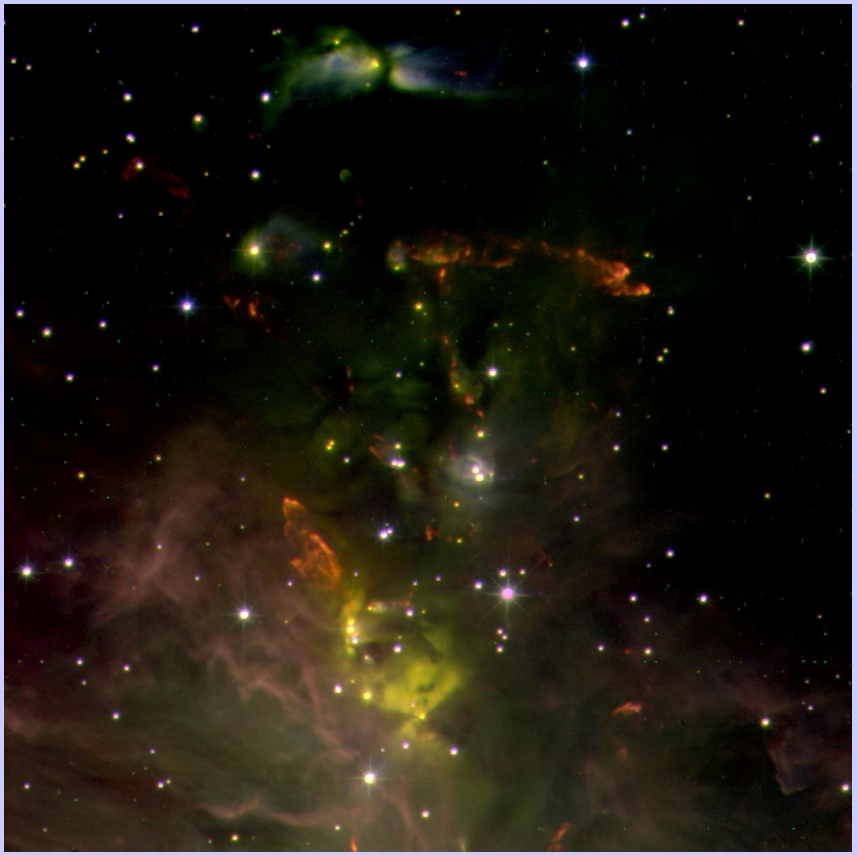
http://surveys.roe.ac.uk/wsa



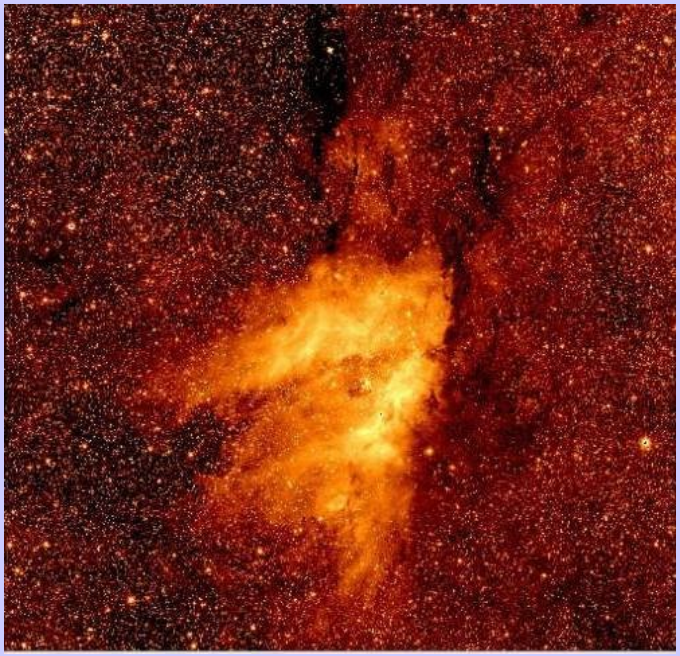


WFCAM pix

NGC 891



ORION

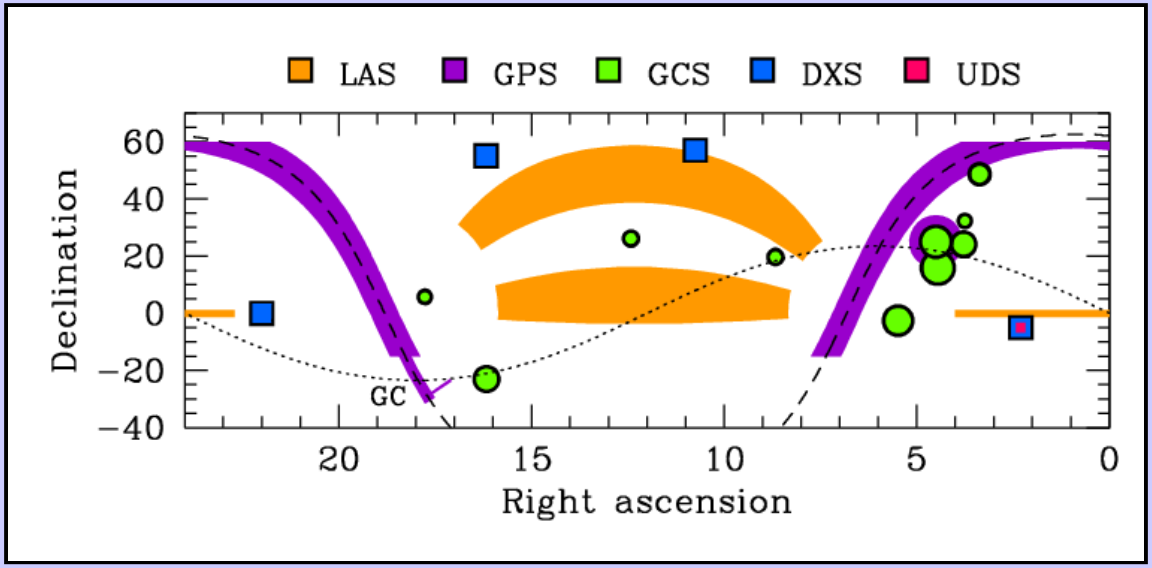


M17

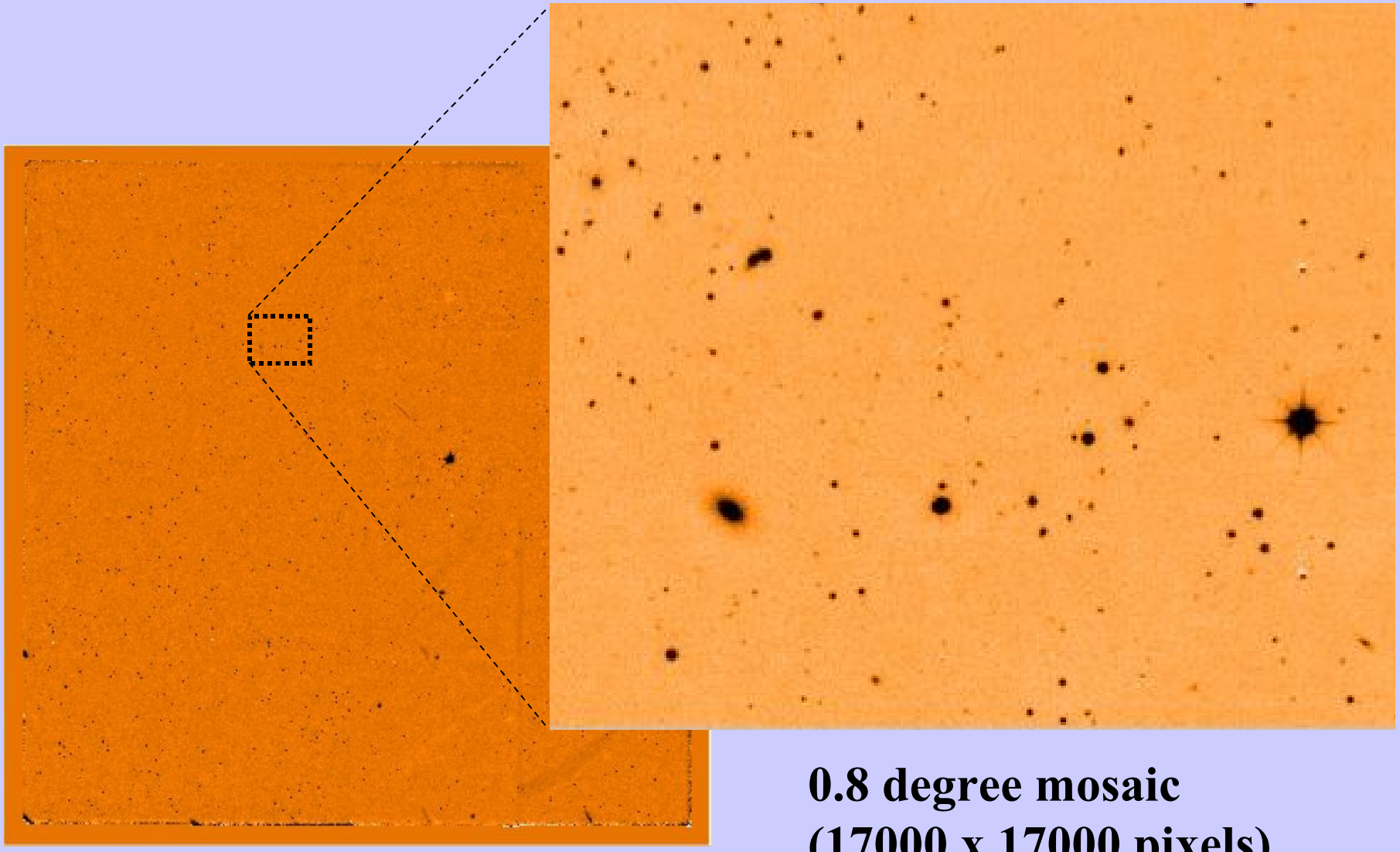
M104

UKIDSS design

Large Area Survey	LAS	YJHK	18.2K	4028 s.d.	262n	ExGal
Deep Extragalactic Survey	DXS	JK	20.8	35	118	ExGal
Ultra Deep Survey	UDS	JHK	22.8	0.77	296	ExGal
Galactic Plane Survey	GPS	JHK	19.0	1868	186	Gal
Galactic Clusters Survey	GCS	ZYJHK	18.6	1067	84	Gal



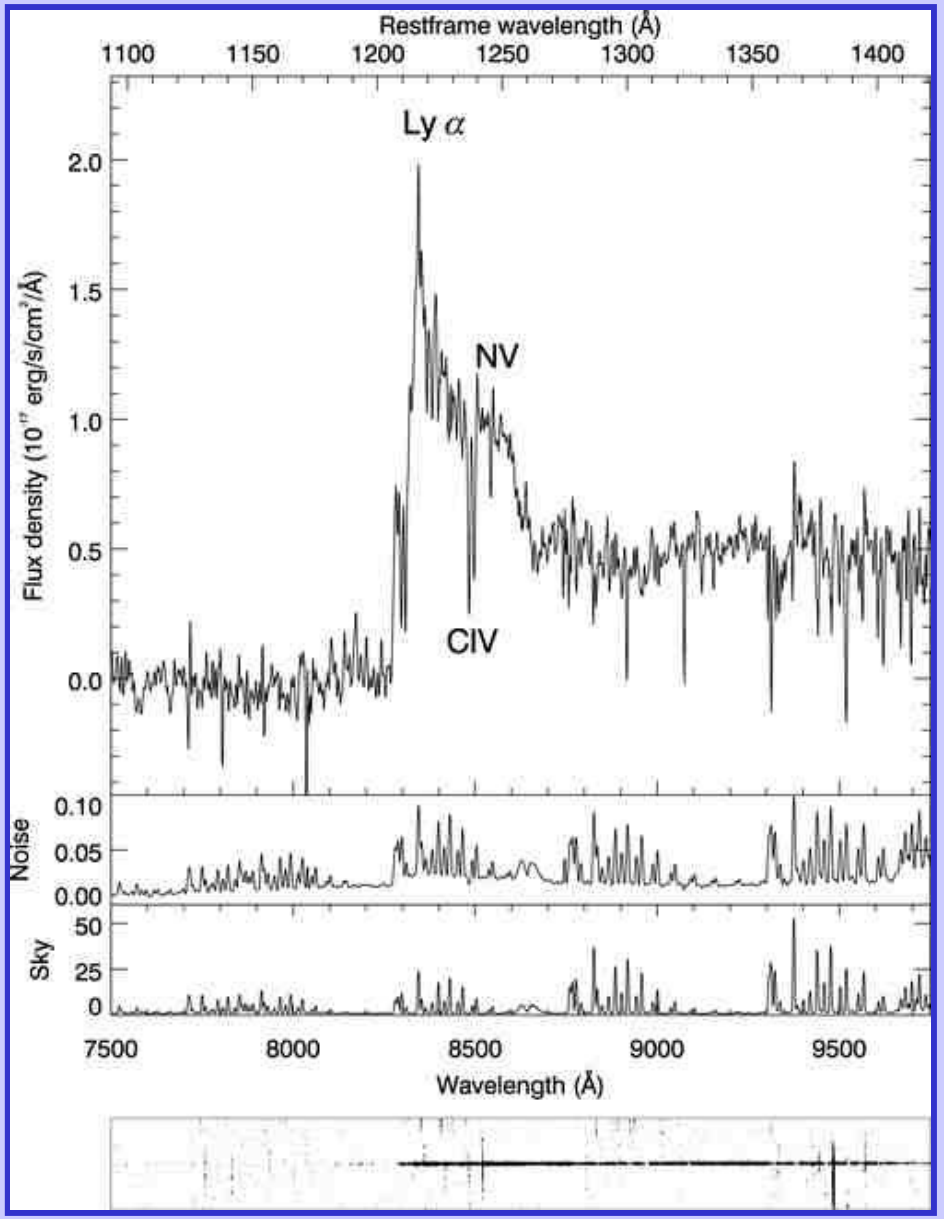
scary amounts of data



0.8 degree mosaic
(17000 x 17000 pixels)

z=6 quasar

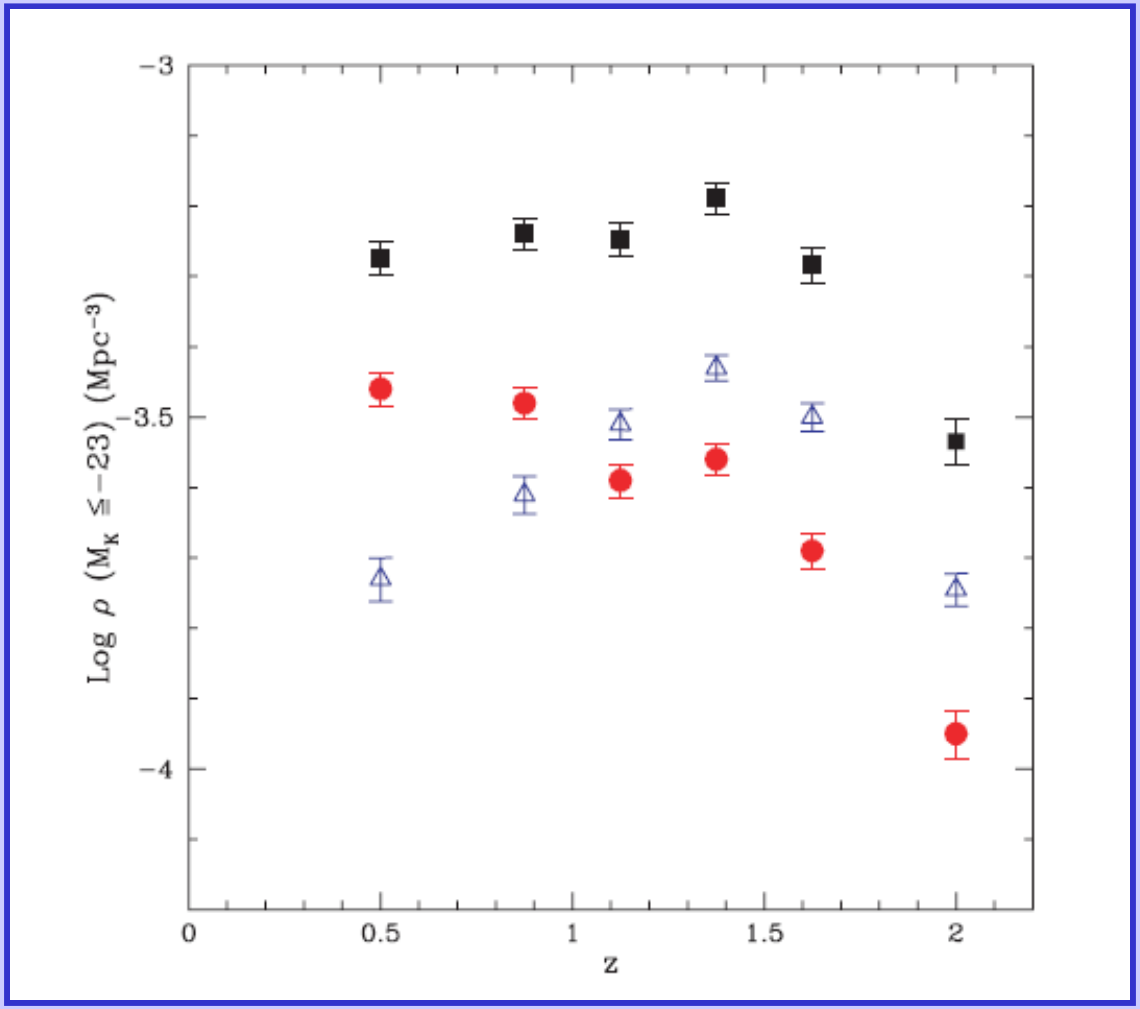
- ULAS J0203+0012
- z=5.86
- From DR1
  - only 106 sq.deg.



Venemans et al 2007

z=2 evolution

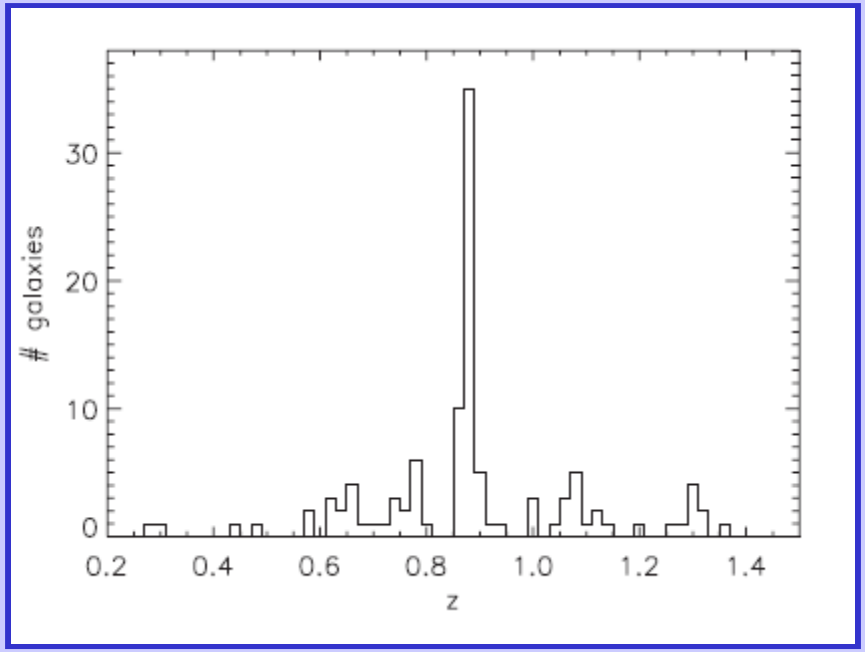
- Blue *vs* red luminous gals evolve differently
- UDS EDR data



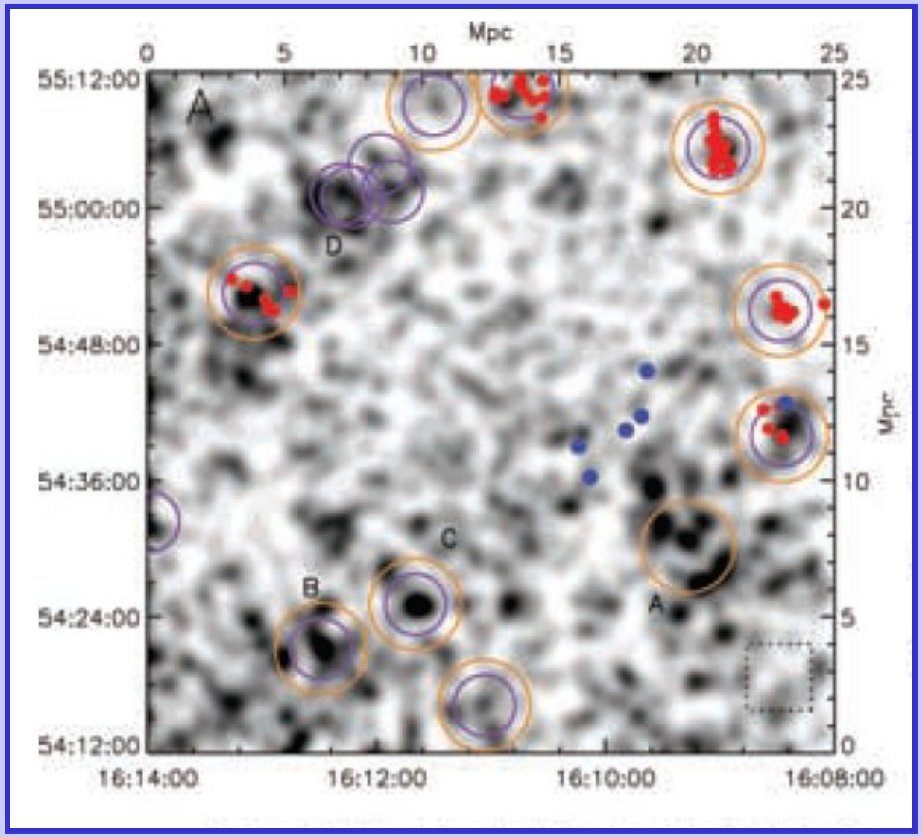
Cirasuolo et al 2007

z=1 supercluster

Swinbank et al 2007



redshift distbn of candidate clusters

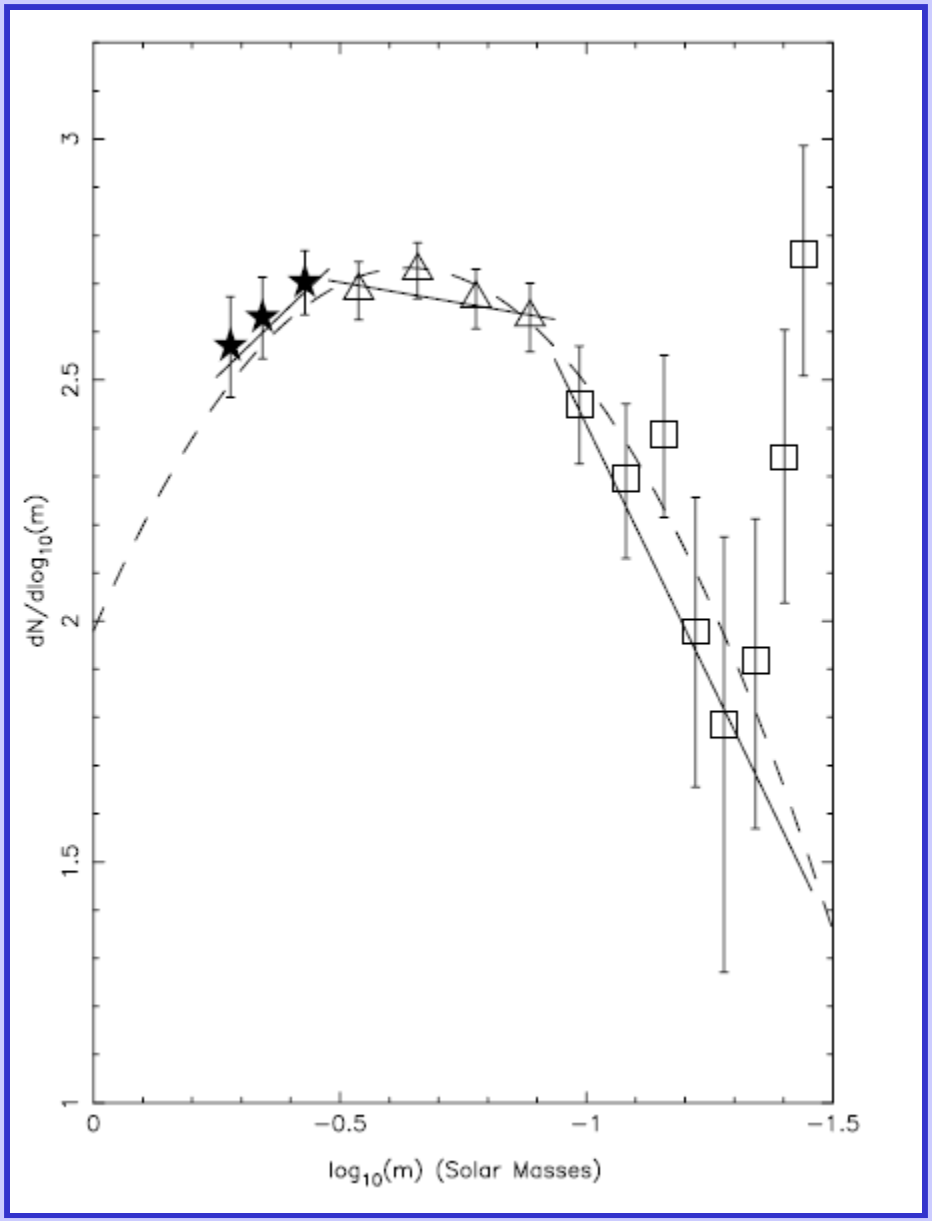


structure 30 Mpc across

colour selected surface density map

134pc substellar MF

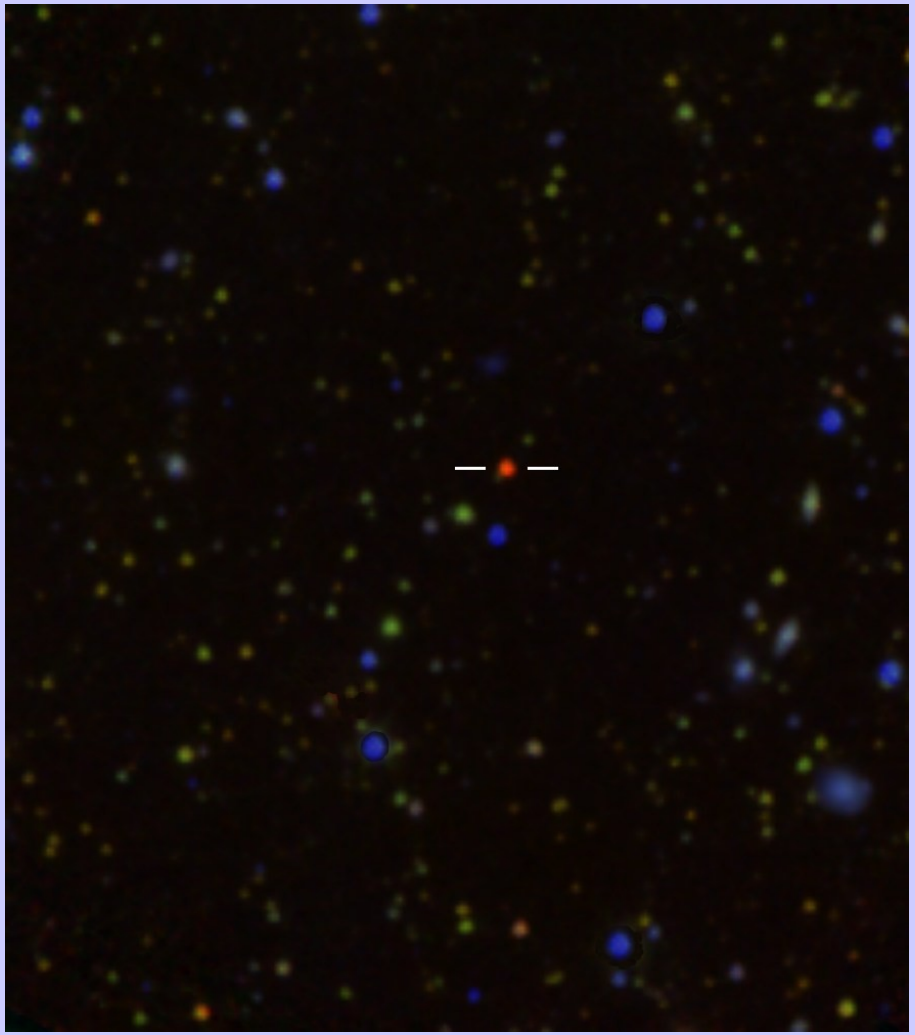
- Pleiades GCS-DR1
- 5 band selection
- 73 new BDs
- MF Gaussian



Lodieu et al 2007

20 pc Brown Dwarf

- ULAS J0034-00
- Coolest known dwarf (T8.5)
- T~600K
- M~15-36 M-Jup



blue = Z
green = Spitzer 3.6um
red = Spitzer 4.5 um

Warren et al 2007

Demo

WSA Home
Start Here
Data Overview
Known Issues
the Surveys
Schema browser
Data access
Login
Archive Listing
GetImage
MultiGetImage
Region

Region search

Use this form to search around a given position or object name. For help on using this form see **region help**.

Database release to use: UKIDSSDR1PLUS

Choose the programme/survey & table you wish to search: UKIDSS Large Area Survey, LAS

source table (merged catalogue)

RA or Galactic Long.: 355.0
Dec or Galactic Lat.: 0.00
Coordinate System: J2000

WFAU

- SQL Query Results

tabase
:46 BST 2008 [2 active, 25 total]

and wait for your results to appear below...

/Lat: 0.00 Coord. Sys (B1950,J2000 or Galactic): J

isplay 1 arcmin image cut-outs around the RA/Dec of the object.

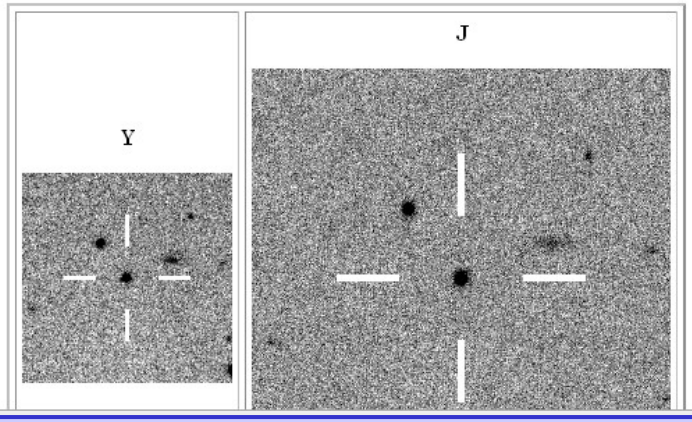
frameSetID	ra	dec	sigRa	sigDec	epoch
33791702333	+354.9942751	-0.0567320	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0068879	-0.0561711	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0516556	-0.0550098	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+354.9430954	-0.0549847	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+354.9708035	-0.0514278	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0430845	-0.0512383	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+354.9608219	-0.0510039	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0276330	-0.0519756	-9.999995E+008	-9.999995E+008	-9.9999950E+008

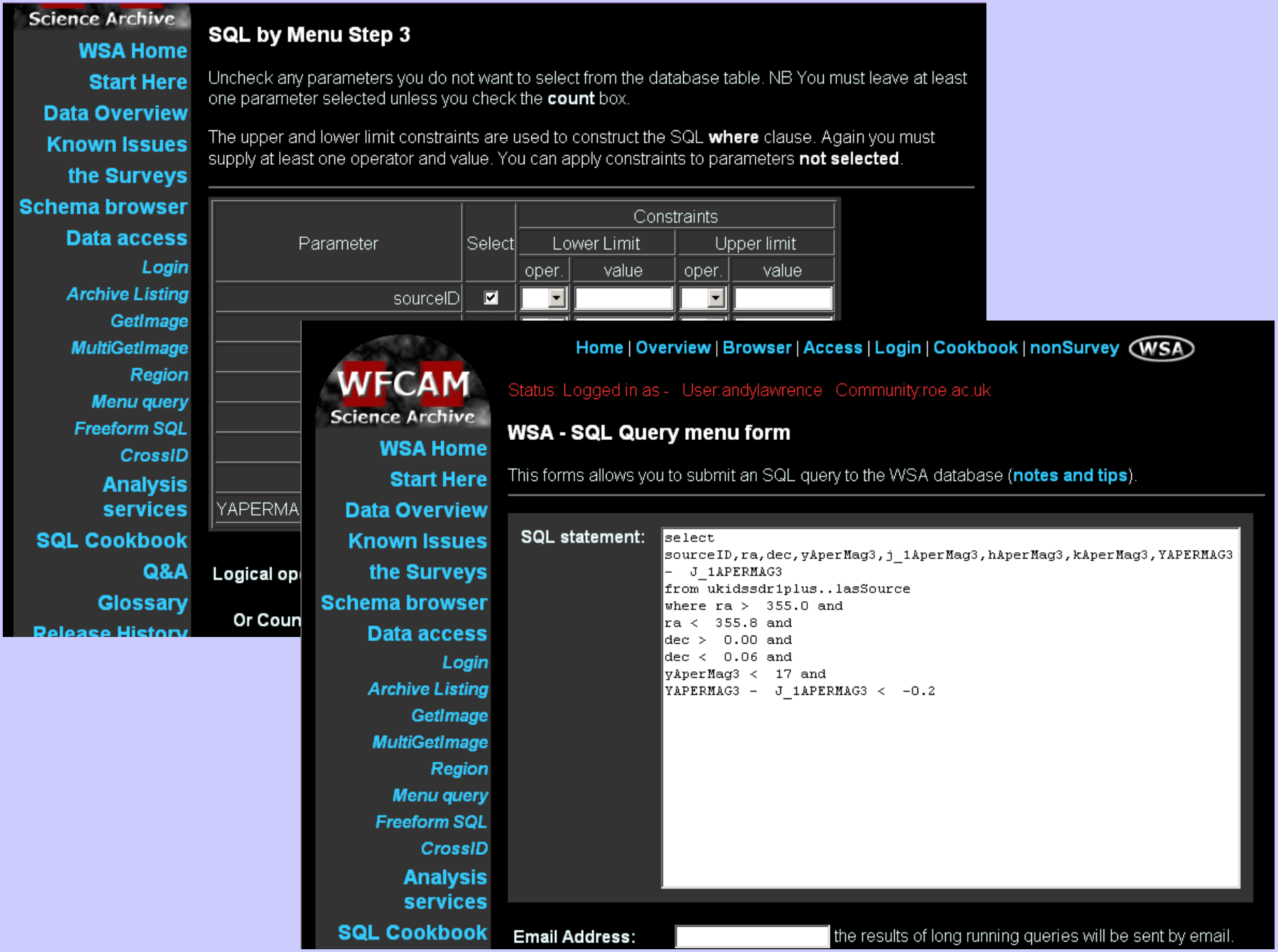
GetImage cut-out results

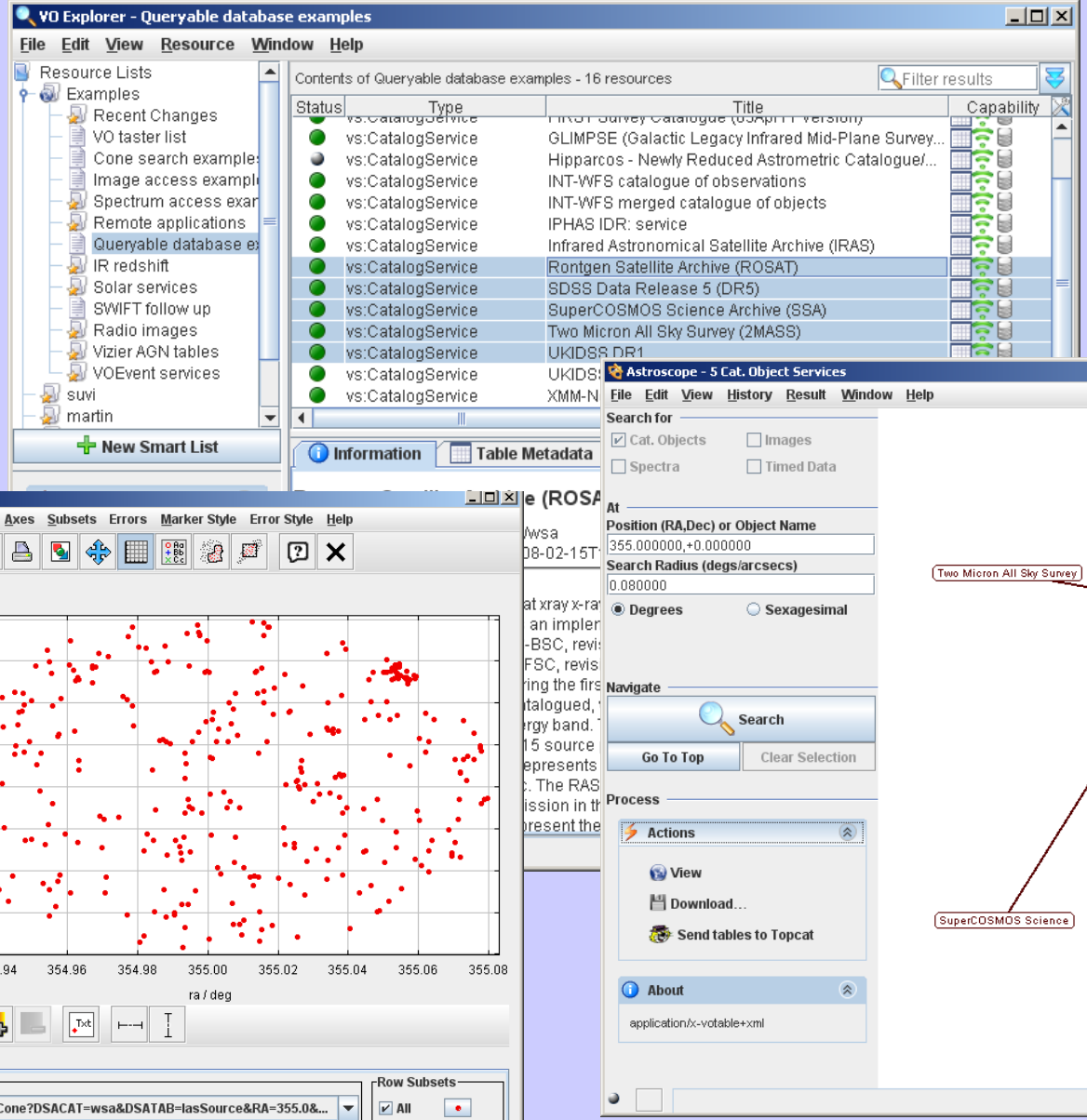
J2000 coords: RA: 232.5028291 Dec:6.919786
Programme: UKIDSS Large Area Survey, LAS
Filter: all
Connecting to database: UKIDSSDR2PLUS

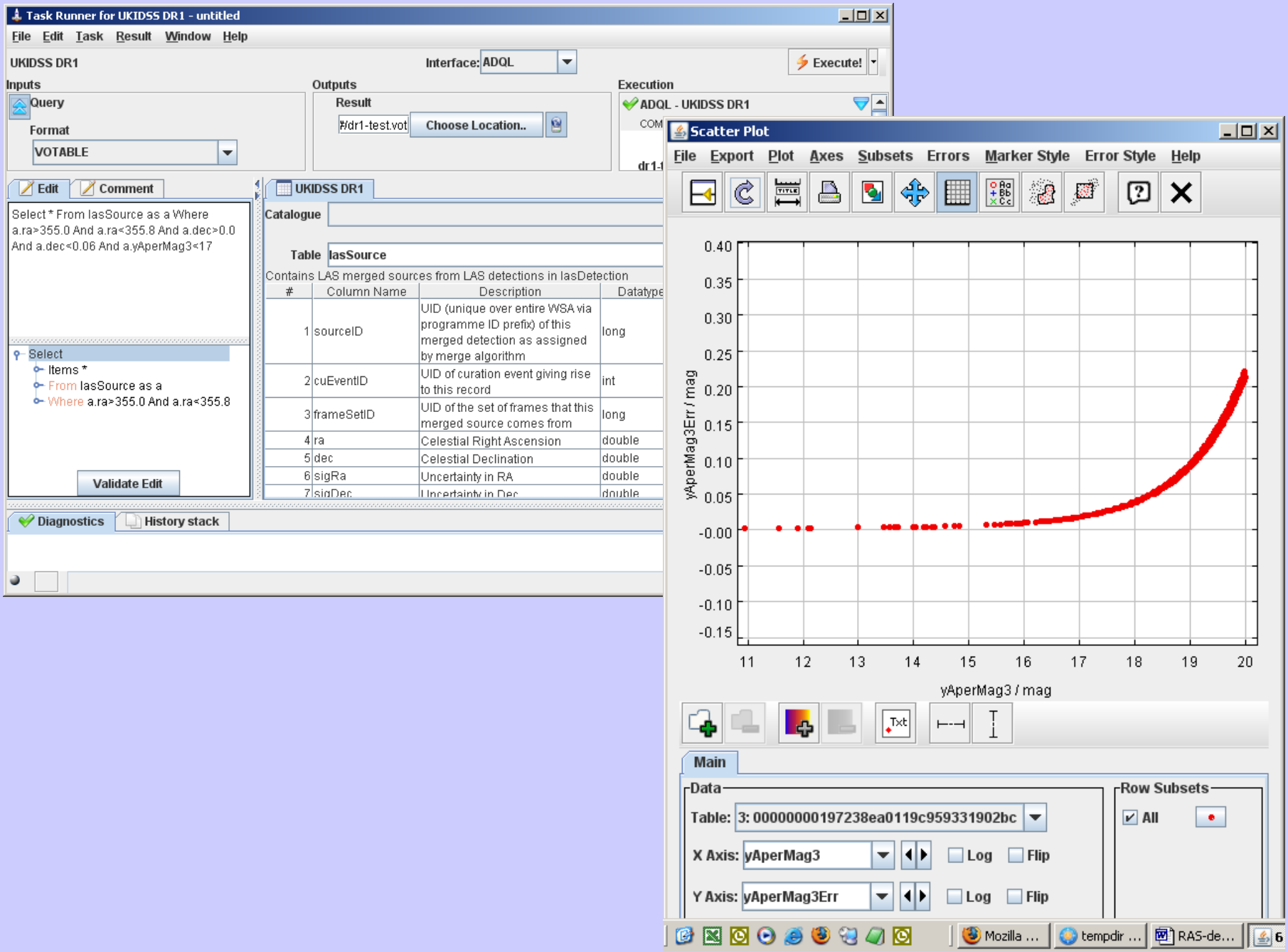
Link	multiframeID	frametype	obstype	filterid	shortname	dateObs	extNum
show	976960	stack	OBJECT	2	Y	2006-06-10 09:06:14.0	5
show	988076	leavstack	OBJECT	3	J	2006-06-10 09:30:23.7	5
show	987086	stack	OBJECT	4	H	2006-06-10 09:54:50.0	5
show	983032	stack	OBJECT	5	K	2006-06-10 10:19:17.1	5

4 rows returned.




Science Archive
WSA Home
Start Here
Data Overview
Known Issues
the Surveys
Schema browser
Data access
Login
Archive Listing
GetImage
MultiGetImage
Region
Menu query
Freeform SQL
CrossID
Analysis services
SQL Cookbook
Q&A
Glossary
Release History
SQL by Menu Step 3
Uncheck any parameters you do not want to select from the database table. NB You must leave at least one parameter selected unless you check the count box.
The upper and lower limit constraints are used to construct the SQL where clause. Again you must supply at least one operator and value. You can apply constraints to parameters not selected.
Parameter
Select
Constraints
Lower Limit
Upper limit
oper.
value
oper.
value
sourceID
YAPERMA
Logical op
Or Coun
WFCAM
Science Archive
WSA Home
Start Here
Data Overview
Known Issues
the Surveys
Schema browser
Data access
Login
Archive Listing
GetImage
MultiGetImage
Region
Menu query
Freeform SQL
CrossID
Analysis services
SQL Cookbook
Home | Overview | Browser | Access | Login | Cookbook | nonSurvey
WSA
Status: Logged in as - User:andylawrence Community:roe.ac.uk
WSA - SQL Query menu form
This forms allows you to submit an SQL query to the WSA database (notes and tips).
SQL statement:
select
sourceID,ra,dec,yAperMag3,j_1AperMag3,hAperMag3,kAperMag3,YAPERMAG3 - J_1APERMAG3
from ukidssdr1plus..lasSource
where ra > 355.0 and
ra < 355.8 and
dec > 0.00 and
dec < 0.06 and
yAperMag3 < 17 and
YAPERMAG3 - J_1APERMAG3 < -0.2
Email Address:
the results of long running queries will be sent by email.


VO Explorer - Queryable database examples
File Edit View Resource Window Help
Resource Lists
Examples
Recent Changes
VO taster list
Cone search example
Image access exampl
Spectrum access exar
Remote applications
Queryable database ex
IR redshift
Solar services
SWIFT follow up
Radio images
Vizier AGN tables
VOEvent services
suvi
martin
New Smart List
Contents of Queryable database examples - 16 resources
Filter results
Status
Type
Title
Capability
vs:CatalogService
GLIMPSE (Galactic Legacy Infrared Mid-Plane Survey...
Hipparcos - Newly Reduced Astrometric Catalogue/...
INT-WFS catalogue of observations
INT-WFS merged catalogue of objects
IPHAS IDR: service
Infrared Astronomical Satellite Archive (IRAS)
Rontgen Satellite Archive (ROSAT)
SDSS Data Release 5 (DR5)
SuperCOSMOS Science Archive (SSA)
Two Micron All Sky Survey (2MASS)
UKIDSS DR1
Information
Table Metadata
Astroscope - 5 Cat. Object Services
File Edit View History Result Window Help
Search for
Cat. Objects
Images
Spectra
Timed Data
At
Position (RA,Dec) or Object Name
355.000000,+0.000000
Search Radius (degs/arcsecs)
0.080000
Degrees
Sexagesimal
Navigate
Search
Go To Top
Clear Selection
Process
Actions
View
Download...
Send tables to Topcat
About
application/x-votable+xml
SDSS Data Release 5 (DR5) - 651 results
Two Micron All Sky Survey
Cat. Objects
UKIDSS DR1 - 1 search
Search Results
wsa, lasSource: cone search - 351 results
SuperCOSMOS Science
Scatter Plot
File Export Plot Axes Subsets Errors Marker Style Error Style Help
x10-2
dec / deg
ra / deg
354.92
354.94
354.96
354.98
355.00
355.02
355.04
355.06
355.08
Main
Data
Table: 3: SubmitCone?DSACAT=wsa&DSATAB=lasSource&RA=355.0&...
X Axis: ra
Y Axis: dec
Log
Flip
Row Subsets
All
Potential: 351 Included: 351 Visible: 351
Position:


Task Runner for UKIDSS DR1 - untitled
File Edit Task Result Window Help
UKIDSS DR1
Interface: ADQL
Execute!
Inputs
Query
Format
VOTABLE
Outputs
Result
#/dr1-test.vot
Choose Location..
Execution
ADQL - UKIDSS DR1
Edit
Comment
Select * From lasSource as a Where a.ra>355.0 And a.ra<355.8 And a.dec>0.0 And a.dec<0.06 And a.yAperMag3<17
Select
Items *
From lasSource as a
Where a.ra>355.0 And a.ra<355.8
Validate Edit
UKIDSS DR1
Catalogue
Table lasSource
Contains LAS merged sources from LAS detections in lasDetection
Column Name Description Datatype
1 sourceID UID (unique over entire WSA via programme ID prefix) of this merged detection as assigned by merge algorithm long
2 cuEventID UID of curation event giving rise to this record int
3 frameSetID UID of the set of frames that this merged source comes from long
4 ra Celestial Right Ascension double
5 dec Celestial Declination double
6 sigRa Uncertainty in RA double
7 sigDec Uncertainty in Dec double
Diagnostics
History stack
Scatter Plot
File Export Plot Axes Subsets Errors Marker Style Error Style Help
yAperMag3Err / mag
yAperMag3 / mag
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
-0.00
-0.05
-0.10
-0.15
11
12
13
14
15
16
17
18
19
20
Main
Data
Table: 3: 00000000197238ea0119c959331902bc
X Axis: yAperMag3
Log
Flip
Y Axis: yAperMag3Err
Log
Flip
Row Subsets
All
Mozilla ...
tempdir ...
RAS-de...
6

```
#!/usr/bin/python
"""
Sends a query to WFCAM Science Archive; saves result to file on local disk.

Usage:    python wsa_gps.py
will write a file named wsa_gps_res.vot to the current directory.
History:   20071212 Written by E. A. Gonzalez-Solares
"""

from time import sleep
from astrogrid import acr, DSA, MySpace

# Uncomment if automatic login is not enabled
# acr.login('ukidss')

# Define SQL here
# This query selects for each source, the x and y position in the detector as well as the
# size of the detector in which it was detected and the pixel scale. Only sources which are
# more than 10 arcsec away from the chip edges are returned in a search box
#
# NOTE: If the 'top 100' clause is removed then see below and save the output to a file in MySpace.
sql="""SELECT top 100
        s.sourceID, s.ra, s."dec", s.jmhPnt, s.pStar, s.pGalaxy, s.pNoise, s.pSaturated,
        s.jAperMag3, s.jAperMag3Err, s.jClass, s.hAperMag3, s.hAperMag3Err, s.hClass,
        s.k_1AperMag3, s.k_1AperMag3Err, s.k_1Class, d.x, d.y, m.xSize, m.ySize, c.xPixSize,
        c.yPixSize
    FROM
        gpsSource AS s, gpsDetection AS d, MultiframeDetector AS m, CurrentAstrometry AS c
    WHERE
        s.k_1ObjId = d.objID AND d.multiframeID = m.multiframeID AND d.extNum = m.extNum AND
        d.multiframeID = c.multiframeID AND d.extNum = c.extNum AND
        s.ra between 310.8 AND 313.0 AND s."dec" between 43.14 AND 44.0 AND
        d.x*c.xPixSize>10 AND d.y*c.yPixSize>10 AND
        (m.xSize-d.x)*c.xPixSize>10 AND (m.ySize-d.y)*c.yPixSize>10"""

# Define the enpoint service
dsa=DSA('ivo://wfau.roe.ac.uk/ukidssDR2-dsa/ceaApplication')

# Write all the SQL in one line
sql = ' '.join(sql.split())

# Submit
r=dsa.query(sql)

# For large queries better use a file in MySpace
# r = dsa.query(sql, saveAs='#ukidss/wsa_gps_res.vot')

# Wait until query status is completed
while r.status()<>'COMPLETED':
        sleep(10)

# Save results to file
open('wsa_gps_res.vot','w').write(r.results()[0])

# If the file is saved in MySpace then do
# open('wsa_gps_res.vot','w').write(urllib2.urlopen(r.results()[0]).read())
```

AstroGrid Python script

Next Step : Bring the code to the data

sociology of astronomy

increasing standardisation

- common user instruments (AAT...)
- standardised data formats (FITs ...)
- standardised reduction packages (Starlink...)
- collectivised data collection (SDSS...)
- common access methods and s/w (VO..)
- standardised analysis tools (VO++..)

- does this make us *the Borg*
  or *happy shoppers* ?

facilities vs experiments

- Old : Facility ==> many small users
- New : Experiment ==> one team
  particle physics style
- Or : Data services ==> many small users
  need a data infrastructure