Drowning in Data

VO to the rescue

- Data Deluge
- Tower of Babel
- We are the Borg
- Web 2.0 Astronomy
- Demo

including UKIDSS interlude

ROYAL OBSERVATORY EDINBURGH

Data Deluge

AstroGrid

big science

- key science goals need large datasets
  - 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

scary data ?

ROYAL OBSERVATORY EDINBURGH

- map the sky : 0.1" x 16 bits = 100 TB
- process to find objects : billion row tables
- repeat as required versus wavelength and epoch...

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

AstroGrid

bottlenecks

ROYAL OBSERVATORY EDINBURGH

- 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 provide hardware and tools

service economy

Astro Grid

VO geometry

- not a warehouse
- not a hierarchy
- not a peer-to-peer system

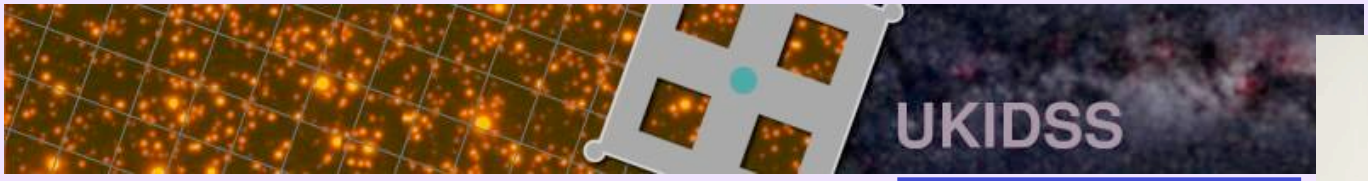
- small set of *service centres*
  and large population of *end users*

pretty much like *shopping*...

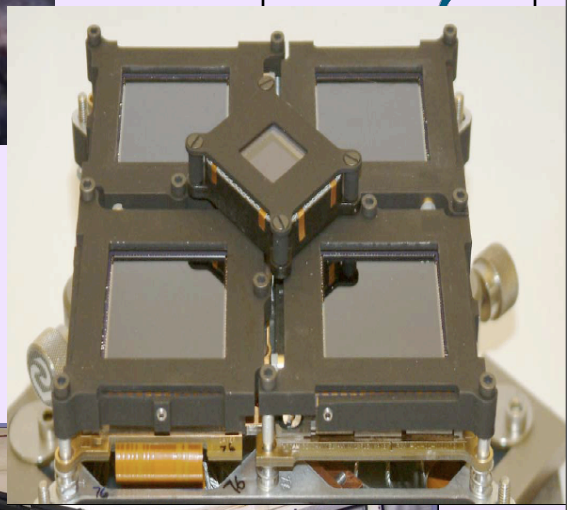
ROYAL OBSERVATORY EDINBURGH

UKIDSS

Astro Grid

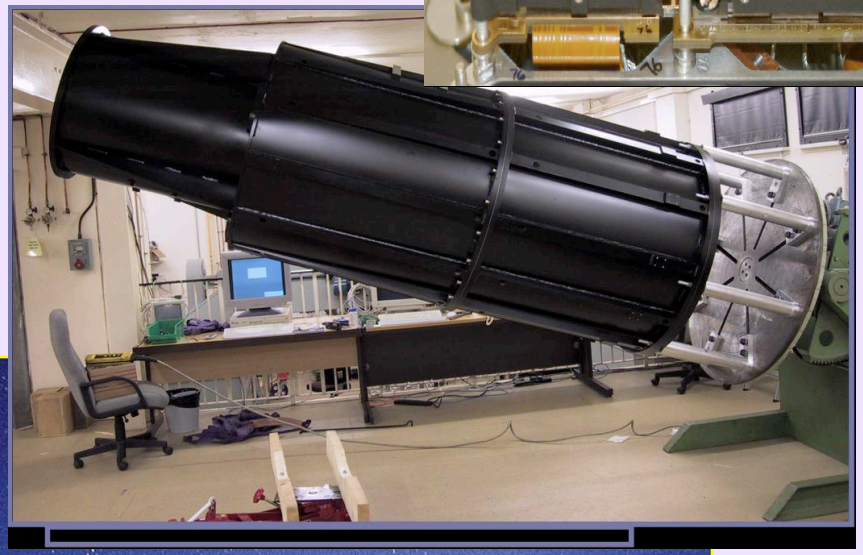


Lawrence *et al* 2007



- 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

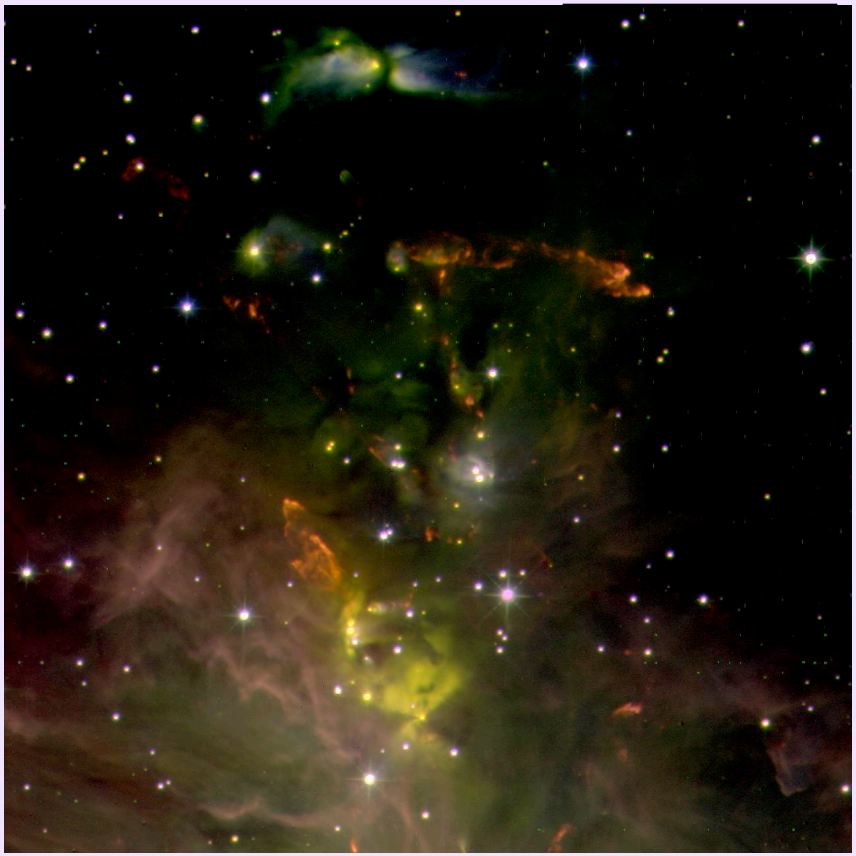
http://surveys.roe.ac.uk/wsa



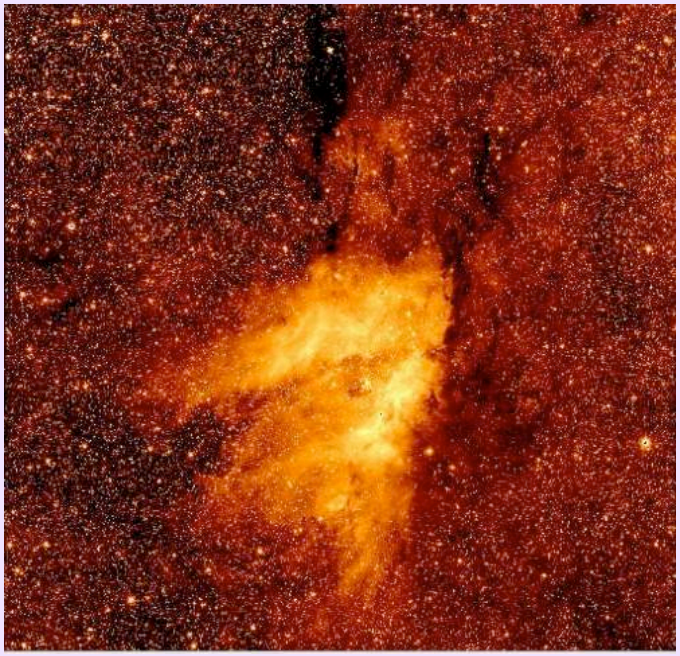
Astro Grid

WFCAM pix

NGC 891



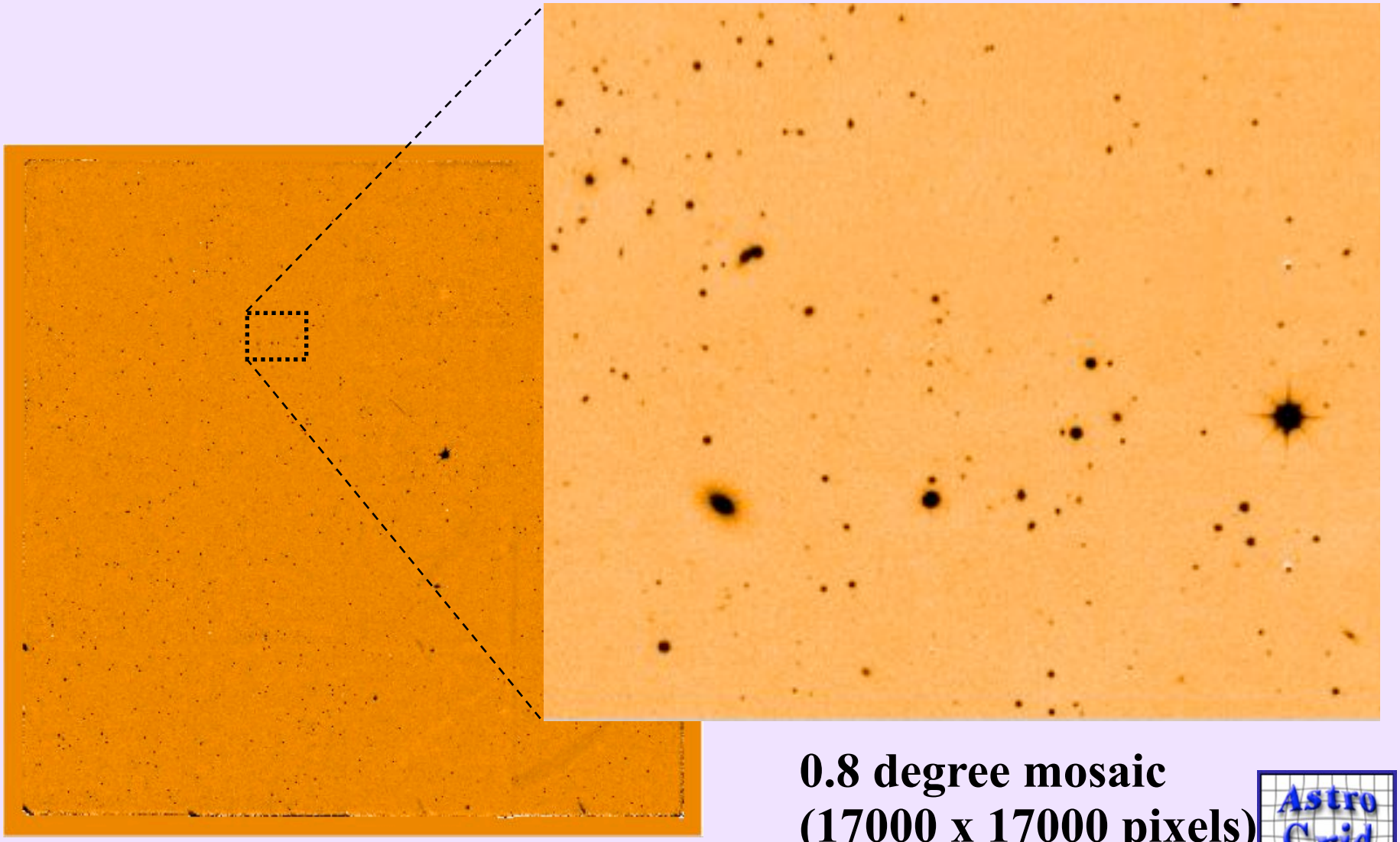
ORION



M17

M104

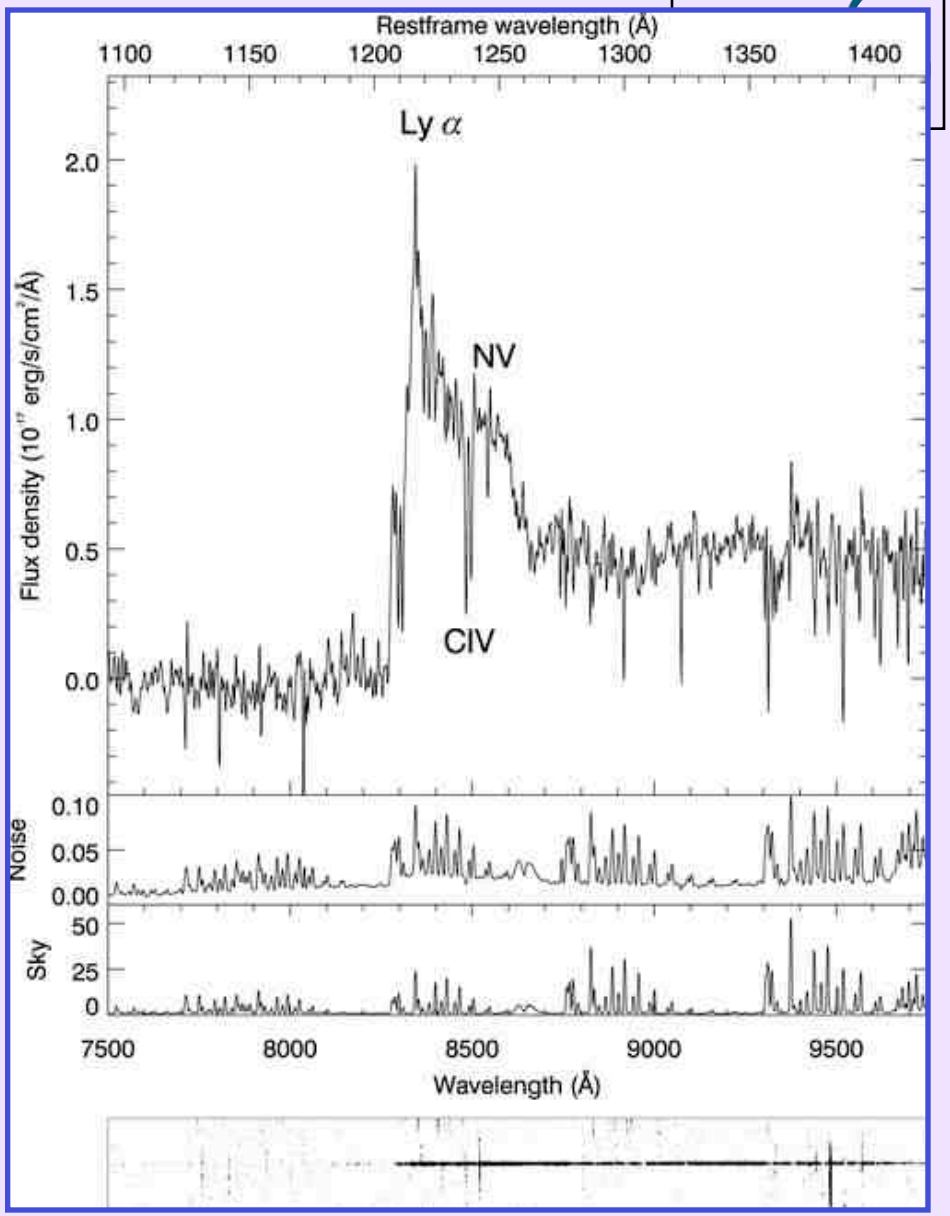
scary amounts of data



0.8 degree mosaic
(17000 x 17000 pixels)

z=6 quasar

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

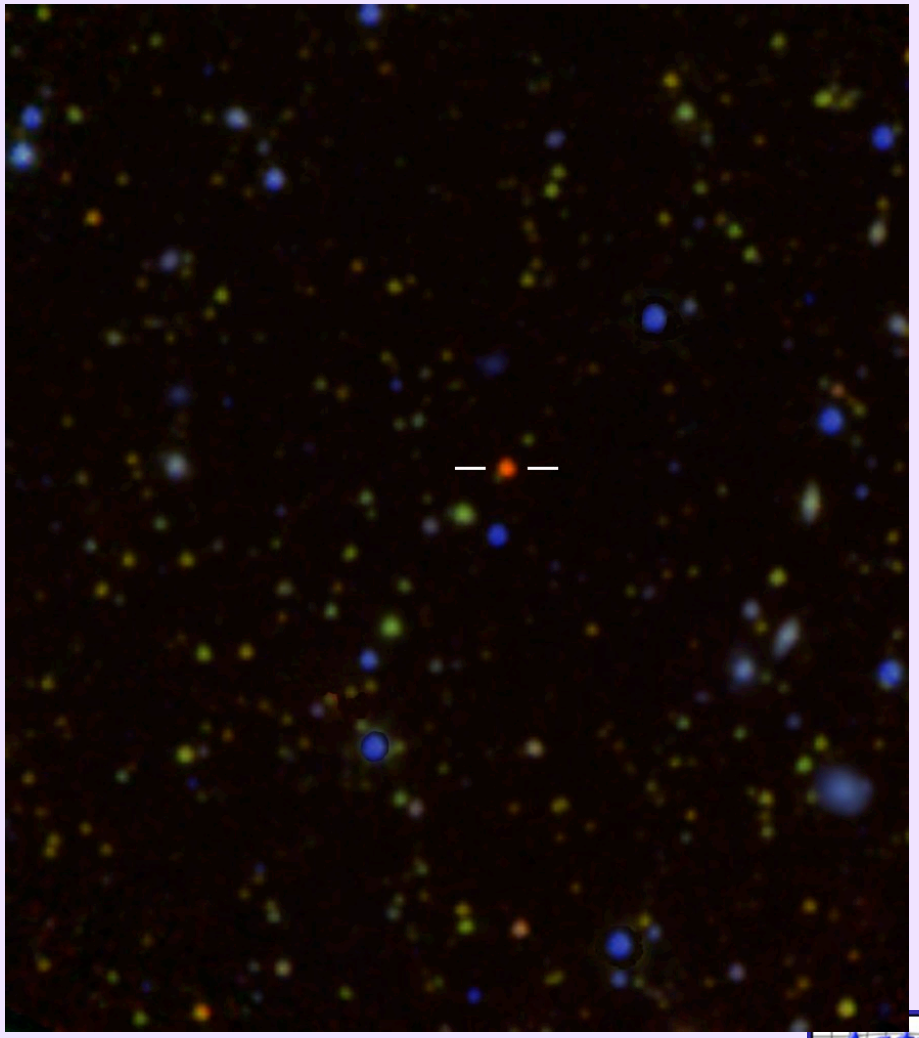


Venemans et al 2007

Astro Grid

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

The Tower of Babel

Diversity

- problem is not just number of *bits* ...
- ....but the number of *archives*

data rich future

- heritage
  - Schmidt, IRAS, Hipparcos
- current hits
  - VLT, SDSS, 2MASS, HST, Chandra, XMM, WMAP, UKIDSS
- coming up :
  - VISTA, ALMA, JWST, Planck, Herschel
- cross fingers :
  - LSST, ELT, Lisa, Darwin,SKA, XEUS, etc.
- plus lots more

- issue is archive *interoperability*
  - need **standards** and **transparent infrastructure**

standards
International VO Alliance

ROYAL OBSERVATORY EDINBURGH

- formal process based on W3C
- key standards agreed
  - formats
  - service metadata
  - data access protocols
  - column semantics
  - s/w interfaces
  - storage addressing

What is the VO ?

- agreed *standards*
- inter-operable *data collections*
- inter-operable *software modules*

- no central VO-command

- its not a thing
- its a way of life

The VO vision

- web all docs in the world inside your PC
- VO all databases in the world inside your PC

ROYAL OBSERVATORY EDINBURGH

We are the Borg

AstroGrid

increasing collectivisation

- 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

Web 2.0 Astronomy

The wisdom of the crowd ?

- Is this is all too rigid ?
  - life dominated by big missions and data centres
  - the IVOA dictates and you must obey
- Why can't the VO just *emerge* ?
  - all the new postdocs are smarter than the greybeards
- Is there a Web 2.0 style VO ?

Web vs Web 2.0

- Web : world becomes transparent
  - but clear divide between *creators* and *readers*
  - and between *servers* and *clients*
- Web 2.0 : users create, adjust, vote
  - blogs, tagging, wiki, Digging etc
  - note these all rely on a background infrastructure...
- What is the astro equivalent ?
  - annotate resources
  - write your own tools and share them

ROYAL OBSERVATORY EDINBURGH

Demo

Astro Grid

ROYAL OBSERVATORY EDINBURGH

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.

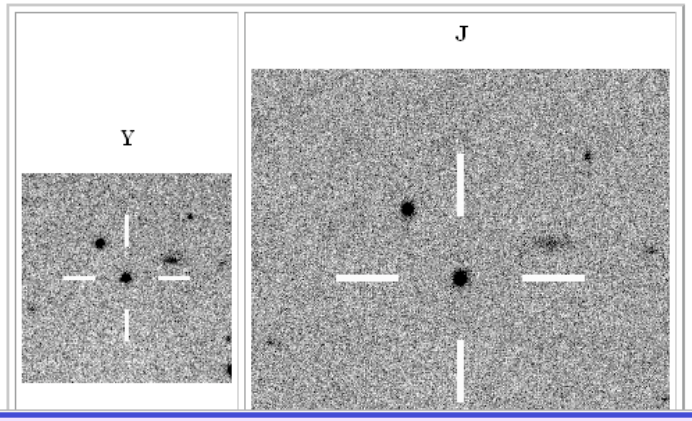
frameSetID	ra	dec	sigRa	sigDec	epoch
33791702333	+354.9942751	-0.0567320	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0068879	-0.0561711	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+355.0516556	-0.0550098	-9.999995E+008	-9.999995E+008	-9.9999950E+008
33791702333	+354.9430954	-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

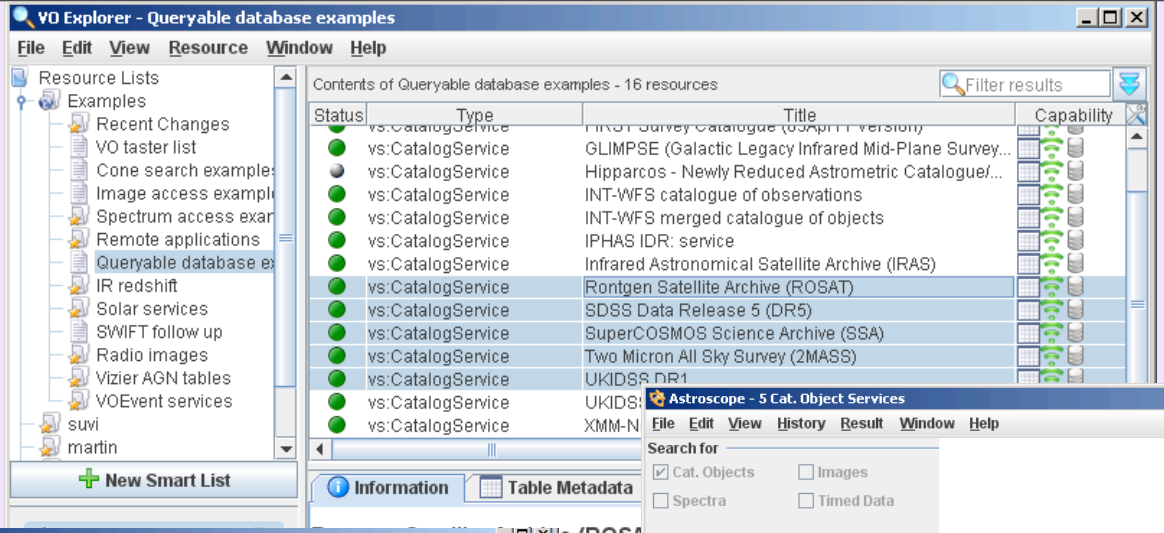
Link	multiframeID	frametype	obstype	filterid	shortname	dateObs	extNum
show	976960	stack	OBJECT	2	Y	2006-06-10 09:06:14.0	5
show	988076	leavstack	OBJECT	3	J	2006-06-10 09:30:23.7	5
show	987086	stack	OBJECT	4	H	2006-06-10 09:54:50.0	5
show	983032	stack	OBJECT	5	K	2006-06-10 10:19:17.1	5

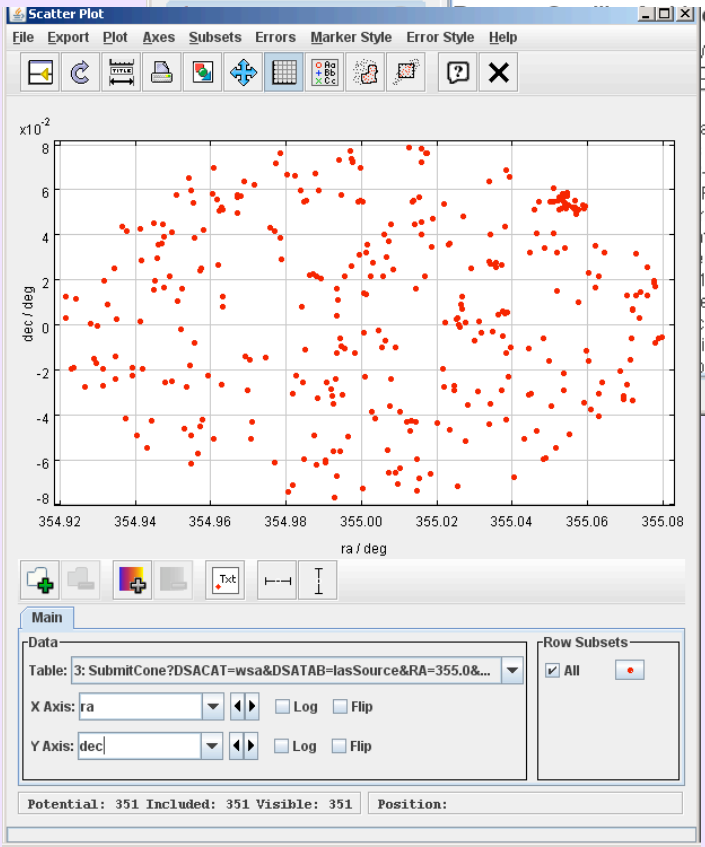
4 rows returned.

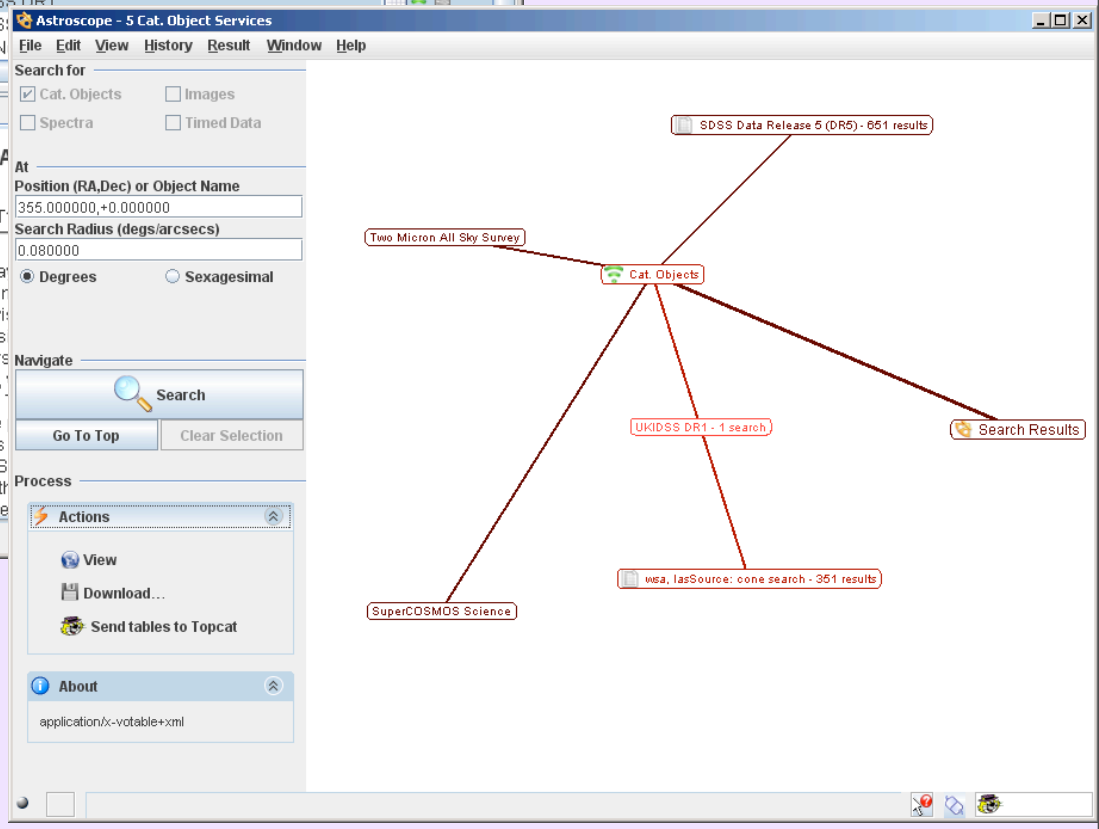


Astro Grid


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
ROYAL OBSERVATORY EDINBURGH
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.

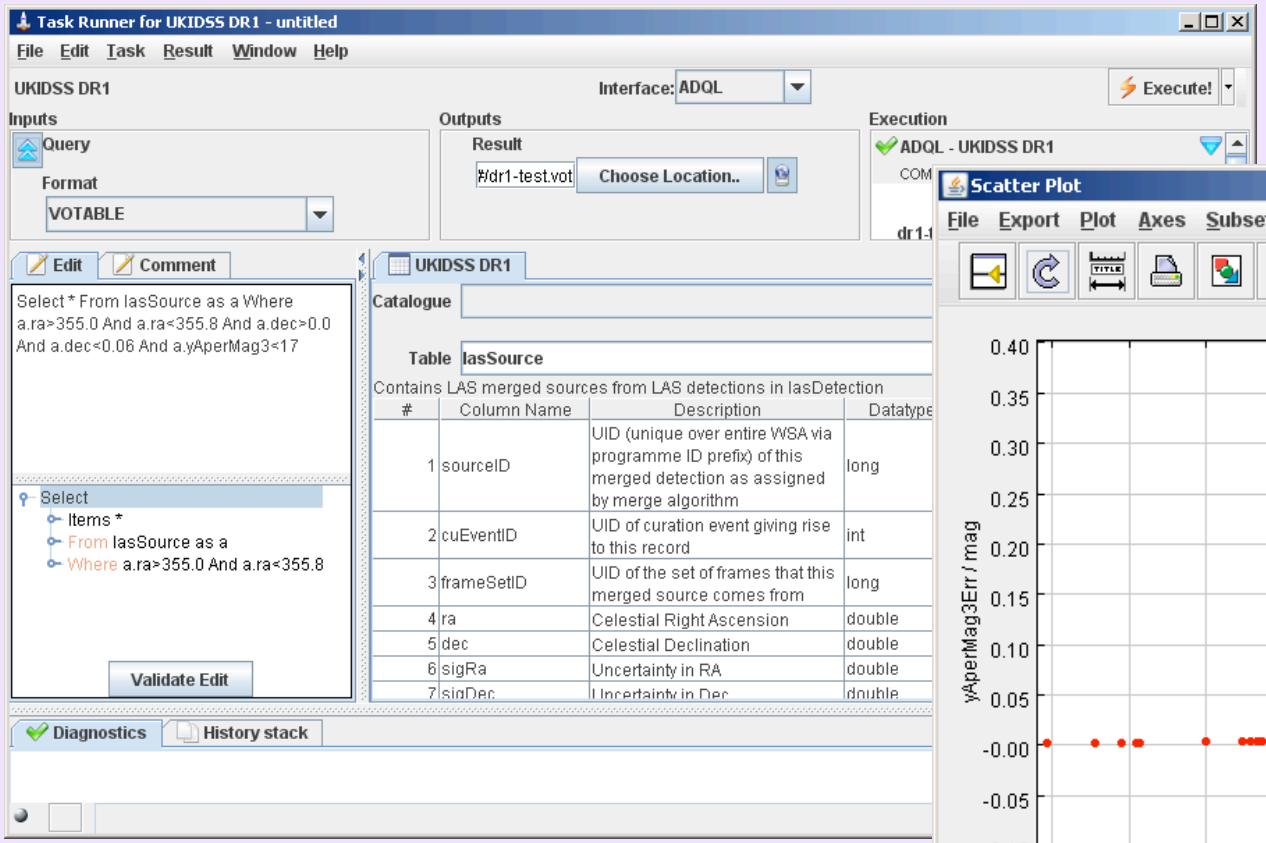


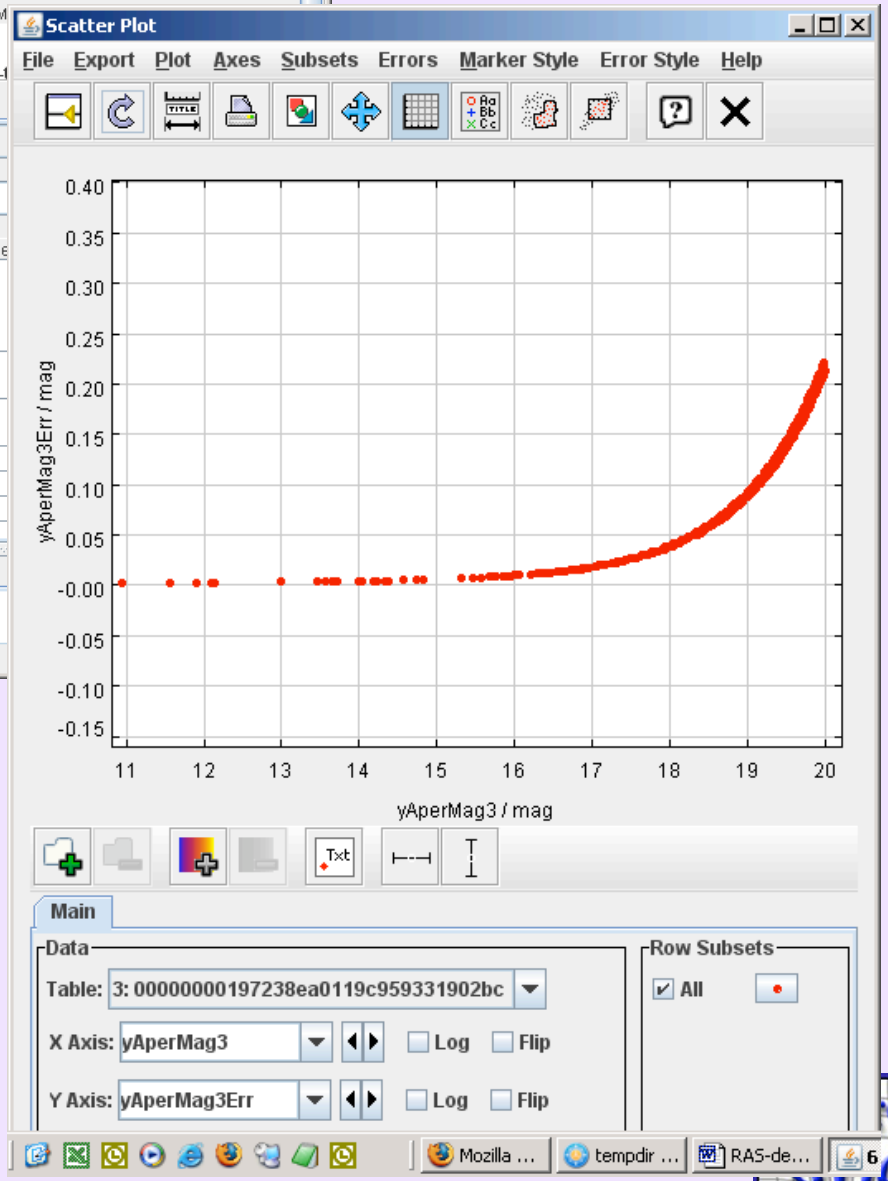




www.astrogrid.org

query multiple databases simultaneously





generic query builder

```
#!/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