

# Lecture 5b: Data Mining

Peter Wheatley

# Data archives

- Most astronomical data now available via archives
  - Raw data and high-level products usually available
  - Data reduction software often specific to individual missions/instruments
- MAST – Multimission Archive at STScI
  - HST and ultraviolet missions
  - <http://archive.stsci.edu/index.html>
- HEASARC - High Energy Astrophysics Science Archive Research Center
  - X-ray and gamma-ray missions
  - <http://heasarc.gsfc.nasa.gov/>
- ESO Science Archive Facility
  - All ESO instruments
  - [http://archive.eso.org/eso/eso\\_archive\\_main.html](http://archive.eso.org/eso/eso_archive_main.html)
- And many others...



- [FAQ](#)
- [High-Level Science Products](#) ▶
- [Software](#) ▶
- [FITS](#) ▶
- [Archive Manual](#)
- [Related Sites](#)
- [NASA Datacenters](#) ▶
- [MAST Services](#)
- [MAST and the VO](#) ▶
- [Newsletters & Reports](#) ▶
- [Data Use Policy](#)
- [Dataset Identifiers](#)
- [Acknowledgments](#)



The Multimission Archive at STScI is a NASA funded project to support and provide to the astronomical community a variety of astronomical data archives, with the primary focus on scientifically related data sets in the optical, ultraviolet, and near-infrared parts of the spectrum.

**Search MAST for a Target or Mission**

Enter [Target name \(or Coordinates\)](#):

Resolver:  SIMBAD  NED  Don't Resolve

and/or [Band/Data Type\(s\)](#): [more options](#)

	Extreme UV	Far UV	Near UV	Optical	Near IR	Radio
Images	<input type="checkbox"/>					
Spectra	<input type="checkbox"/>					
Other	<input type="checkbox"/>					

[Help](#)



WWW  MAST

Note: STScI may provide links to Web pages that are not part of the STScI, AURA, NASA, or ESA domain. These sites are managed by organizations, companies or individuals not under our control, and neither STScI, AURA, NASA, nor ESA are responsible for the information or links you may find there. We provide these links as a convenience and the presence of these links is not an endorsement of the site.

## NEWS

### December 09, 2011:

WFC3 Persistence Project Releases Correction Files

### December 08, 2011:

New "Condition Flag" available in Kepler Target Search Forms

### December 07, 2011:

Kepler announces Accelerated Data Release schedule

### December 03, 2011:

HST/NICMOS Paschen-alpha Survey of the Galactic Center High Level Science Products now available

### November 28, 2011:

Enhanced Kepler Target Search Interface now available



## Missions

- [Hubble](#)
- [Hubble Legacy Archive](#)
- [HSTonline](#)
- [DSS](#)
- [GALEX](#)
- [JWST](#)
- [KEPLER](#)
- [XMM-OM](#)
- [BEFS \(ORFEUS\)](#)
- [Copernicus](#)



# HEASARC Browse

Other Browse interfaces:

[Notification Service](#) | [Batch](#) | [Correlation](#) | [Index of all tables](#) | [Keyword Search](#)

Query File And Session Upk

**Main Search Form** > Search Results > Choose Data Products

Start Search

Reset

Detailed Mission/Catalog Search

## 1. Do you want to search around a position ... ?

(If you want to search on parameters other than object name or coordinates, select "Detailed Mission/Catalog Search".)

**Object Name Or Coordinates:**

and/or

Select  
Local  
File:

Choose File no file selected

e.g. Cyg X-1 or 12 00 00, 4 12 6 or  
Cyg X-2; 12.235, 15.345 (Note use of semi-colons  
(;) to separate multiple object names or coordinate  
pairs)

File should contain objects and/or coordinate pairs one  
per line or separated by semi-colons.

**Coordinate System:** J2000

**Search Radius:** Default

arcmin

Default uses the optimum radius for each catalog searched.

... and/or search by date?

**Observation Dates:**

YYYY-MM-DD hh:mm:ss or MJD: DDDDD.ddd

The time portion of the date is optional. Separate multiple dates/ranges with semicolons (;).  
Range operator is '..'. (e.g. 1992-12-31; 48980.5; 1995-01-15 12:00:00; 1997-03-20 .. 2000-10-18)

## 2. What missions and catalogs do you want to search? (Bold text indicates mission is active)

### Most Requested Missions

Chandra **[CXC,CSC]**

Fermi

ROSAT

RXTE

Suzaku

Swift

WMAP

XMM-Newton **[XSA]**

### Other X-Ray and EUV Missions

Ariel V

ASCA

BeppoSAX

BBXRT/Astro-1

Copernicus

Einstein

FIVE

EXOSAT



# ESO Archive Query Form

[ESO Archive Overview](#) [Help Page](#) [FAQ](#)

[Archive Facility HOME](#) [ESO HOME](#)

If you would like to query the Archive for instrument specific parameters, please use the [dedicated query forms](#).  
To search for **reduced Data Products**, please have a look at the [ESO Data Products](#) page and the [Advanced Data Products](#) query form.

The checkboxes on the right of the parameters define whether or not they will be displayed on the query result page.

[query Help](#) [Status of Requests](#)

## Target, Program and Scheduling Information

**Target Name**   Resolved by SIMBAD

**RA**  **DEC**  **J2000**

**Search Box**  **Input** RA(h) DEC(deg)

**Output**  Sexagesimal (h, deg)

[List of Targets](#)  no file selected

**Night**   (DD MM YYYY)

OR give a query range using the following start/end dates:

**Start**  12 hrs [UT]  **End**  12 hrs [UT]

**Program ID**   **Program Type**  Any

**PI CoI**   **SV**  Any

**Title**

## Observing Information

Imaging	Spectroscopy	Interferometry	Other
<input type="button" value="ALL"/> <input type="button" value="NONE"/>			
<input type="checkbox"/> VLT/FORS1	<input type="checkbox"/> VLT/CRIRES	<input type="button" value="NONE"/>	<input type="checkbox"/> APEX/HET
<input type="checkbox"/> VLT/FORS2	<input type="checkbox"/> VLT/FORS1	<input type="checkbox"/> VLT/VINCI	<input type="checkbox"/> APEX/BOL
<input type="checkbox"/> VLT/HAWKI	<input type="checkbox"/> VLT/FORS2	<input type="checkbox"/> VLT/MIDI	<input type="checkbox"/>
<input type="checkbox"/> VLT/ISAAC	<input type="checkbox"/> VLT/GIRAFFE	<input type="checkbox"/>	UKIRT/WFCAM
<input type="checkbox"/> VLT/NACO	<input type="checkbox"/> VLT/ISAAC	VLT/AMBER	<input type="checkbox"/> LGSF
<input type="checkbox"/> VLT/VIMOS	<input type="checkbox"/> VLT/NACO	<b>Polarimetry</b>	<input type="checkbox"/> MASCOT
<input type="checkbox"/> VLT/VISIR	<input type="checkbox"/> VLT/SINFONI	<input type="button" value="ALL"/> <input type="button" value="NONE"/>	
<input type="checkbox"/> VISTA/VIRCAM	<input type="checkbox"/> VLT/UVES	<input type="button" value="NONE"/>	
<input type="checkbox"/>	<input type="checkbox"/> VLT/VIMOS	<input type="checkbox"/> VLT/FORS1	
VST/OMEGACAM	<input type="checkbox"/> VLT/VISIR	<input type="checkbox"/> VLT/FORS2	

**Category**

SCIENCE  
 CALIB

**Data Product Info**

**Type**  Any

User defined input:

**Mode**  Any

User defined input:

**Dataset ID**

**Orig Name**

**Release Date**

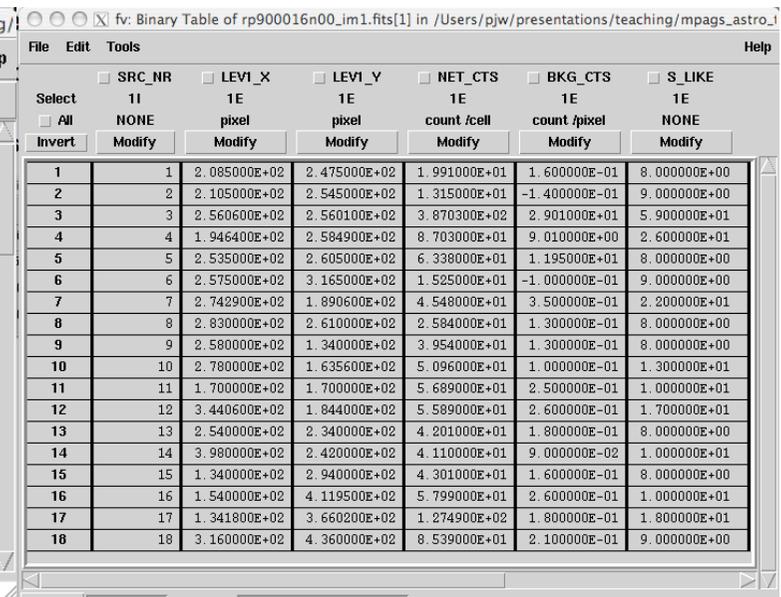
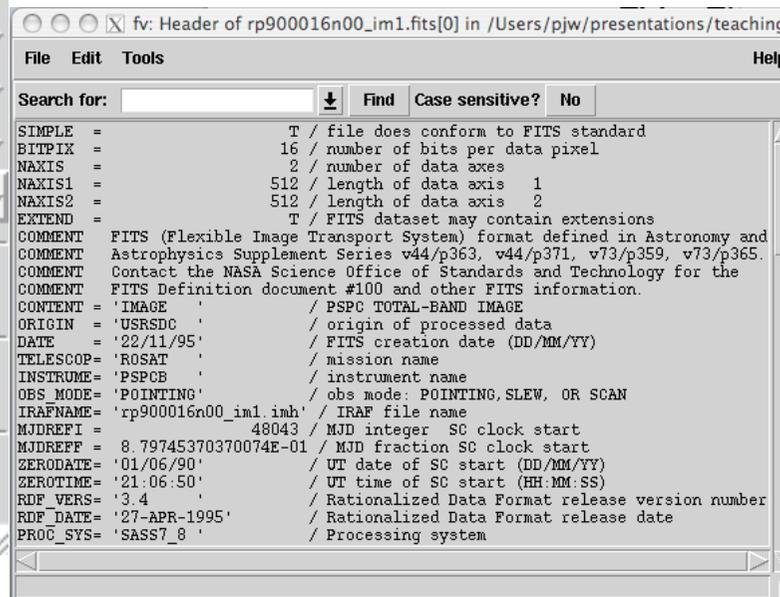
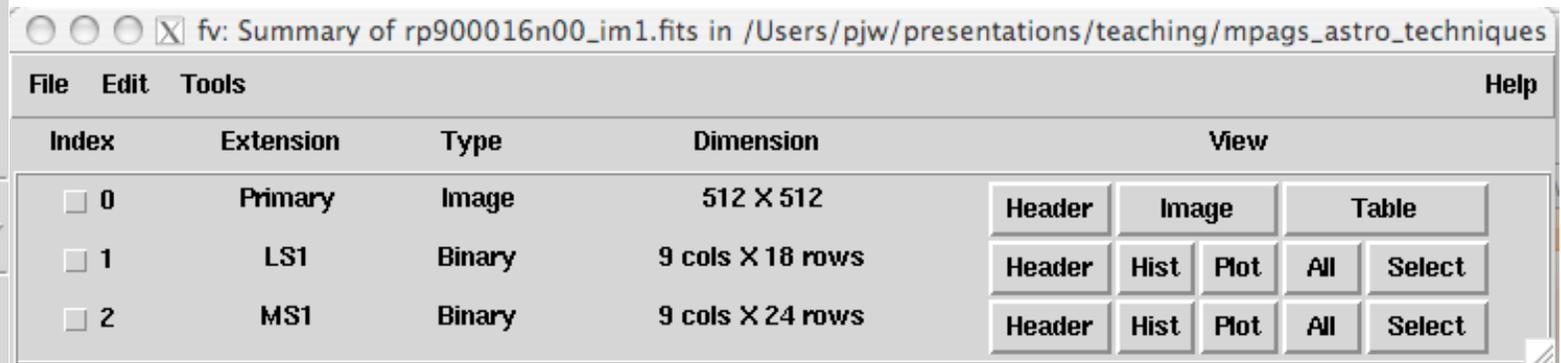
**OB Name**

# Data file formats

- ASCII – text files
  - Commonly used for small / simple datasets
- FITS - Flexible Image Transport System
  - The most common format for astronomical data
  - Ascii headers and binary data storage
  - Designed for images, but extended for 2d (and even 3d) tables
- HDS - Hierarchical Data System
  - UK Starlink standard, still used by many starlink tools
  - Highly versatile, with arbitrarily complex data structures
  - No longer widely used
- VOTable – virtual observatory table
  - XML standard for astronomical data tables
  - Designed for easy exchange of data for Virtual Observatory tools
  - Software can *understand* rather than just *read* (e.g. defn of exposure time)

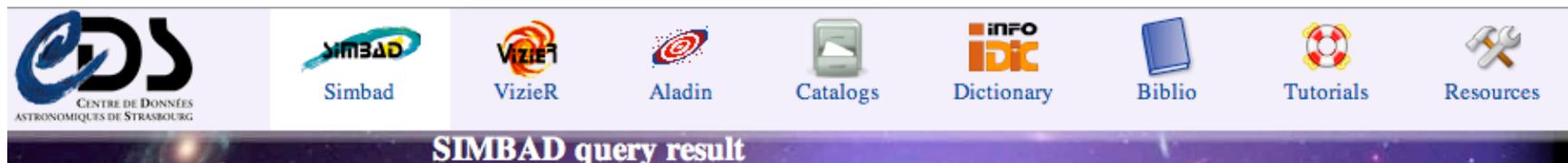
# Accessing FITS files

- cfitsio - set of sub-routines for reading/writing fits files
  - Often built into scripting languages, e.g. python
- FV- Useful tool for viewing contents of FITS files
  - Part of the HEASOFT software, or available stand-alone



# Astronomical catalogues

- CDS Simbad – data and bibliography for objects mentioned in papers



[other query modes :](#)
[Identifier query](#)
[Coordinate query](#)
[Criteria query](#)
[Reference query](#)
[Basic query](#)
[Script submission](#)
[Output options](#)
[Help](#)

Object query : vega

C.D.S. - SIMBAD4 rel 1.186 -  
2011.12.16CET08:14:51

Available data : [Basic data](#) • [Identifiers](#) • [Plot & images](#) • [Bibliography](#) • [Measurements](#) • [External archives](#) • [Notes](#) • [Annotations](#)

## Basic data :

**NAME VEGA -- Variable Star**

query around with radius  arcmin

Other object types:

\*  
(\* , AG , ASCC , BD , CSI , FK5 , GAT , GC , GCRV , GEN# , GJ , HD , HGAM , HIC , HIP , HR , JP11 , N30 , 8pc , PLX , PMC , F , \*\* ( ADS , CDM , IDS , WDS ) , IR ( IRAS , IRC , 2MASS , RAFGL ) , PM\* ( LSPM , LTT , NLTT ) , V\* ( V\* , CSV ,

ICRS coord. (ep=J2000) : 18 36 56.33635 +38 47 01.2802 ( Optical ) [ 3.51 2.81 90 ] A [2007A&A...474..653V](#)

FK5 coord. (ep=J2000 eq=2000) : 18 36 56.336 +38 47 01.28 ( Optical ) [ 3.51 2.81 90 ] A [2007A&A...474..653V](#)

FK4 coord. (ep=B1950 eq=1950) : 18 35 14.67 +38 44 09.8 ( Optical ) [ 20.31 16.25 90 ] A [2007A&A...474..653V](#)

Gal coord. (ep=J2000) : 067.4482 +19.2373 ( Optical ) [ 3.51 2.81 90 ] A [2007A&A...474..653V](#)

Proper motions *mas/yr* [error ellipse]: 200.94 286.23 [0.32 0.40 0] A [2007A&A...474..653V](#)

Radial velocity / Redshift / cz : V(km/s) -13.9 [0.9] / z(-) -0.000046 [0.000003] / cz -13.90 [0.90] (~) A [1979IAUS](#)

Parallax *mas*: 130.23 [0.36] A [2007A&A...474..653V](#)

Spectral type: A0V C ~

Fluxes (7) : B 0.03 [-] C ~

V 0.03 [-] C ~

# Astronomical catalogues

- CDS Vizier –inc. sky survey catalogues and tables from papers

**CDS** CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

**SIMBAD** Simbad

**VIZIER** VizierR

**Aladin**

**Catalogs**

**INFO IDIC** Dictionary

**Biblio**

**Tutorials**

**Resources**

**VizieR Service**

**new CDS cross-match service : Service to cross-identify sources in tables.**

**Search Criteria**

**Keywords**

hd189733

**Preferences**

max: 50

HTML Table

All columns

► Compute

**Mirrors**

CDS, France

**Find catalogs among 9530 available**

Clear  Find...

Expand search

? *Catalog, author's name, word(s) from title, description, etc. e.g.: AGN, Veron, I/239, or bibcodes...*

► **Search catalogs by column descriptions (UCD)** ?

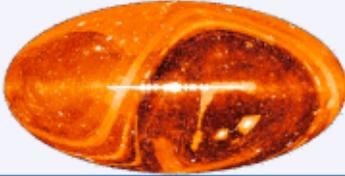
Wavelength	Mission	Astronomy
Radio	AKARI	AGN
IR	ANS	Abundances
optical	ASCA	Ages
UV	BeppoSAX	Associations
EUV	CGRO	Atomic_Data
X-ray	COBE	BL_Lac_objects
Gamma-ray	Chandra	Binaries:cataclysmic

**Search by Position across 9899 tables**

Target Name (resolved by [Sesame](#)) or Position: Clear

Target dimension: J2000  2 arcmin  Go!

Radius  Box size



**~ 0 matching catalogs** Find Catalogs

[More about Vizier](#)

Browsing modes: [Designation](#), [Acronyms](#), [Favorites](#), [Dates](#), [Image\\_spectra](#), [Kohonen](#)

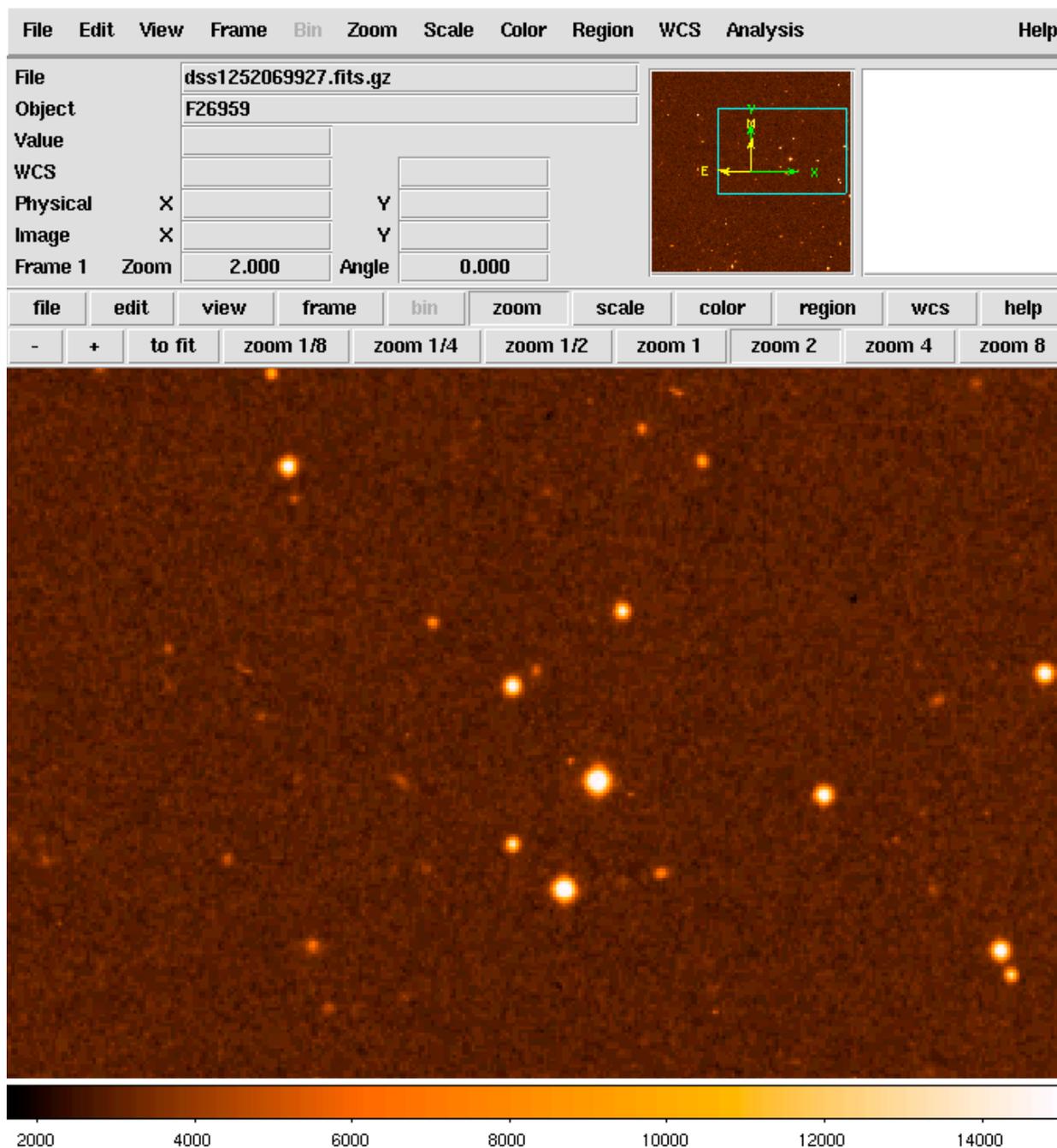
Or list [the large surveys](#)

Return to the old Vizier version [here](#)

©UDS/CNRS Contact: 

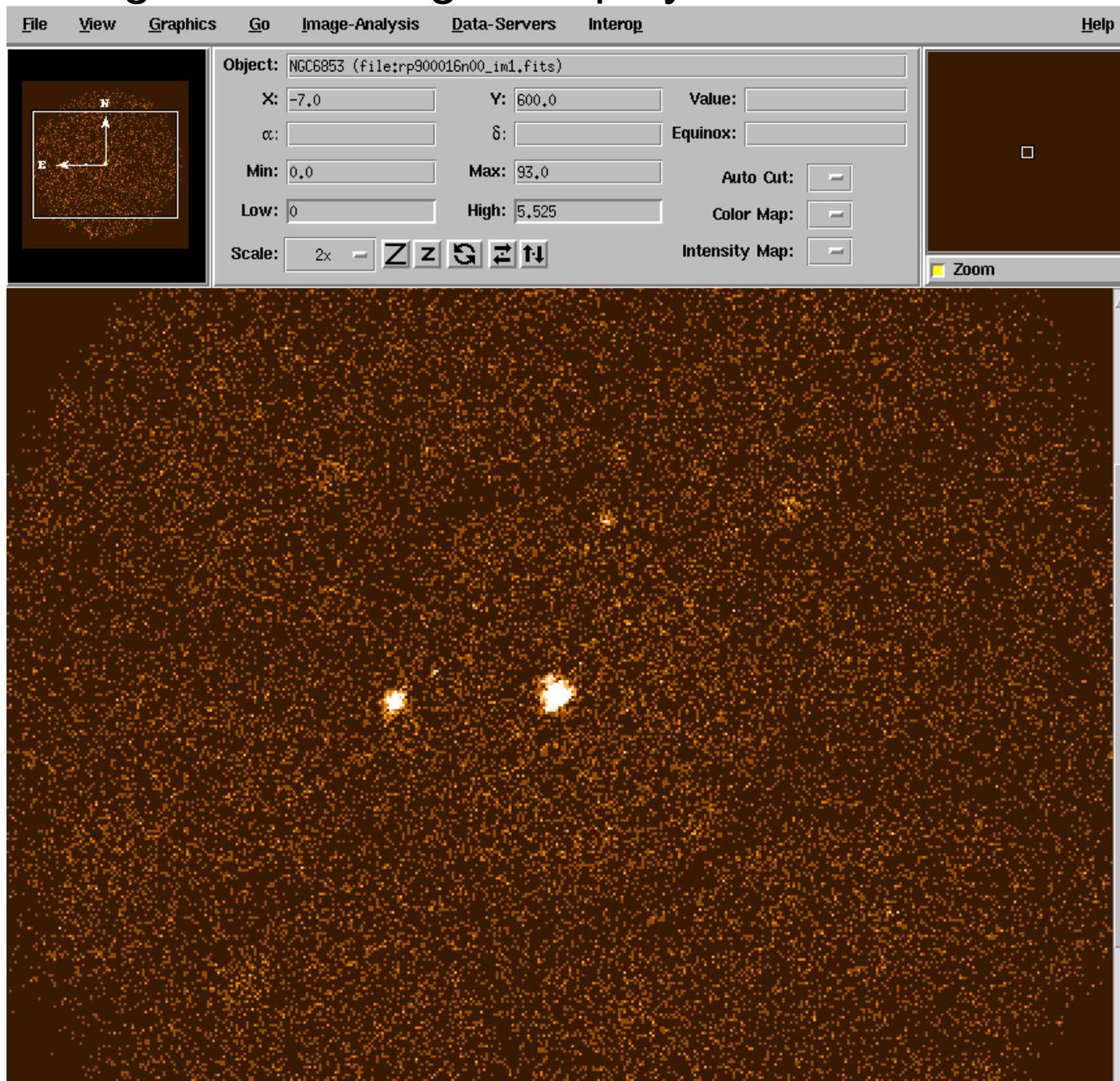
# Image and catalogue display

- ds9
  - Image display and comparison
  - Image and catalogue download
  - Catalogue overlay



# Image and catalogue display

- Gaia
  - Starlink software
  - Includes photometry and astrometry software



## Lecture 5 Assignment, Part 1 of 2

- Use SIMBAD to determine the nature of the star HD189733 and any companion objects
- Download from HEASARC a ROSAT PSPC image of the star HD189733
- Display the ROSAT image aligned with an optical image of the same field (downloaded from within ds9 or Gaia)
- Overlay SIMBAD sources (also downloaded within ds9/Gaia)
- Send us a screenshot
- Determine the name and nature of the famous nebula nearby

# TOPCAT & STIL(TS)

TOPCAT Java GUI

<http://www.star.bris.ac.uk/~mbt/topcat/>

STIL library for dealing with tabular data

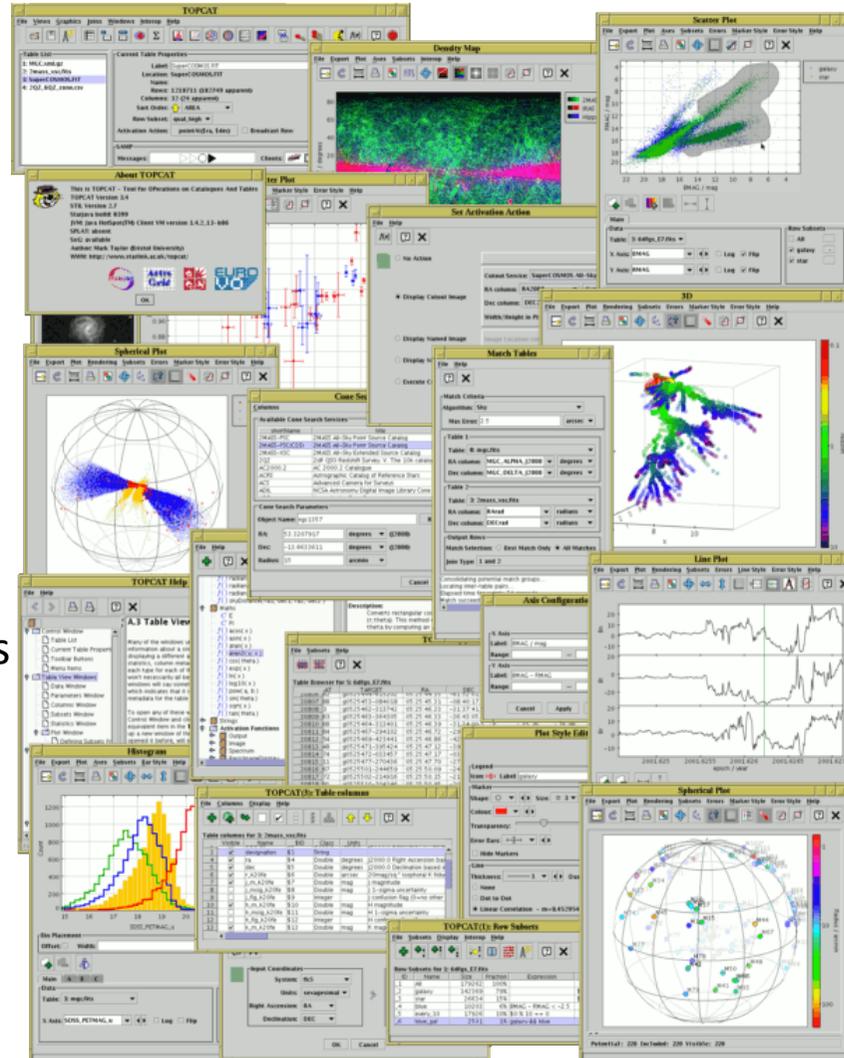
<http://www.star.bristol.ac.uk/~mbt/stil>

STILTS command-line tool

<http://www.star.bristol.ac.uk/~mbt/stilts/>

Key features:

- Import/export catalogs in various formats
- Cross-matching and merging
- Plotting and manipulating data
- Search and retrieve data from archives



# TOPCAT : VO retrieval

The screenshot displays the TOPCAT Cone Search window. The main window is titled "Cone Search" and contains the following elements:

- Available Cone Search Services:**
  - Registry: [http://registry.astrogrid.org/astrogrid-registry/services/RegistryQueryv1\\_0](http://registry.astrogrid.org/astrogrid-registry/services/RegistryQueryv1_0)
  - Keywords: iphas
  - Match Fields:  Short Name  Title  Subjects  ID  Publisher  Description
  - Accept Resource Lists
  - Buttons: Cancel Query, Submit Query
- Search Results Table:**

Short Name	Title	
J/A+A/480/409	IPHAS symbiotic stars candidates (Corradi+, 2008)	Binarie
J/A+A/504/291	Candidate Planetary Nebulae in IPHAS catalog (Viironen+, 2009)	Photor
J/MNRAS/384/1277	H(alpha) emission line sources from IPHAS (Witham+, 2008)	Photor
J/MNRAS/397/1685	Galactic plane IPHAS-POSS1 proper motion survey (Deacon+, 2009)	Proper
	IPHAS IDR: service	IPHAS
- Cone Search Parameters:**
  - Cone Search URL: <http://vizier.u-strasbg.fr/viz-bin/votable/-A?-source=J/MNRAS/384/1277&>
  - Object Name: KR Aur
  - RA: 93.933 degrees (J2000)
  - Dec: 28.585722 degrees (J2000)
  - Radius: 50 degrees
  - Accept Sky Positions
  - Buttons: Resolve, OK

In the background, a smaller window titled "Table List" is visible, showing a table with 12 rows and 910 M columns.

# Table Data

add custom columns

TOPCAT(1): Table Columns

File Columns Display Help

Table Columns for 1: KR Aur-J\_MNRAS\_384\_1277-50d

Visible	Name	\$ID	Class	Units	Description	UCD
<input type="checkbox"/>	Index	\$0	Long		Table row index	
<input checked="" type="checkbox"/>	_r	\$1	Double	deg	Distance from center (RA)J2000=06 15 43.92, (DEJ)2000=+28 35 08.6)	POS_ANG_DIST_GENER
<input checked="" type="checkbox"/>	_RAJ2000	\$2	Double	deg	Right ascension (FK5) Equinox=J2000. (computed by VizieR, not part of the original data)	POS_EQ_RA_MAIN
<input checked="" type="checkbox"/>	_DEJ2000	\$3	Double	deg	Declination (FK5) Equinox=J2000. (computed by VizieR, not part of the original data)	POS_EQ_DEC_MAIN
<input checked="" type="checkbox"/>	IPHAS	\$4	String		IPHAS source designation, JHHMMSS.ss+DDMMSS.s (1)	ID_MAIN:1
<input checked="" type="checkbox"/>	r'mag	\$5	Float	mag	Sloan r' band magnitude	PHOT_SDSS_R
<input checked="" type="checkbox"/>	e_r'mag	\$6	Float	mag	Sloan r' band magnitude error	ERROR
<input checked="" type="checkbox"/>	r'-i'	\$7	Float	mag	Sloan r'-i' colour index	PHOT_SDSS_R-I
<input checked="" type="checkbox"/>	e_r'-i'	\$8	Float	mag	Sloan r'-i' colour index error	ERROR
<input checked="" type="checkbox"/>	r'-Ha	\$9	Float	mag	Sloan r'-H(alpha) colour index	PHOT_CL_R-HALPHA
<input checked="" type="checkbox"/>	e_r'-Ha	\$10	Float	mag	Sloan r'-H(alpha) colour index error	ERROR
<input checked="" type="checkbox"/>	Flag	\$11	String		[cksmo ] Photometry/Cross-match flag (2)	CODE_MISC
<input checked="" type="checkbox"/>	KW99	\$12	String		Details from Kohoutek \& Wehmeyer (1999, Cat. III/205)	DATA_LINK
<input checked="" type="checkbox"/>	RAJ2000	\$13	String	"h:m:s"	Right ascension (J2000.0)	POS_EQ_RA

TOPCAT(1): Table Browser

File Subsets Help

Table Browser for 1: KR Aur-J\_MNRAS\_384\_1277-50d

	_r	_RAJ2000	_DEJ2000	IPHAS	r'mag	e_r'mag	r'-i'	e_r'-i'	r'-Ha	e_r'-Ha	Flag
1	49.89964	34.39933	56.81556	J021735.84+564856.0	17.912	0.016	0.863	0.021	0.715	0.02	
2	49.98747	34.41121	57.44617	J021738.69+572646.2	18.432	0.025	1.013	0.031	0.913	0.03	
3	49.864	34.52233	57.02453	J021805.36+570128.3	16.709	0.007	0.748	0.01	0.492	0.01	
4	49.87278	34.56196	57.22292	J021814.87+571322.5	17.899	0.016	0.978	0.021	0.796	0.02	
5	49.76696	34.72971	57.11914	J021855.13+570708.9	17.222	0.01	0.816	0.013	0.639	0.013	s
6	49.75811	34.74792	57.12506	J021859.50+570730.2	18.657	0.028	0.996	0.034	0.85	0.034	s
7	49.88995	34.8	58.10786	J021912.00+580628.3	18.013	0.01	0.111	0.031	0.13	0.023	
8	49.65691	34.80183	56.64975	J021912.44+563859.1	16.785	0.007	0.702	0.01	0.518	0.01	
9	49.70676	34.89779	57.30825	J021935.47+571829.7	16.83	0.008	0.109	0.014	0.138	0.012	s
10	49.53083	35.03762	56.66333	J022009.03+563948.0	18.585	0.024	1.012	0.03	1.016	0.028	
11	49.80962	35.27662	59.02044	J022106.39+590113.6	17.029	0.009	1.008	0.011	0.946	0.011	
12	49.92993	35.5885	60.37658	J022221.24+602235.7	17.896	0.016	1.032	0.019	0.634	0.021	
13	49.17729	35.78283	56.96867	J022307.88+565807.2	18.349	0.028	0.915	0.028	0.926	0.034	
14	49.96145	35.89017	61.16406	J022333.64+610950.6	15.933	0.005	0.905	0.006	0.621	0.006	
15	49.75364	35.90437	60.26744	J022337.05+601602.8	14.	0.002	0.585	0.003	0.489	0.003	
16	49.2297	35.92779	57.70722	J022342.67+574226.0	18.611	0.054	1.02	0.062	1.14	0.061	
17	49.54987	35.93512	59.37764	J022344.43+592239.5	13.967	0.002	0.575	0.003	0.415	0.003	
18	49.7565	36.09854	60.69911	J022423.65+604156.8	16.253	0.006	0.828	0.008	0.519	0.008	
19	49.99653	36.35683	62.19936	J022525.64+621157.7	18.692	0.031	1.359	0.035	1.062	0.04	

# TOPCAT Example

3 tables loaded

The screenshot displays the TOPCAT software interface with several windows open. The main window shows a 'Table List' with three tables: '1: KR Aur-J\_MNRAS\_384\_127', '2: KR Aur-IX\_10A-50d', and '4: match(1,2)'. The 'Current Table Properties' window shows details for the selected table. The 'Match Tables' window is open, showing match criteria and a list of matched pairs. The 'Scatter Plot' window shows a plot of  $r-H\alpha$  / mag versus  $r-I$  / mag. The 'Table Browser' window shows a table with columns for RA, Dec, IPHAS, r'mag, e\_r'mag, r'-I, e\_r'-I, r'-Ha, e\_r'-Ha, and Flag.

Table 1:

Table	RA column	Dec column
1: KR Aur-J_MNRAS_384_1277-50d	_RAJ2000	_DEJ2000

Table 2:

Table	RA column	Dec column
2: KR Aur-IX_10A-50d	_RAJ2000	_DEJ2000

Table Browser for 4: match(1,2)

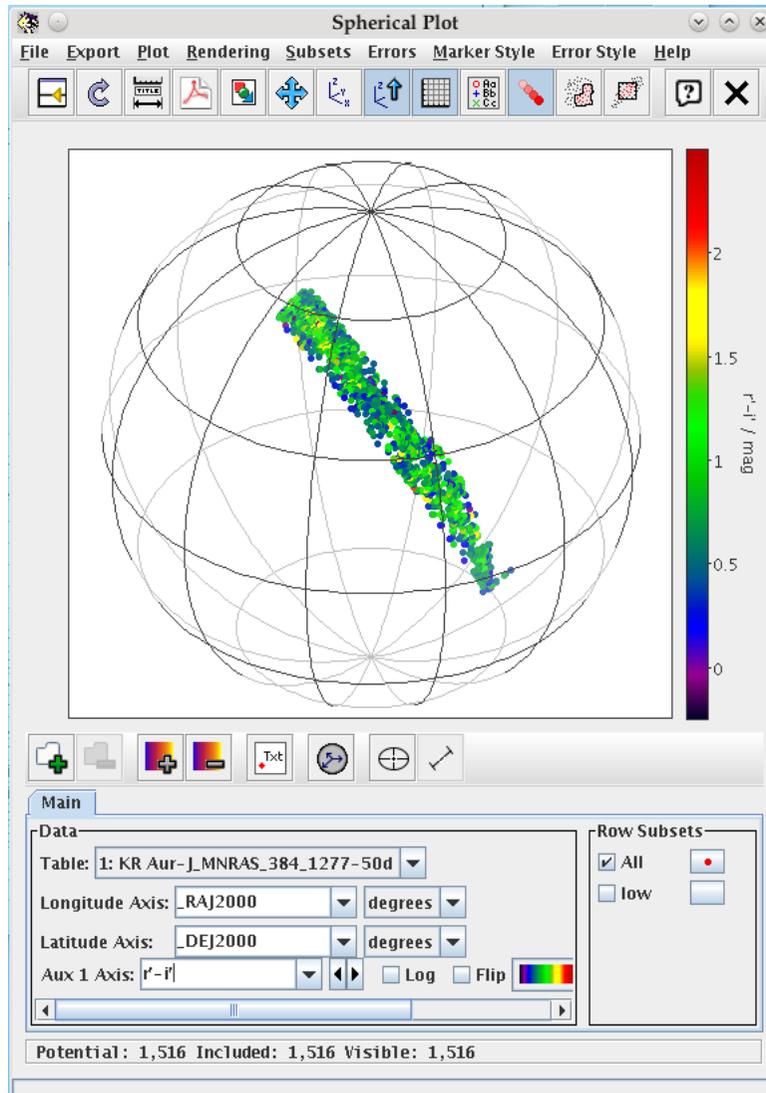
	_r_1	_RAJ2000_1	_DEJ2000_1	IPHAS	r'mag	e_r'mag	r'-I	e_r'-I	r'-Ha	e_r'-Ha	Flag
1	6.12958	86.9515	28.58647	J054748.36+283511.3	15.797	0.005	0.447	0.007	0.669	0.006	s
2	13.18456	94.18008	15.40314	J061643.22+152411.3	15.627	0.002	0.528	0.004	0.887	0.004	s

scatter plot

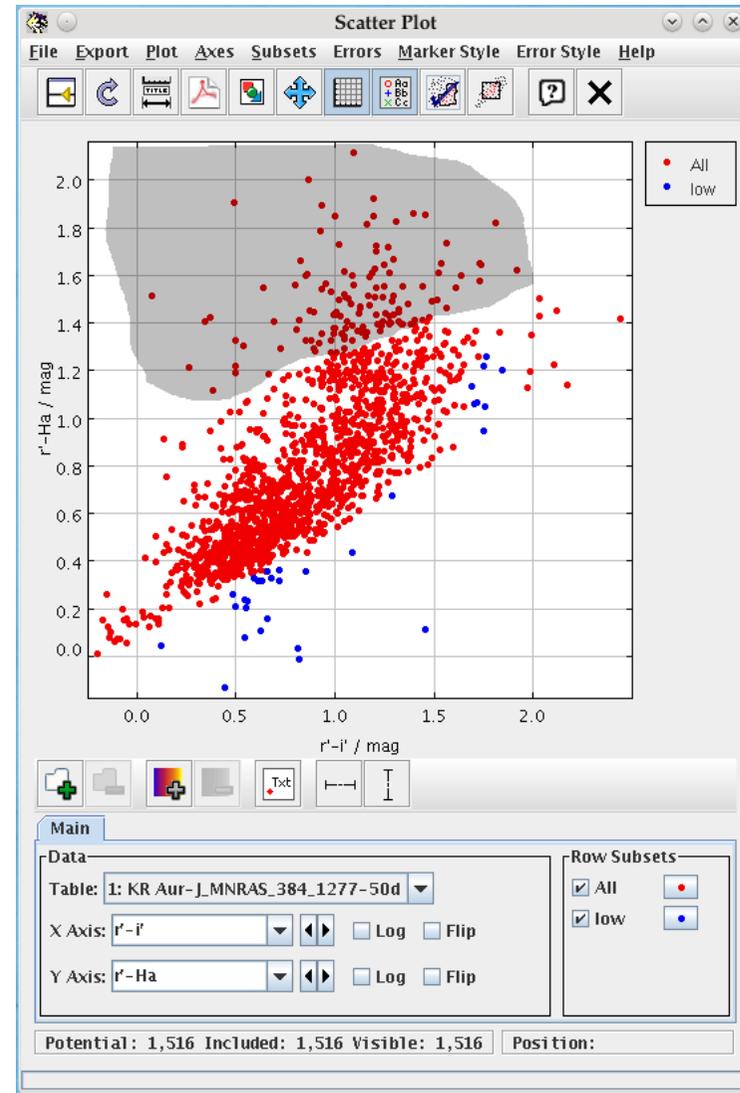
sky matched pairs

table matching

# TOPCAT



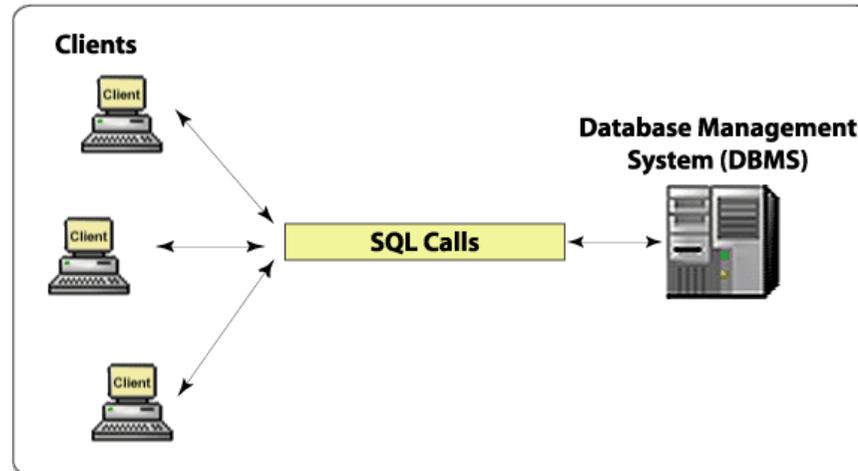
sky plot, colour coded



subsets

# SQL Queries

- **SQL** refers to query syntax allowing you to search and interact with large databases stored by a SQL-compatible server
- The **Structured Query Language** is very popular, originating outside of astronomy, and can handle very large databases plus many clients (commercial and open source implementations available)



- Many astronomical databases powered by SQL even if you may query it by other means (a webform or VO query)
- When simpler frontends fail, or once tables involved become very large, you may need to resort to performing your own SQL queries

# Database lingo

---

- Tables with rows and columns = Database with records and fields

ID	Source Class	Magnitude	Filter
1	1	12.34	1
2	3	16.89	2

- Records in separate tables may be connected to each other:

Class ID	Value
1	Star
2	Galaxy
3	Unknown

*Relational database*

Filter ID	Value
1	U
2	B
3	V
4	R

# SQL Queries : SDSS

- SDSS prime and early example of a large astronomical data resource that offered multiple ways of retrieving data, including custom SQL queries
- Tutorial covering performing SQL queries on SDSS data:  
<http://skyserver.sdss3.org/dr8/en/help/howto/search/>



The screenshot shows the Sloan Digital Sky Survey / SkyServer DR8 website. The page is titled "SQL Tutorial" and features a navigation menu with links for Home, Data, Schema, Education, Astronomy, SDSS, Contact Us, Download, Site Search, and Help. The main content area is divided into two columns. The left column contains a list of tutorial sections: 1. Introduction, 2. A Simple Query, 3. Common Searches, 4. More Samples, 5. Multiple Tables, 6. Aggregate Fcns., 7. Group By, 8. Order By, 9. Views, 10. Functions, and 11. Conclusion. The right column contains the text of the tutorial, starting with a note about the tutorial's purpose and a section titled "Searching for Data: A Tutorial" which explains how to use the SkyServer database to find specific objects.

Sloan Digital Sky Survey / SkyServer  
DR8

Home Data Schema Education Astronomy SDSS Contact Us Download Site Search Help

**SQL Tutorial**

NOTE: This is a step-by-step tutorial for those new to SQL. If you already know some SQL and want a more in-depth introduction to how SQL applies to SkyServer, see [Using SQL with SkyServer](#).

Another great way to learn SQL is to modify and submit queries that other people have already written. To jump right in to modifying queries written by scientists, see the [Sample SQL Queries](#).

**Searching for Data: A Tutorial**

Did you know that you can search through SkyServer's database for only the objects you are interested in? For example, you might want to find all the bright blue galaxies for which we have obtained spectra. Looking through all 469 million objects with the Navigation tool could take years. But using the [right search](#), you could find all the bright blue galaxies with spectra in seconds!

If you know how to retrieve the right data, you can ask millions of different questions. Searching SkyServer is an incredibly powerful tool for conducting astronomy research; in fact, many professional astronomers search the data in the same way you can.

In this tutorial, you will learn how to search SkyServer to answer many useful astronomical questions. Do the tutorial at your own pace, and use the links to the right to return to points later in the tutorial. As you work, you may also want to look at the [Sample Queries](#), which are examples of real queries that other people have written to search the database.

Click Next to start learning.

**SQL Tutorial**

[Back to Help](#)

1. Introduction
2. A Simple Query
3. Common Searches
4. More Samples
5. Multiple Tables
6. Aggregate Fcns.
7. Group By
8. Order By
9. Views
10. Functions
11. Conclusion

**NEXT**

# Table browser

Sloan Digital Sky Survey / SkyServer

DR8

Home Data Schema Education Astronomy SDSS Contact Us Download Site Search Help

## Schema Browser

FieldProfile  
FileGroupMap  
FIRST  
Frame  
galSpecExtra  
galSpecIndx  
galSpecInfo  
galSpecLine  
HalfSpace  
History  
IndexMap  
Inventory  
LoadHistory  
Mask  
MaskedObject  
Neighbors  
PartitionMap  
PhotoObjAll  
PhotoObjDR7  
PhotoPrimaryDR7  
PhotoProfile  
Photoz  
PhotozRF  
PhotozRFTemplateCoeff  
PhotozTemplateCoeff  
Plate2Target  
PlateX  
ProfileDefs  
ProperMotions  
PubHistory  
RC3  
RecentQueries  
Region  
Region2Box  
RegionArcs  
RegionPatch  
Rmatrix  
ROSAT  
Run  
RunShift  
SDSSConstants  
sdssImagingHalfSpaces  
sdssPolygon2Field  
sdssPolygons  
sdssTargetParam  
sdssTileAll  
sdssTiledTargetAll  
sdssTilingGeometry  
sdssTilingInfo  
sdssTilingRun  
Sector  
segueTargetAll  
SiteConstants  
SiteDBs  
SiteDiagnostics  
SpecDR7

TABLE **PhotoObjAll**

The full photometric catalog quantities for SDSS imaging.

This table contains one entry per detection, with the associated photometric parameters measured by PHOTO, and astrometrically and photometrically calibrated.

The table has the following views:

- **PhotoObj**: all primary and secondary objects; essentially this is the view you should use unless you want a specific type of object.
- **PhotoPrimary**: all photo objects that are primary (the best version of the object).
  - **Star**: Primary objects that are classified as stars.
  - **Galaxy**: Primary objects that are classified as galaxies.
  - **Sky**: Primary objects which are sky samples.
  - **Unknown**: Primary objects which are noOne of the above
- **PhotoSecondary**: all photo objects that are secondary (secondary detections)
- **PhotoFamily**: all photo objects which are neither primary nor secondary (blended)

The table has indices that cover the popular columns.

name	type	length	unit	ucd	description
objID	bigint	8			Unique SDSS identifier composed from [skyVersion, rerun, run, camcol, field, obj].
skyVersion	tinyint	1			Layer of catalog (currently only one layer, 0; 0-15 available)
run	smallint	2			Run number
rerun	smallint	2			Rerun number
camcol	tinyint	1			Camera column
field	smallint	2			Field number
obj	smallint	2			The object id within a field. Usually changes between reruns of the same field
mode	tinyint	1			1: primary, 2: secondary, 3: other
nChild	smallint	2			Number of children if this is a composite object that has been debbled. BRIGHT (in a flags sense) objects also have nchild == 1, the non-BRIGHT sibling.
type	smallint	2			Type classification of the object (star, galaxy, cosmic ray, etc.)
clean	int	4			Clean photometry flag (1=clean, 0=unclean).
probPSF	real	4			Probability that the object is a star. Currently 0 if type == 3 (galaxy), 1 if type == 6 (star).
					Flag to indicate whether object is inside a mask and

## Schema Browser

Need to know what Tables are available and which fields are stored in those tables

# Basic Query

---

**SELECT**

ra, dec

which fields do you want returned?

**FROM**

specObj

from which table?

**WHERE**

ra BETWEEN 140 and 141

**AND**

dec BETWEEN 20 and 21

**AND**

class='STAR'

boolean constraints  
on fields that records  
must satisfy

# Relational Query

---

## SELECT

```
specObj.fiberID,  
PhotoObj.modelMag_u,  
PhotoObj.modelMag_g,  
PhotoObj.modelMag_r,  
PhotoObj.modelMag_i,  
PhotoObj.modelMag_z,  
PhotoObj.ra, PhotoObj.dec,  
specObj.z, PhotoObj.ObjID
```

which fields do you want returned?  
[Table.Field]

## FROM

```
PhotoObj, specObj
```

from which tables?

## WHERE

```
specObj.bestObjid = PhotoObj.ObjID
```

need to ensure we  
compare same object

## AND

```
specObj.class = 'qso'
```

## AND

```
specObj.zWarning = 0
```

## AND

```
specobj.z between 0.3 and 0.4
```

boolean constraints  
on fields that records  
must satisfy  
[Table.field = Value]

# Result

Your SQL command was:

```
SELECT
specObj.fiberID, PhotoObj.modelMag_u, PhotoObj.modelMag_g,
PhotoObj.modelMag_r, PhotoObj.modelMag_i, PhotoObj.modelMag_z,
PhotoObj.ra, PhotoObj.dec, specObj.z, PhotoObj.ObjID
FROM
PhotoObj, specObj
WHERE
specObj.bestObjid = PhotoObj.ObjID
AND
specObj.class = 'qso'
AND
specObj.zWarning = 0
AND
specObj.z between 0.3 and 0.4
```

fiberID	modelMag_u	modelMag_g	modelMag_r	modelMag_i	modelMag_z	ra	dec	z	ObjID
225	21.608122	20.919027	20.106173	19.625843	19.161301	215.97274128	-0.96643655	0.391129	1237648720172745195
640	20.437267	20.066856	19.869516	19.683071	19.237711	195.66483383	0.40463547	0.374446	1237648674510995921
221	14.929588	12.896034	12.344924	12.255587	13.574942	6.40105152	-0.46343099	0.383457	1237657190371164165
218	20.620367	19.742416	18.909544	18.444504	17.982805	6.55951741	-0.98912516	0.372924	1237657189834358887
512	17.007906	16.777861	16.785234	16.746496	16.370136	23.57580286	0.26019863	0.399601	1237657071157444676
189	18.119993	17.833719	17.874315	17.67757	17.250725	33.24829998	-0.50817531	0.394522	1237657070087962632
151	20.841267	20.093006	18.863449	18.384741	17.953033	337.42612345	-0.71003756	0.374328	1237656567586095748
192	19.786795	19.552595	19.321449	19.210716	18.773647	233.74375738	2.70391221	0.389854	1237654879671550196
332	19.542379	19.182779	18.850351	18.602491	18.349787	143.53824274	2.5436351	0.381111	1237654599949942970
513	18.413069	18.271866	18.284575	18.26753	17.632139	145.17885592	2.26594343	0.385768	1237653665256767497
85	20.295122	19.745787	19.058146	18.573439	18.002174	166.94072826	60.1539464	0.397834	1237653616398696678
42	20.057171	19.952911	19.487123	19.267277	18.884089	348.67848826	-10.51547783	0.377506	1237653436008759550
382	19.270554	19.143511	18.722548	18.43214	18.025404	49.6711526	-6.67914789	0.390237	1237652901306499242
601	20.580416	20.056355	19.482079	19.14917	18.849903	45.02323702	-7.20430752	0.374826	1237652901304467725
585	20.303259	19.847921	19.321499	19.023911	18.627033	28.93484124	-8.68499508	0.381934	1237652901297520731

# Powerful Queries

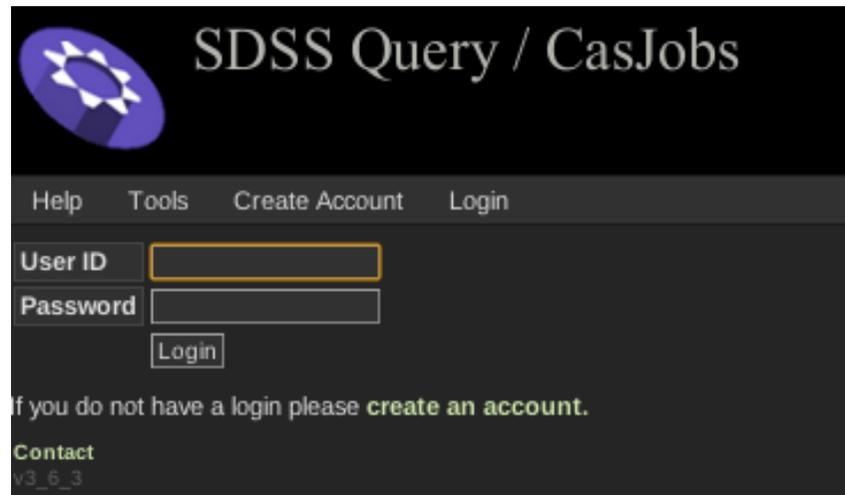
---

```
SELECT TOP 10
run,camCol,rerun,field,objID,ra,dec
FROM Galaxy
WHERE ( ( flags & (dbo.fPhotoFlags('BINNED1')
| dbo.fPhotoFlags('BINNED2')
| dbo.fPhotoFlags('BINNED4')) ) > 0
and ( flags & (dbo.fPhotoFlags('BLENDED')
| dbo.fPhotoFlags('NODEBLEND')
| dbo.fPhotoFlags('CHILD')) ) != dbo.fPhotoFlags('BLENDED')
and ( flags & (dbo.fPhotoFlags('EDGE')
| dbo.fPhotoFlags('SATURATED')) ) = 0
and petroMag_i > 17.5
and (petroMag_r > 15.5 or petroR50_r > 2)
and (petroMag_r > 0 and g > 0 and r > 0 and i > 0)
and ( (petroMag_r-extinction_r) < 19.2
and (petroMag_r - extinction_r <
(13.1 + (7/3) * (dered_g - dered_r) + 4 * (dered_r - dered_i)
- 4 * 0.18) )
and ( (dered_r - dered_i - (dered_g - dered_r)/4 - 0.18) < 0.2)
and ( (dered_r - dered_i - (dered_g - dered_r)/4 - 0.18) > -0.2)
-- dered_ quantities already include reddening
and ( (petroMag_r - extinction_r +
2.5 * LOG10(2 * 3.1415 * petroR50_r * petroR50_r)) < 24.2) )
or ( (petroMag_r - extinction_r < 19.5)
and ( (dered_r - dered_i - (dered_g - dered_r)/4 - 0.18) > (0.45 - 4 *
(dered_g - dered_r)) )
and ( (dered_g - dered_r) > (1.35 + 0.25 * (dered_r - dered_i)) ) )
and ( (petroMag_r - extinction_r +
2.5 * LOG10(2 * 3.1415 * petroR50_r * petroR50_r) ) < 23.3 ) )
```

# CasJobs

---

- No limit on number of records returned or the time it takes for your Query to run
- You can also upload and create your own Tables, then join those in SQL queries



The screenshot shows the SDSS Query / CasJobs interface. At the top left is a purple circular logo with a white gear-like pattern. To its right, the text "SDSS Query / CasJobs" is displayed in a light grey font. Below the logo and title is a navigation bar with links for "Help", "Tools", "Create Account", and "Login". The main content area contains a login form with two input fields: "User ID" and "Password". Below these fields is a "Login" button. A message below the form reads: "If you do not have a login please [create an account](#)." At the bottom left, there is a "Contact" link and the version number "v3\_6\_3".

- Note: it is very easy to write a very complex Query so appreciate what you are asking from the server
- The SDSS/SQL model is used by others, such as GALEX, MAST, VISTA etc

# Other example ; UKIDSS @ WFAU

The screenshot displays the WFCAM Science Archive website. At the top, there is a navigation bar with links: Home | Overview | Browser | Access | Login | Cookbook | nonSurvey, and a WSA logo. The left sidebar contains a vertical menu with various links: WSA Home, Start Here, Data Overview, Known Issues, the Surveys, Schema browser, Data access, Login, Archive Listing, GetImage, ColourImage, MultiGetImage, Region, Menu query, Freeform SQL, CrossID, Analysis services, SQL Cookbook, Q&A, Glossary, Release History, non-Survey, Gallery, Publications, Monitor, Downtime, and Links. The main content area is titled 'Freeform SQL Query' and includes a status message: 'Status: Not logged in. Please reload this page if you have logged in and are not seeing the correct login status.' Below this is a note: 'The GPS and UDS are not publically accessible in UKIDSSDR7plus. Users wishing to access GPS/UDS data should use the DR6plus/DR5plus databases respectively.' The form itself has a dropdown menu for 'Database release to use:' set to 'UKIDSSDR7PLUS'. It features an 'Upload SQL query from file into this form:' section with a 'Choose File' button and 'No file chosen' text, and an 'Upload' button. Below this is a text area for 'or enter SQL statement:'. At the bottom of the form, there is a 'Submit' button and instructions: 'ensure one of the file formats is selected below if you want to save your results.' The 'Email Address:' field is partially obscured by a redaction box, with the text 'the results of long running queries will be sent by email.' The 'Data Format:' section has radio buttons for: HTML table summary (results are NOT saved to file), ASCII FILE (downloadable with HTML table summary on-screen), FITS FILE (downloadable with HTML table summary on-screen), and VOTable FILE (downloadable with HTML table summary on-screen). A note in italics states: '( The number of rows written to the downloadable files is subject to an upper limit see data access )'. The 'File Compression:' section has radio buttons for: (none) and GZIP. At the bottom left of the page, there is a WFAU logo and the text 'IFA ROE'.

# Lecture 5 Assignment (part 2 of 2)

---

- Using TOPCAT Retrieve a 5' cone of data from the SDSS DR7 photometry (`sdsdr7-sda` service, `PhotoObjAll` catalog) centered on 01:30:56, +43:55:49
- Merge the resulting table with 2MASS (`2mass-psc`) using a sky matching radius of 2''
- Make a scatter plot of SDSS (u-g) colour versus 2MASS (J-K) colour, limited to sources brighter than SDSS  $g=20$
- Consider the source that is most red in the near-infrared and comment on its optical colour relative to other sources in the field