



**Data Mining,  
ADQL,  
&  
The National Virtual  
Observatory's  
OpenSkyQuery Utility**

**by Richard “Doc” Kinne, KQR  
2008 AAVSO Fall Conference  
Nantucket, MA**

# Most Folks Thoughts on Data...





# Real Situation in Five Years...



# Need for Standard Database Operations

- For Business
- For Astronomy



# SQL – The Win!

- Standard Query Language
- Used With...
  - SQL Anywhere – Commercial
  - PostgreSQL
  - mySQL – used by Amazon.com...and the AAVSO!

# SQL – What Is It?

- A Means of Interacting with, and creating, a Standard Relational Database.
- Relational Database? A linked series of 2-dimensional tables – linked, stacked Excel spreadsheets.
- SQL – a language that allows you to create, modify, read, and query those tables of data.

# SQL Example Syntax

- create database varstar;  
use varstar;  
create table observation  
{  
  name varchar(30),  
  mag float,  
  dec varchar(30),  
  ra varchar(30),  
  jd float not null  
} type=InnoDB;



# SQL Example Syntax

- use varstar;  
insert into observation values ("Z Uma", 7.4,  
"57:52:18.0", "11:56:30.2", 2454728.4);



# SQL Example Syntax

- use varstar;  
delete from observation where name = "Z UMa";

# ADQL

- ADQL – Astronomical Data Query Language.
- A Subset of SQL 92
- Differences?
  - Intended for read-only queries only. Actually very simple.
  - Astronomical Extensions have been added.
    - Spatial areas: Regions & Catalog crossmatching

# ADLQ – Example Syntax

- Select o.objId, o.ra, o.r, o.type, t.objId from SDSS:PhotoPrimary o, TWOMASS:Photoprimary, t where  $xmatch(o,t) < 2.5$  and Region ('Circle J2000 181.3 -0.76 6.5') and o.type=3;
- What the heck does this mean?



# ADLQ – Example Explanation

- Select o.objId, o.ra, o.r, o.type, t.objId  
from SDSS:PhotoPrimary o, TWOMASS:Photoprimary, t  
where xmatch(o,t) < 2.5  
and Region ('Circle J2000 181.3 -0.76 6.5')  
and o.type=3;
- SDSS:Photoprimary is a table now called “o”
- TWOMASS:Photoprimary is a table now called “t”
- Select the objId, ra, r, and type columns out of “o”
- Select the objId column out of “t”
- These are the columns in the two tables we're going to look at.

# ADLQ – Astronomical Extension

- Select o.objId, o.ra, o.r, o.type, t.objId from SDSS:PhotoPrimary o, TWOMASS:Photoprimary, t where  $xmatch(o,t) < 2.5$  and Region ('Circle J2000 181.3 -0.76 6.5') and o.type=3;
- XMATCH(o,t)<2.5
  - Function will calculate the chi-squared probability that the objects in these two catalogs are actually the same object.
  - <2.5 means that the detection will be rejected if the standard deviation between the two is more than 2.5.

# ADLQ – Astronomical Extensions

- Select o.objId, o.ra, o.r, o.type, t.objId from SDSS:PhotoPrimary o, TWOMASS:Photoprimary, t where  $xmatch(o,t) < 2.5$  and Region ('Circle J2000 181.3 -0.76 6.5') and o.type=3;
- Region ('Circle J2000 181.3 -0.76 6.5') – does the position for the current record fall in this region of the sky?
- Circle J2000, RA, DEC, r of circle in dec. degrees.



# OpenSkyQuery

- OpenSkyQuery is a Data Mining Utility of the Virtual Observatory.
- OpenSkyNode supports *distributed queries*! DBs are hosted & maintained by their institutions but readable by everyone.
- In order for a DB to be part of it, it needs to have a SkyNode interface and be published to the VO Registry.
- Once this is done, it appears on a list of databases that can be searched via OpenSkyQuery.
- OpenSkyQuery is at <http://openskyquery.net/>

# OpenSkyQuery – A Simple Example



Open SkyQuery

Home

Simple Query

Advanced Query

Import

Tutorial

National Virtual Observatory

## Simple Query Form

Select a single catalog to query, or up to three catalogs to cross-match, then press Next.

Catalog 1 : Catalog



Catalog 2 : Catalog



Catalog 3 : Catalog

Next -- >

### Nodes

- Rosat
- XMM
- GALEX
- GALEXGR1
- DLS
- RC3
- GSC2
- NBCKEDR1
- SDSS
- SDSSDR2
- SDSSDR3
- SDSSDR4
- TwoDf
- Twoqz
- TWOSLAQLRGEDR



Sponsored all or in part by the National Science Foundation under Cooperative Agreement AST0122449 with The Johns Hopkins University.

Developed in collaboration with the International Virtual Observatory Alliance.

Contact the NVO Help Desk to report problems and suggestions.

Last Modified: Friday, April 07, 2006 at 4:24:01 PM \$Name: v1\_0\_9\_7 \$ Revision 1.16





# OpenSkyQuery – A Simple Example

**Simple Query Form**

Catalog 1 : SDSS    Catalog 2 : FIRST    Catalog 3 : TWOMASS

**CrossMatch :**

SDSS    FIRST    TWOMASS (NOTE: For dropouts, uncheck the catalog! )

Sigma :

**Region :**

Ra :  deg.    Dec :  deg.    Radius :  arcmin

**Select parameters and constraints :**

SDSS			
Select	Min	Attr	Max
<input checked="" type="checkbox"/>	<input type="text"/>	u	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	g	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	r	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	i	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	z	<input type="text"/>

FIRST			
Select	Min	Attr	Max
<input checked="" type="checkbox"/>	<input type="text"/>	fint	<input type="text"/>

TWOMASS			
Select	Min	Attr	Max
<input checked="" type="checkbox"/>	<input type="text"/>	j_m	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	h_m	<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text"/>	k_m	<input type="text"/>

**Check and run your query:**

```
SELECT
a.objid,a.ra,a.dec,a.u,a.g,a.r,a.i,a.z,b.objid,b.ra,b.dec,b.fint,c.objid,c
FROM SDSS:photoprimary a,FIRST:photoprimary b,TWOMASS:photoprimary c
WHERE XMATCH(a,b,!c)<3.5 AND Region('CIRCLE J2000 181.3 -0.76 30')
```

**Nodes**

- Rosat
- XMM
- GALEX
- GALEXGR1
- DLS
- RC3
- GSC2
- NBCKDEDR1
- SDSS
- SDSSDR2
- SDSSDR3
- SDSSDR4
- TwoDf
- Twoqz
- TWOSLAQLRGEDR
- TWOSLAQQSOEDR
- USNOB
- GOODS
- HDFN
- HDFS
- UDF
- TWOMASS
- IRAS
- PSCz
- FIRST
- NVSS
- SUMSS
- LCATheory
- NDWFS
- SAGE



# OpenSkyQuery – A Simple Example

HTML

44 row(s)

sdss_objid	sdss_ra	sdss_dec	sdss_u	sdss_g	sdss_r	sdss_i	sdss_z	fit
587722982277513779	181.280964329843	-0.775161775579332	26.59848	23.48722	22.30811	21.48366	21.06459	
588848898839478741	181.241753429695	-0.895253982374432	22.75468	22.23607	21.08645	20.3742	20.21942	
588848899376415277	181.395560517706	-0.618968107715356	22.60993	22.11696	20.3791	19.39062	18.85863	
588848899376415271	181.393855203689	-0.61742038695532	22.54627	22.63787	21.0656	20.0659	19.78067	
588848899376415277	181.395560517706	-0.618968107715356	22.60993	22.11696	20.3791	19.39062	18.85863	
588848898839413066	181.158631722352	-0.873301373473849	21.84024	22.21916	21.38232	20.8104	21.38801	
588848898839544419	181.449753252085	-0.860114981262776	23.96865	21.61563	20.03169	19.22029	18.84507	
588848898839543917	181.452000649341	-0.860159402334619	23.65605	22.46628	21.36011	20.83372	20.81125	
588848899376349612	181.222654063796	-0.574898585088707	25.42542	22.21821	21.58511	21.27709	21.3112	
588848899376349206	181.257280892827	-0.537363871624612	19.77114	19.62425	19.36931	19.3394	19.29524	
588848899376284517	181.100151529517	-0.615921152282603	26.13067	24.42188	22.47922	21.92833	21.48851	
587722982277644787	181.546871355706	-0.734921765861926	22.5851	22.94553	21.21888	20.39603	19.54589	
588848898839544345	181.408718502499	-0.993146105058267	22.74373	21.7	20.01053	19.32947	18.89978	
588848898839544346	181.408058073042	-0.992703320749044	24.96557	21.72807	20.22829	19.74157	19.07533	
588848898839609792	181.534254735305	-0.873295134776541	24.41793	22.79921	23.10035	21.95683	22.10049	
588848898839610404	181.536192688924	-0.878429757406259	23.38247	22.94795	22.39088	22.20344	21.41267	
587722982277513779	181.280964329843	-0.775161775579332	26.59848	23.48722	22.30811	21.48366	21.06459	



# Resources

- US National Virtual Observatory - <http://www.us-vo.org/>
- *The National Virtual Observatory: Tools and Techniques for Astronomical Research* – Available from the US VO website.
- OpenSkyQuery Tutorial at the website - <http://www.openskyquery.org/>
- Latest ADQL Specification - <http://www.ivoa.net/Documents/latest/ADQL.html>

# Questions?

