SKYMAPPER IN THE VIRTUAL OBSERVATORY:
HOW WILL YOU ACCESS THE TERABYTES?

SIMON MURPHY (HEIDELBERG)
FEAT. **JON SMILLIE** (ANU), **YESHE FENNER** (AAL)
& THE **ALL SKY VIRTUAL OBSERVATORY** PROJECT

ALL SKY VIRTUAL OBSERVATORY

OUTLINE

- Introduction to the Virtual Observatory
- SkyMapper data products
- Complex queries using TAP/ADQL
- Federating data
- Interoperability with SAMP *(if time permits)*
- Feedback - you are the users!

ALL SKY VIRTUAL OBSERVATORY

THE VIRTUAL OBSERVATORY

- What is the Virtual Observatory?
  "all astronomy data products on your computer"

- A set of globally-agreed standards for:
  - *Describing resources*
  - *Publishing resources*
  - *Finding resources*
  - *Accessing resources*

- **Interoperability** is the key concern

ASVO
ALL SKY VIRTUAL OBSERVATORY

THE VIRTUAL OBSERVATORY

LEVEL 2

USER LAYER

Browser Based Apps | *Desktop Apps* | *Script Based Apps*

SAMP | SSO | USING | CDP | WS BP

REGISTRY

Registry Interface

Resource Metadata

VOResource

VODataService

ApplicationRegExt

StandardRegExt

SimpleDALRegExt

VOSI

ADQL

PQL

VO Query Languages

UCD

Semantics

Vocabularies

VO CORE

STC | Utypes | Units

CharDM | SpectrumDM

ObsCoreDM

ObsProvDM | SSLDM

PhotDM | VOEvent

Formats

Resource Identifier | VOTable

SimDM

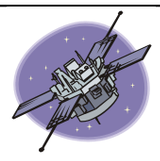
SIAP | SCS | SSAP | TAP | SLAP | SEAP | SimDAL | FAP

DATA ACCESS PROTOCOLS

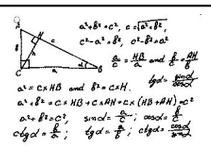
VOSpace | VOPipe | SHARING | UWS

Data and Metadata Collection

Storage | RESOURCE LAYER | *Computation*



PROVIDERS



20101004
IVOA Architecture

ALL SKY VIRTUAL OBSERVATORY

YESHE FENNER

ASTRONOMY AUSTRALIA LIMITED

What is the All Sky Virtual Observatory?

Platform for connecting Australian datasets with users and the world-wide VO network, via IVOA standards and services, and customised web interfaces.

History:

2011: Mid-Term Review of the Decadal Plan 2006-2015 recommended:

> *"Astronomy should build an astronomical data fabric that links high-performance resources through appropriate data middleware and networks to create new opportunities for discovery by Australian researchers based on data flowing from telescopes like SkyMapper, ASKAP and the MWA. A path forward to achieving this goal is for the community to actively engage in the Government's eResearch initiatives, such as NeCTAR and RDSI"*

2012: AAL commissioned an eResearch software development company *Intersect* to conduct a design study into a "Federation of National Astronomy Datasets"

2012: AAL secured NeCTAR funds to take a first step towards this vision, by building the $3.5M All Sky Virtual Observatory

Sponsors: Australian Government Department of Industry Innovation, Science, Research and Tertiary Education; nectar

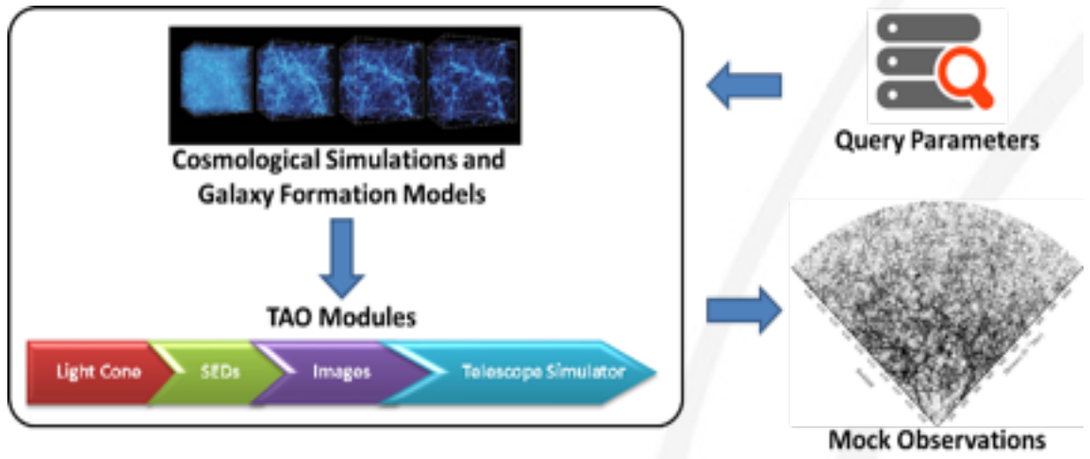
Partners: Astronomy Australia Ltd.; Swinburne University of Technology; NCI National Computational Infrastructure; The Australian National University; Intersect

ASVO

ASVO v1.0

Initial ASVO datasets chosen in consultation with AAL's eResearch Advisory Committee and the community:

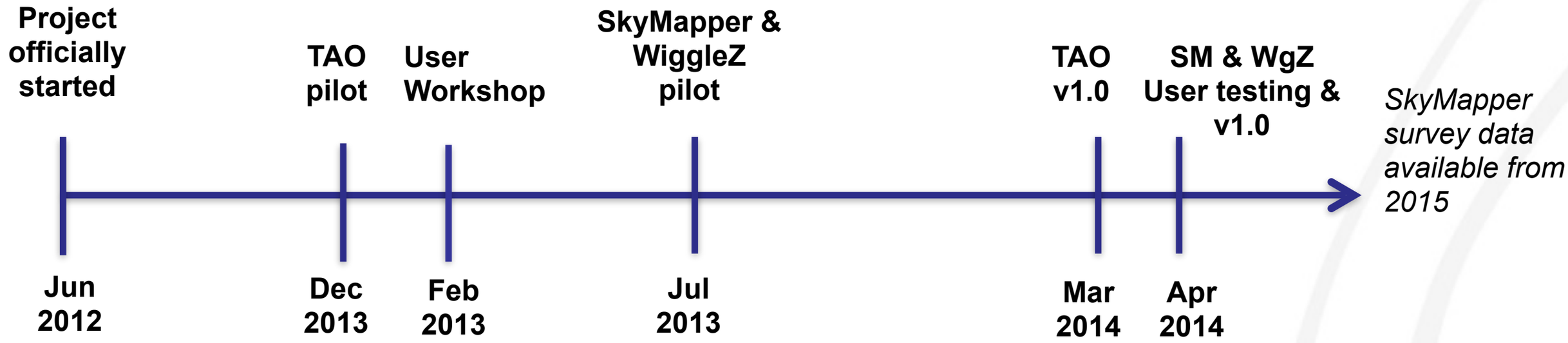
1. The **Theoretical Astrophysical Observatory** (TAO https://tao.asvo.org.au) builds virtual universes based on simulation and model datasets. Developed at Swinburne. Launched 24th March 2014.
2. Optical data from ANU's **SkyMapper Southern Sky Survey** (+ WiggleZ spectroscopic survey). Housed at NCI. First ASVO-SkyMapper public data release ~2015.



TAO schematic

SkyMapper telescope

Timeline and next steps



Future possibilities:

- More functionality for TAO and SkyMapper Nodes
- Data access services for other Australian datasets, e.g. next-gen radio telescopes

THE VIRTUAL OBSERVATORY

- **How is this relevant to science with SkyMapper?**

  - *VO tries to make <u>science</u> as easy as possible*

  - *SkyMapper access primarily through the (AS)VO*

  - *A lot of other data are already in the VO (catalogues, images, spectra, simulations, …)*

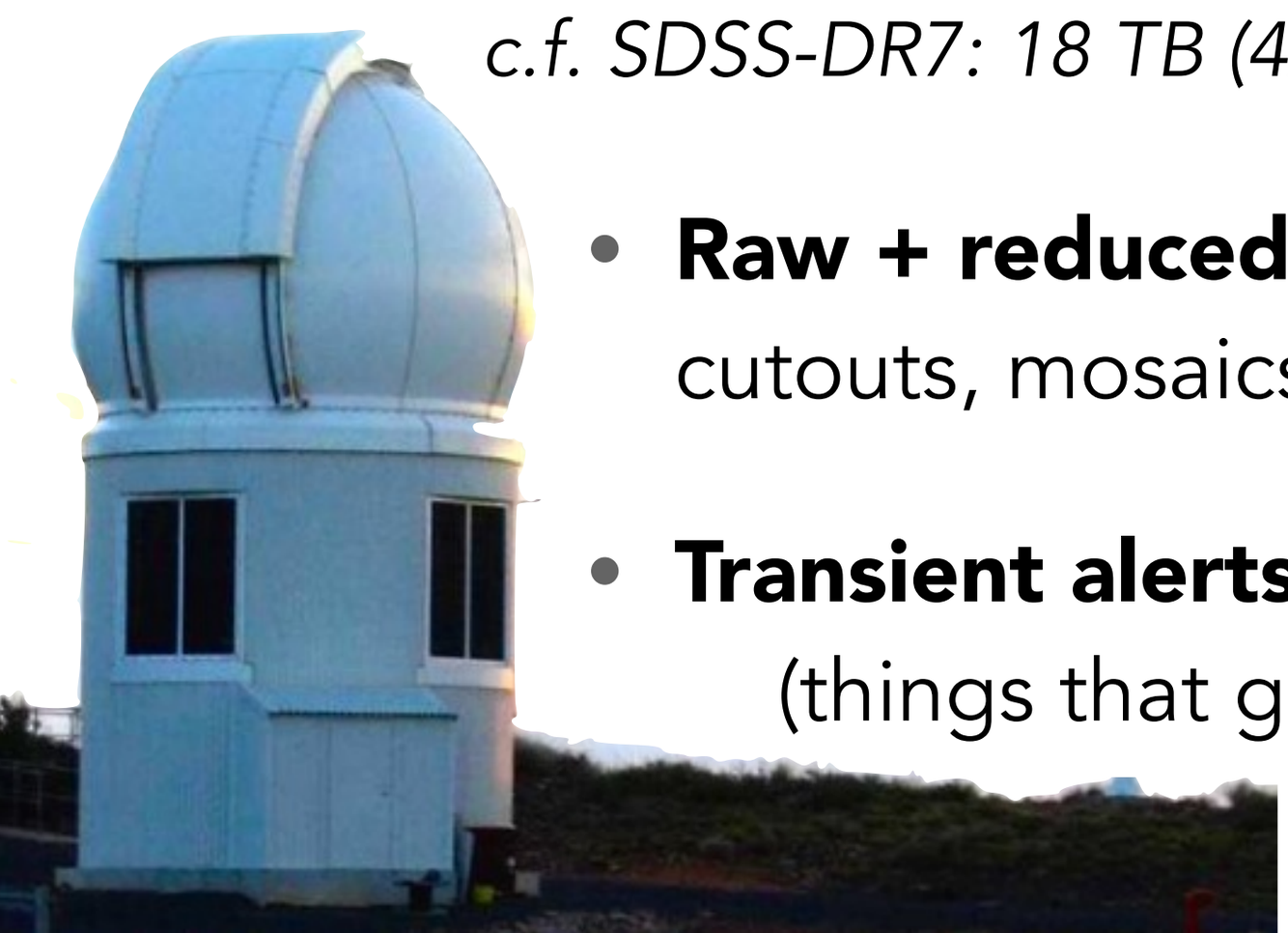
Multi-wavelength, multi-dataset science

ASVO
ALL SKY VIRTUAL OBSERVATORY

SKYMAPPER DATA PRODUCTS

- Shallow and Main Survey data releases (from late 2015?)
- **Catalogue access** first priority, then images
- 10 tables x 10^9 rows x ~10^2 cols → **50-100 TB database**

c.f. SDSS-DR7: 18 TB ($4x10^8$ srcs); DR9: 60 TB (10^9 srcs)



- **Raw + reduced images** (**1-2 PB**), plus cutouts, mosaics, RGB, photometry service?
- **Transient alerts**

  (things that go bump in the night)

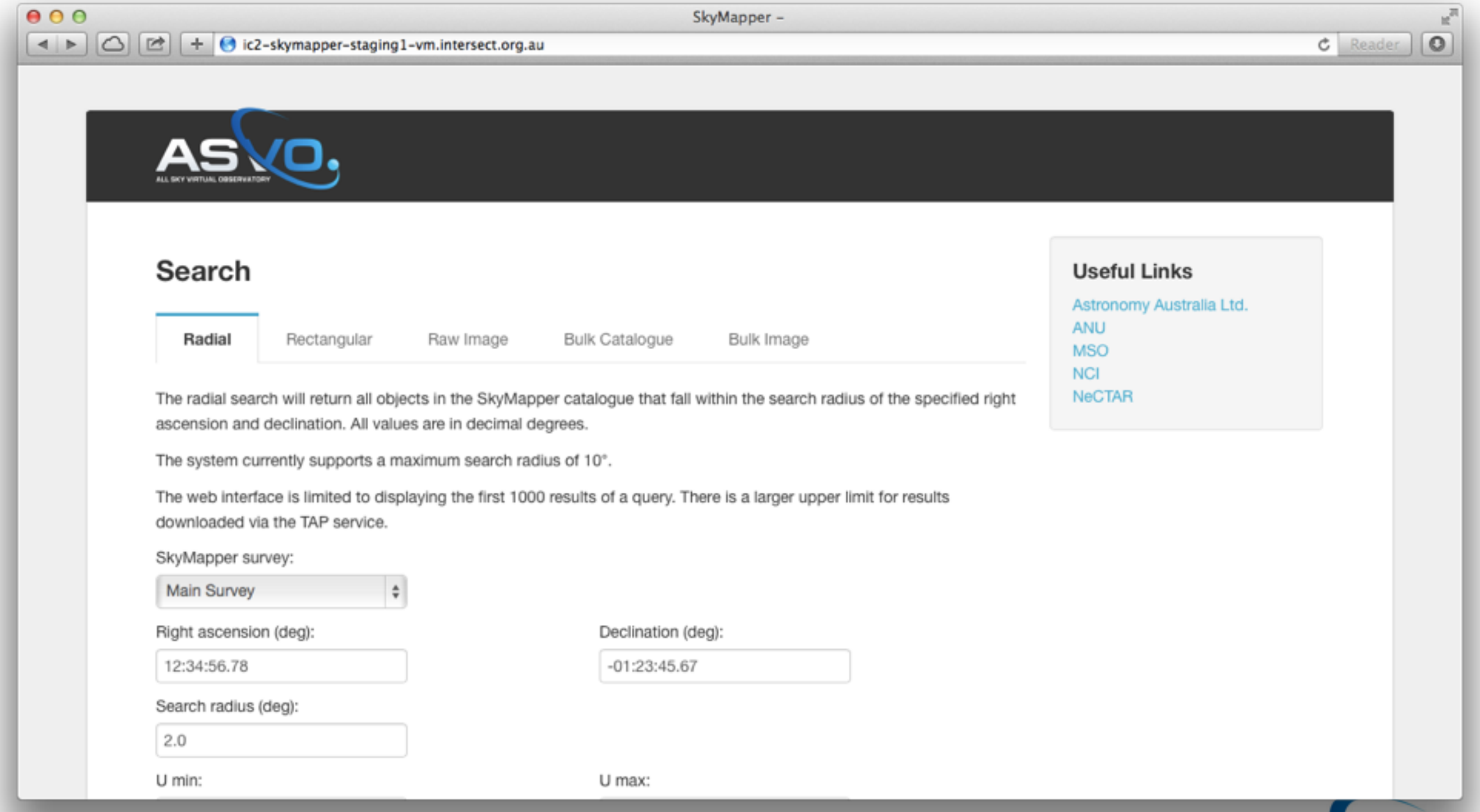
ASVO
ALL SKY VIRTUAL OBSERVATORY

SKYMAPPER DATA PRODUCTS

- Under development within the ASVO project:
  - **High Performance Data Archive** *(internal+release)*
  - Cloud Analysis *(near data, BYO-software)*
  - **Data Exploration Tools**
    - Web forms (simple queries)
    - Direct database access (local SQL)
    - **VO services (images, catalogues)**

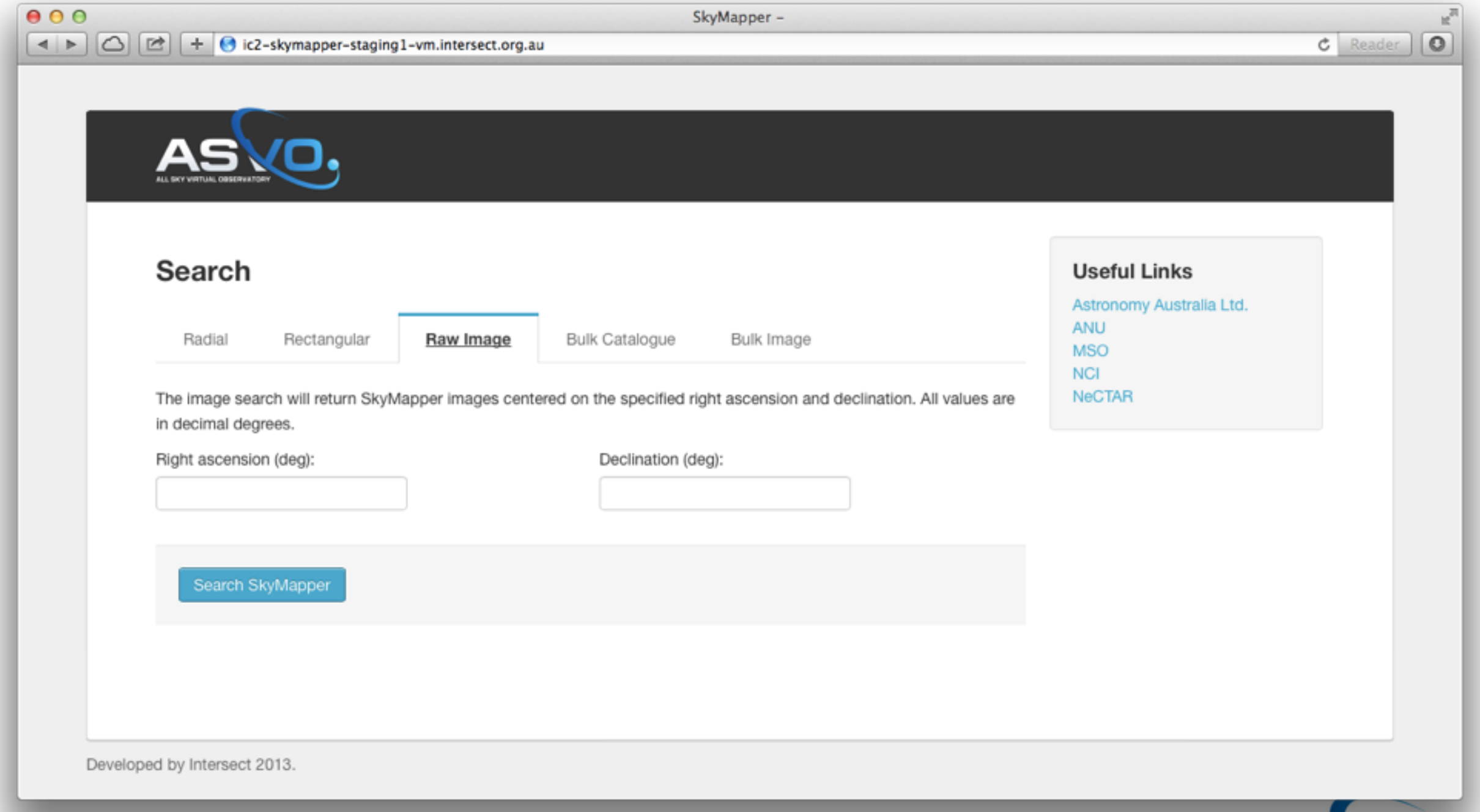
ALL SKY VIRTUAL OBSERVATORY

WEB ACCESS: CONE SEARCH



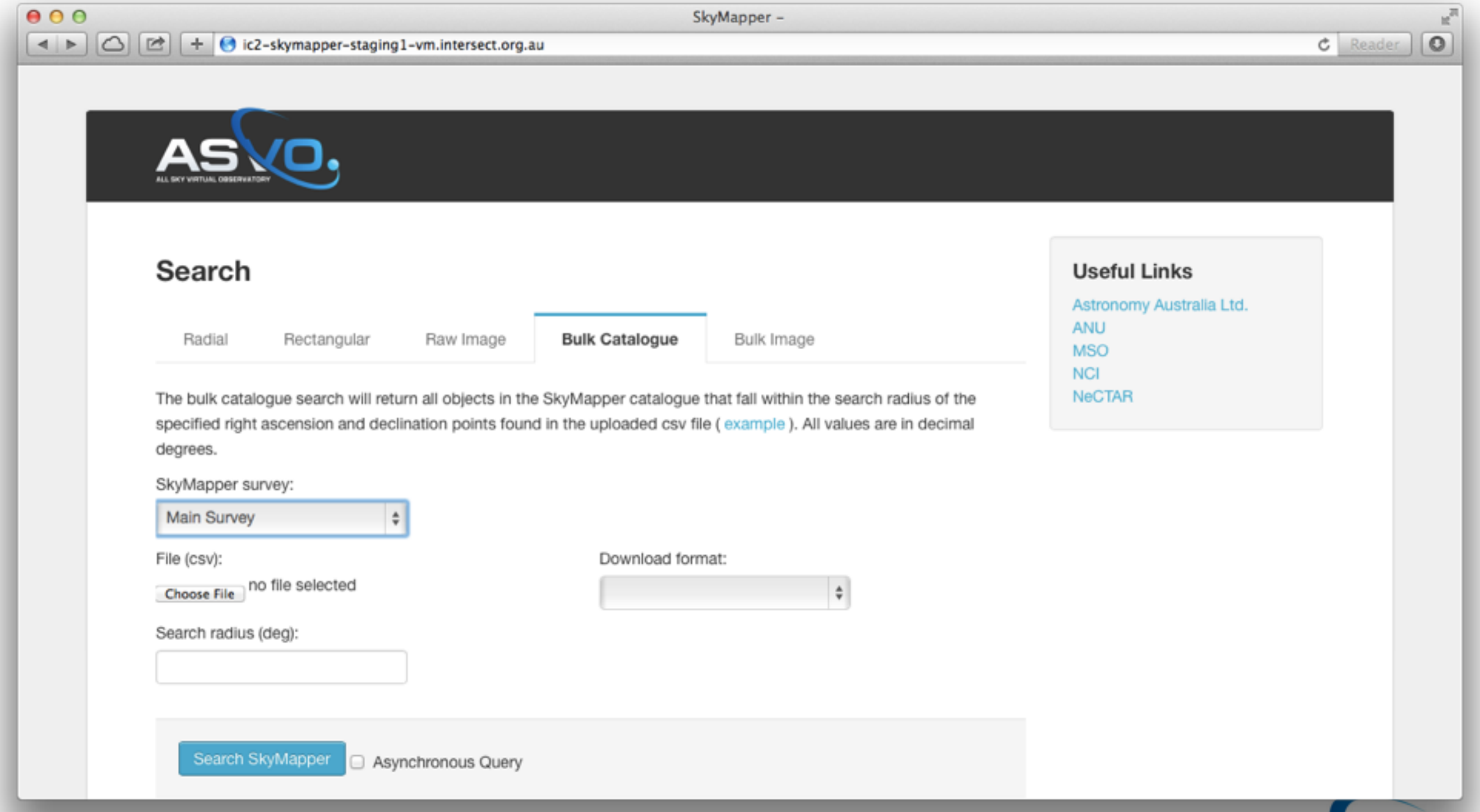
ASVO
ALL SKY VIRTUAL OBSERVATORY

WEB ACCESS: IMAGES



ASVO
ALL SKY VIRTUAL OBSERVATORY

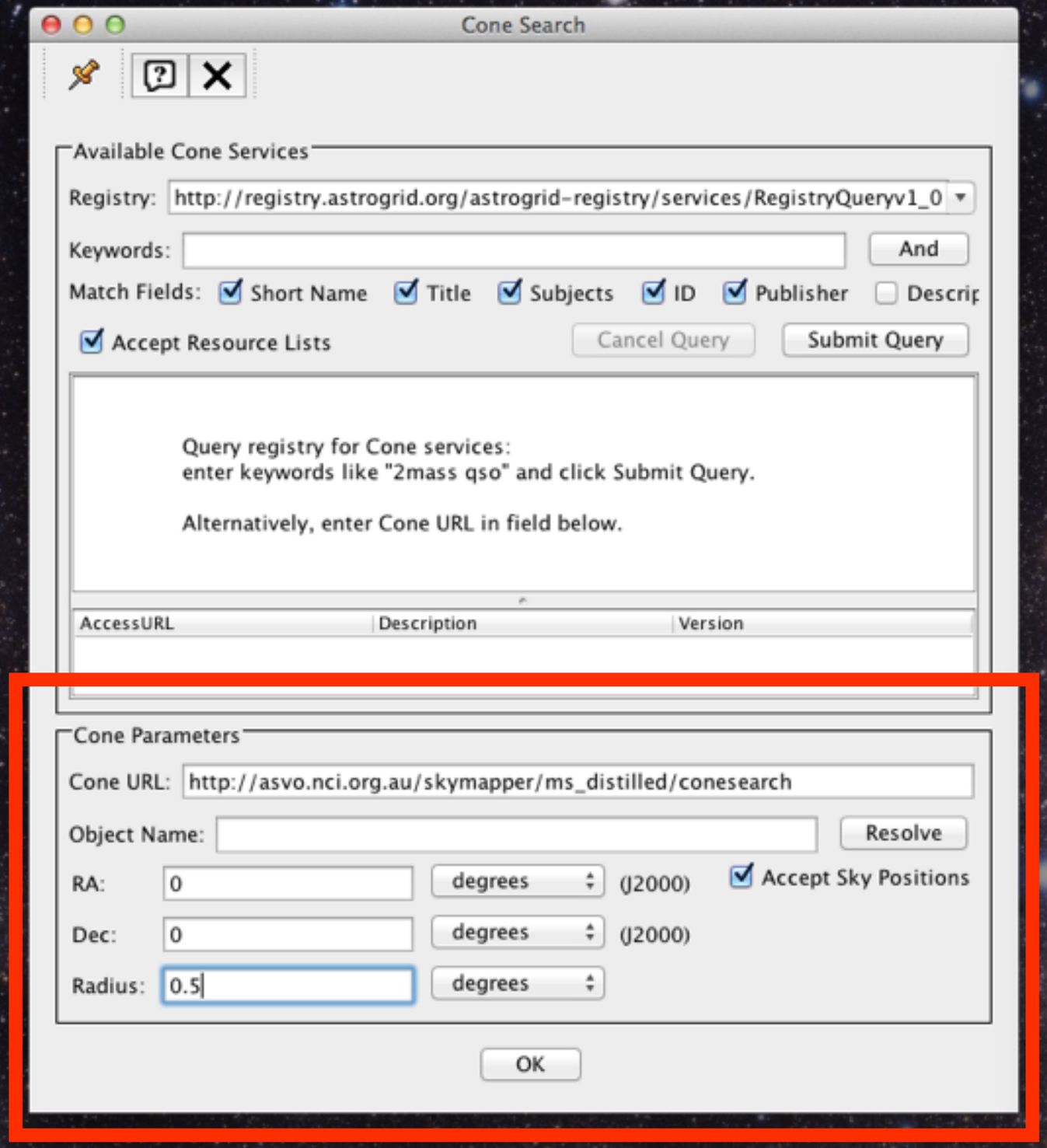
WEB ACCESS: CROSS MATCH

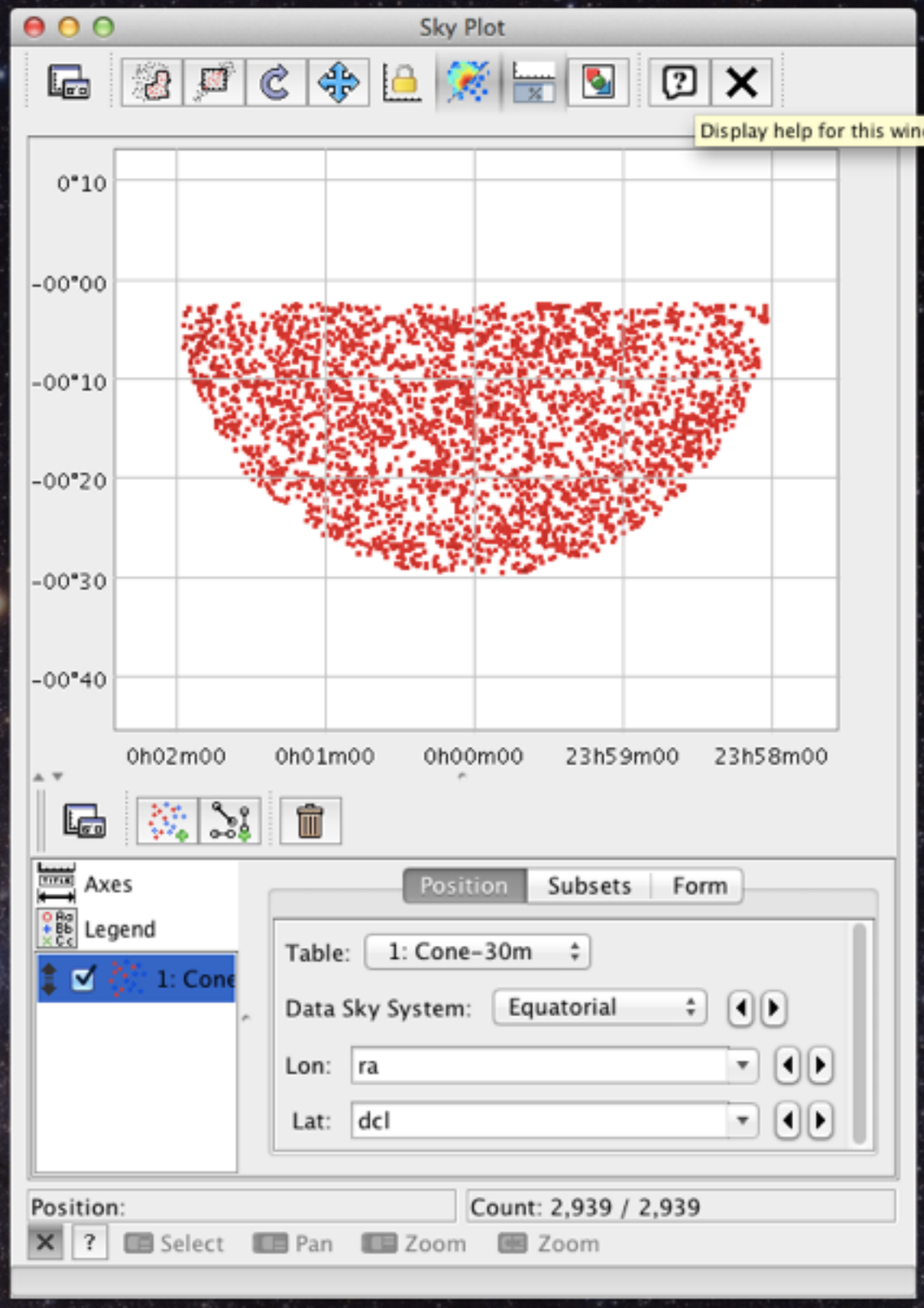


ASVO
ALL SKY VIRTUAL OBSERVATORY

DEMO

SKYMAPPER CONE SEARCH WITH TOPCAT





Cone Search URL (in registry mid-2014):

http://asvo.nci.org.au/skymapper/ms_distilled/conesearch

COMPLEX TABULAR QUERIES

SHOW ME BLUE VARIABLE STARS AT HALO DISTANCES

WHAT OBSERVATIONS WENT INTO THIS CATALOGUE ENTRY?

FIND ALL HIGH PROPER MOTION OBJECTS WITHOUT 2MASS PHOTOMETRY

CONSTRUCT THE SED OF THIS GALAXY FROM UV TO RADIO

IS THERE A DWARF GALAXY IN THIS REGION?

IDENTIFY HIGH-Z QUASARS IN THIS IRREGULAR SKY AREA

ASVO
ALL SKY VIRTUAL OBSERVATORY

COMPLEX TABULAR QUERIES

- VO standard: **TAP (Table Access Protocol)**
- SQL-like access to multi-table databases:
  ADQL (Astronomical Data Query Language)
- Complex spatial queries
  (cross matching, radial/polygonal, set geometry)
- User uploaded tables
- Asynchronous job handling
- **Services & clients already available**

ASVO
ALL SKY VIRTUAL OBSERVATORY

ADQL IN ONE SLIDE

- Similar syntax to SQL (commercial DBs, CASJobs):

```
SELECT new_object_id,ra,dcl,a/b
FROM public.ms_distilled
WHERE sigma_g/mean_g BETWEEN 0.1 AND 0.2
    AND class_star>0.9
    AND transient='false'
```

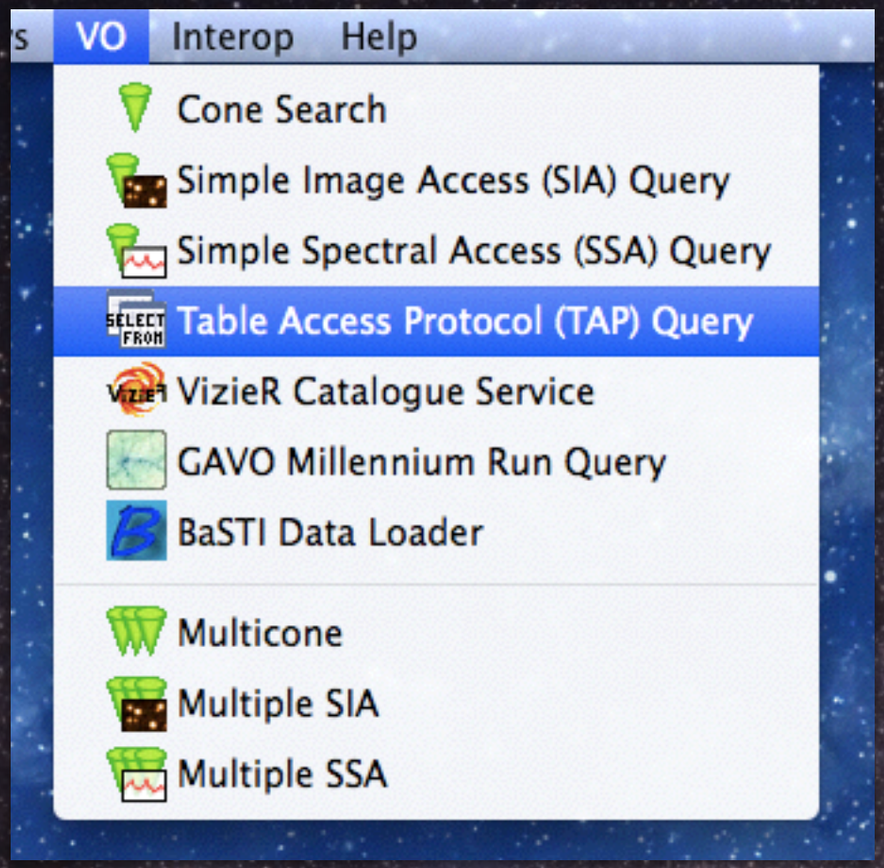
- Mathematical, logical operators, joins, sub queries,…

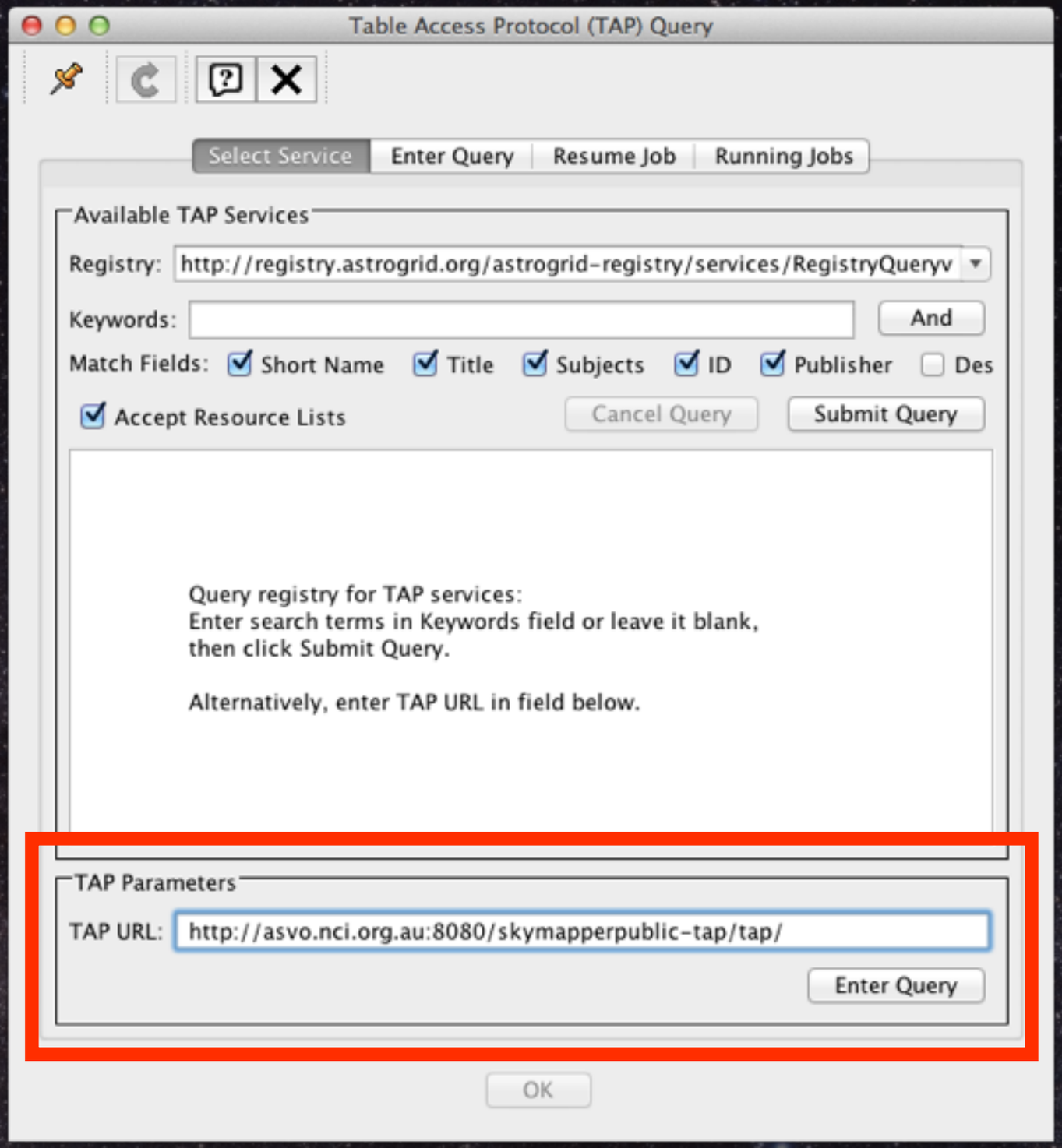
- See **is.gd/ADQLTutorial** and **is.gd/ADQLTutorial2** for a range of example queries

ALL SKY VIRTUAL OBSERVATORY

DEMO

ADQL ACCESS TO SKYMAPPER USING TOPCAT





TAP URL (in registry mid-2014):

http://asvo.nci.org.au:8080/skymapperpublic-tap/tap/

Table Access Protocol (TAP) Query

Select Service | Enter Query | Re

Table Metadata

Service: http://asvo.nci.org.au:8080/skymapper

Table: public.ms_distilled

Columns:

Name	DataType	Indexed	Unit	Desc
new_object_id	VARCHAR	☑		
mean_epoch	INTEGER	☐		
ra	DOUBLE	☐	deg	
dcl	DOUBLE	☐	deg	
pi	DOUBLE	☐		
sigma_pi	DOUBLE	☐		
mu_x	DOUBLE	☐		

Foreign Keys:

Target Table	Links	Description	Utype

Service Capabilities

Query Language: ADQL-2.0 Max Rows: Uploads: unavailable

ADQL Text

☑ Synchronous Examples Clear Parse Errors

```
SELECT *
FROM public.ms_distilled
WHERE class_star>0.9
  AND transient='true'
```

OK

TOPCAT(2): Table Browser

Table Browser for 2: TAP_1_public.ms_distille

	petro_sigma_i	petro_sigma_z	variability	transient	class_star	a	b	theta
1	0.54694	1.85437	0.3540	true	0.99	.35063	4.35618	138.0
2	0.70363	2.91083	0.438	true	0.99	.66424	13.28489	82.1
3	0.75712	3.28075	1.222	true	0.99	.63432	6.06462	73.3
4	0.54283	2.79581	0.059	true	0.99	.20918	4.18355	152.0
5	0.98691	2.07157	0.148	true	0.99	.00979	1.7948	26.7
6	0.80781	1.94733	0.185	true	0.99	.59671	2.06907	103.8
7	0.52152	1.60454	0.286	true	0.99	.66414	13.28274	50.2
8	0.45397	1.50415	0.3540	true	0.99	.99201	1.61614	134.6
9	0.4872	1.31406	0.438	true	0.99	.23055	4.61099	56.3
10	0.79725	2.69053	0.664	true	0.99	.66413	13.28257	115.2
11	0.59495	2.84734	0.075	true	0.99	.34325	6.865	142.5
12	0.51523	2.08333	0.185	true	0.99	.95232	1.77515	66.8

Cone Search

```
1=CONTAINS(POINT('ICRS',mean_ra,mean_dcl,
           CIRCLE('ICRS',0.0,0.0,0.5))
```

Table Access Protocol (TAP) Query

Table M

Service:

Table:

Columns

Name	DataType	Indexed	Unit	Description	UCD	Utype
new_object_id	VARCHAR	☑			meta.id;meta.main	
mean_epoch	INTEGER	☐				
ra	DOUBLE	☐	deg		pos.eq.ra;meta.main	
dcl	DOUBLE	☐	deg		pos.eq.dec;meta.main	
pi	DOUBLE	☐				
sigma_pi	DOUBLE	☐				
mu_x	DOUBLE	☐				

Foreign Keys:

Target Table	Links	Description	Utype

Service Capabilities

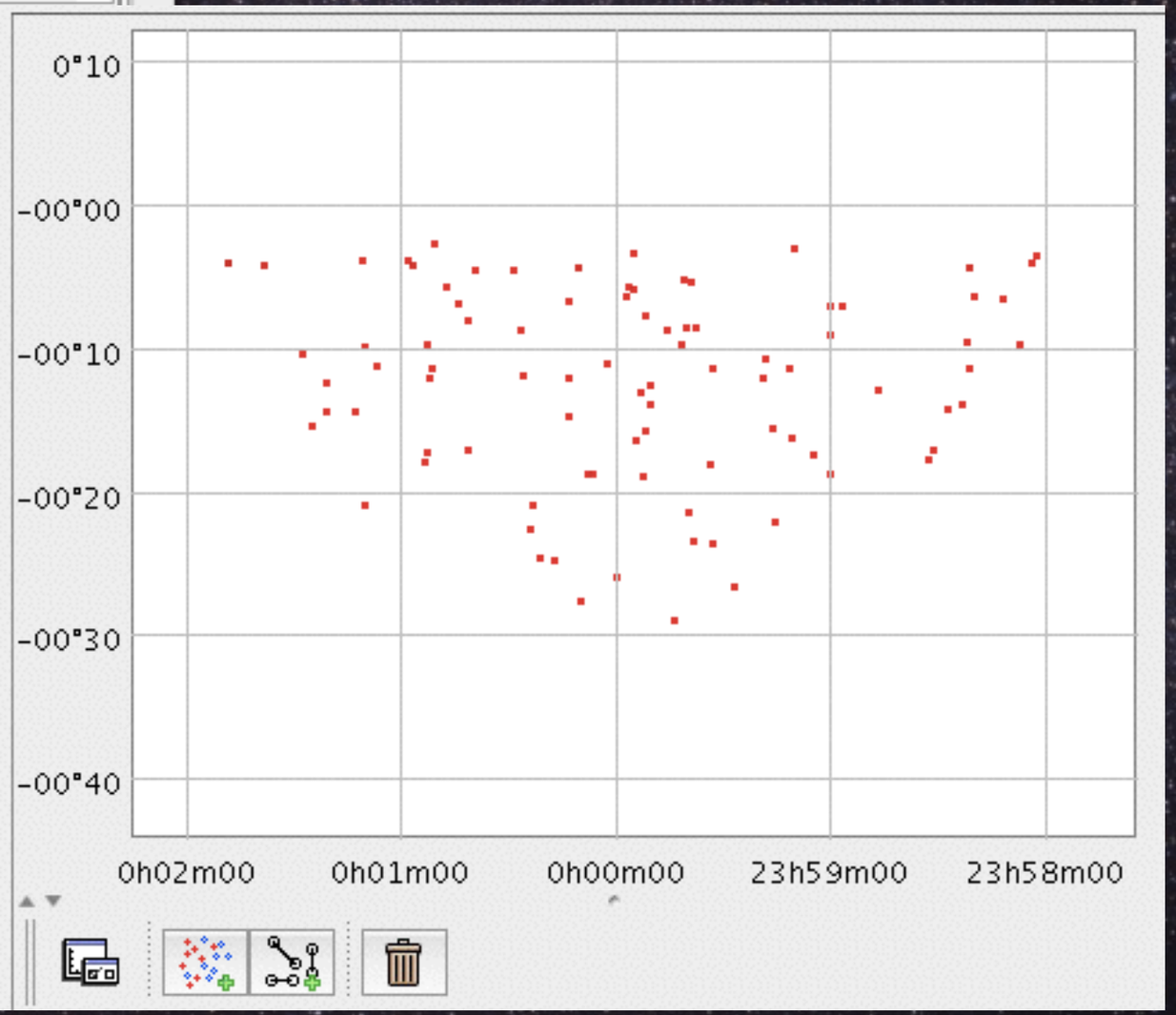
Query Language: ADQL-2.0 Max Rows: Uploads: unavailable

ADQL Text

☑ Synchronous Examples Clear Parse E

```
SELECT *
FROM public.fs_distilled
WHERE 1=CONTAINS(POINT('ICRS', mean_ra, mean_dcl), CIRCLE('ICRS', 0, 0, 0
AND sigma_flux_u/flux_u<0.1 AND n_epochs=3
```

OK



Simple Joins

```
SELECT *
FROM public.fs_photometry AS f
JOIN public.image AS i
ON f.image_id=i.image_id
```

Cross match - a join using CONTAINS

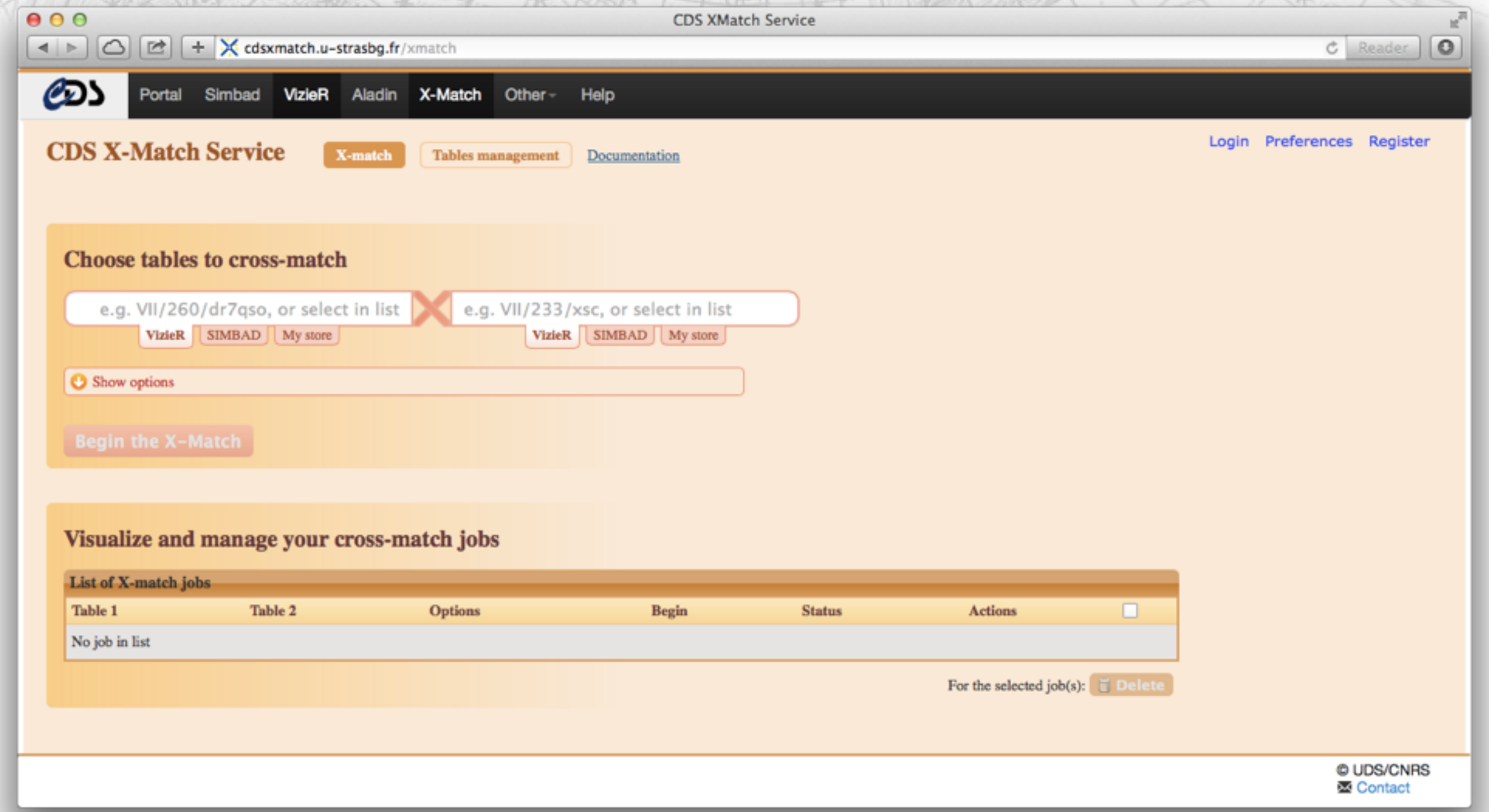
```
SELECT *
FROM public.fs_distilled AS f
JOIN public.ms_distilled AS m
ON 1=CONTAINS(POINT('ICRS',f.mean_ra,f.mean_dcl),
              CIRCLE('ICRS',m.ra,m.dcl,5./3600.))
WHERE f.n_epochs>3 AND m.a/m.b<0.5
```

FEDERATING DATA

- **Datasets more powerful when brought together**
- ASVO already hosts *HiPASS, WiggleZ, MACHO*
- Coming soon: *SkyMapper, GALAH, TAIPAN, FunnelWeb, ASKAP, …, Gaia, SKA,* **others?**
- Allow user to upload/reference their own data
- **TAP can already do much of this**
- The dream: **distributed queries**
  (coming soon to TAP standard, possible to hack now)

ASVO
ALL SKY VIRTUAL OBSERVATORY

NEW: CDS X-MATCH SERVICE



cdsxmatch.u-strasbg.fr/xmatch

ALL SKY VIRTUAL OBSERVATORY

DEMO

QUERYING THE GAVO DATA CENTRE

Table Access Protocol (TAP) Query

Select Service | Enter Query | Resume Job

Available TAP Services

Registry: http://registry.astrogrid.org/astrogrid-registry/

Keywords: gavo

Match Fields: Short Name | Title | Subjects

Accept Resource Lists

Short Name	Title	Subjects
GAVO DC TAP	GAVO Data Center TAP service	Virtual observa

TAP Parameters

TAP URL: http://dc.zah.uni-heidelberg.de/__system__/ta

OK

Table Access Protocol (TAP) Query

Select Service | Enter Query | Resume Job | Running Jobs

Table Metadata

Service: GAVO Data Center TAP service (122 ables)

Table: amanda.nucand — Dete tion parameters of neutrino candidates rec

theossa.data
twomass.data
ucac3.main
ucac3.ppmxlcross
ucac3.icrscorr
ucac4.main
usnob.plates
usnob.data

Column

Name	Description
nualpha	N utrino arrival direction, RA
nudelta	N utrino arrival direction, Declination
nch	N mber of optical modules hit in an event
angular	A gular resolution of a neutrino event
mjd	O servation time, Modified Julian Day
atmonu	S set of candidates used in AMANDA's seven-ye

Foreign Keys:

Target Table	Links	Description	Utype

Service Capabilities

Query Language: ADQL-2.0 Max Rows: 2000 (default) Uploads: Mb

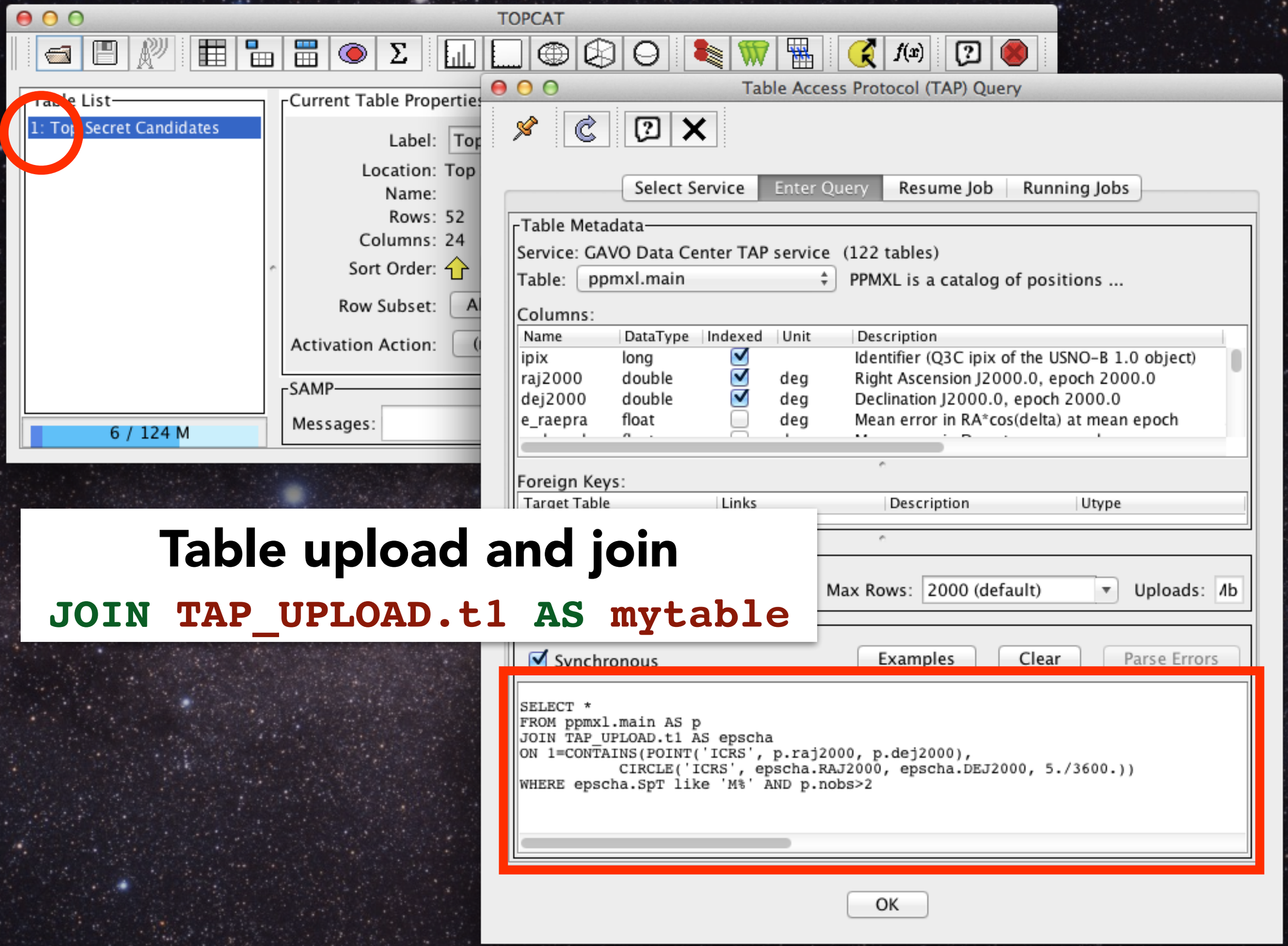
ADQL Text

Synchronous | Examples | Clear | Parse Errors

OK

Table upload and join

```
JOIN TAP_UPLOAD.t1 AS mytable
```



A slightly more complicated query…

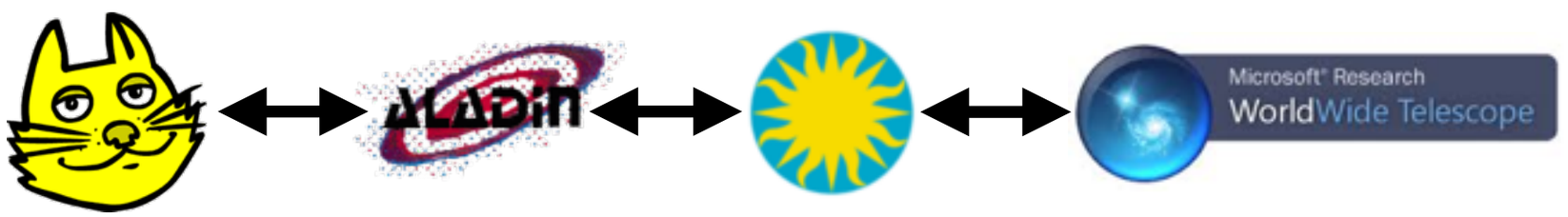
```
SELECT *
FROM (
      SELECT ALL *
      FROM wise.main AS w
      JOIN sdssdr7.sources AS s
      ON 1=CONTAINS(
            POINT('ICRS',s.ra,s.dec),
            CIRCLE('ICRS',w.raj2000,w.dej2000,20./3600.))
      JOIN twomass.data AS t
      ON 1=CONTAINS(
            POINT('ICRS',t.raj2000,t.dej2000),
            CIRCLE('ICRS',w.raj2000,w.dej2000,20./3600.))
      WHERE 1=CONTAINS(
                  POINT('ICRS',w.raj2000,w.dej2000),
                  CIRCLE('ICRS',125.0,15.0,5))
                  AND 0=CONTAINS(
                        POINT('ICRS',t.raj2000,t.dej2000),
                        CIRCLE('ICRS',w.raj2000,w.dej2000,1./3600.))
                  AND NOT EXISTS (
                        SELECT *
                        FROM supercosmos.sources AS sc
                        WHERE 1=CONTAINS(
                              POINT('ICRS',sc.raj2000,sc.dej2000),
                              CIRCLE('ICRS',w.raj2000,w.dej2000,1./3600.))))
AS bigtable
```

```
WHERE w1mag-w2mag>0.5
      AND w2mag<15.5
      AND w2mag-w3mag<2.5
      AND g>22.2
      AND r>22.2
      AND (i>21.3 OR i-z>3)
      AND z<20.5
      AND aflg='0'
      AND xflg='0'
      AND qflg NOT LIKE 'U%'
      AND jmag>9
      AND jmag-w2mag>1.8
      AND z-jmag>2.5
```

See **is.gd/ADQLTutorial2** for more details

INTEROPERABILITY

- Tools can exchange data using **SAMP (Simple Application Messaging Protocol):**
  - tables, row selections, images, sky positions, …
- **Wide support:**
  - Desktop tools (TOPCAT, Aladin, ds9, WWT, SPLAT…)
  - **Languages (Python, Java[script], C,…)**
  - Web pages (VizieR, MAST, Web Profile)



ALL SKY VIRTUAL OBSERVATORY

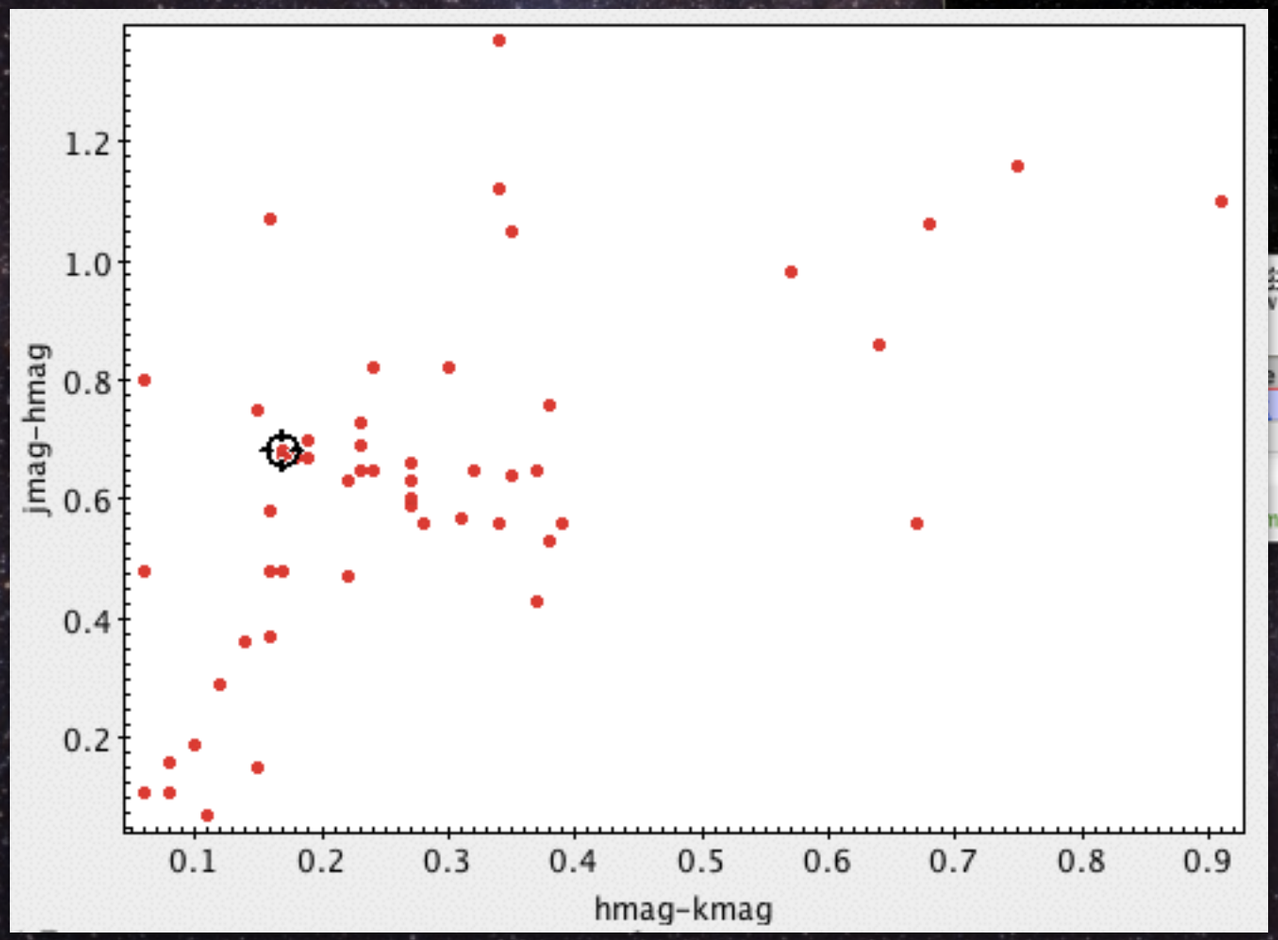
DEMO

INTEROPERABILITY WITH SAMP

TOPCAT(1): Table Brow

Table Browser for 1: Top Secret Candidates

	_RAJ2000	_DEJ2000	recno	ID	Mm	Na
1	142.08792	-78.25972	1			HI
2	164.45542	-69.23333	2		C	CF
3	167.00625	-77.70806	3			VV
4	169.12083	-78.42244	4		p	TY
5	169.64875	-79.59856	5	13	C	2M
6	170.73167	-79.41217	6	14		RX
7	171.32542	-84.95444	7			HI
8	173.45542	-76.31108	8	15		2M
9	174.38042	-76.79972	9			RX
10	175.20708	-74.99428	10	16	C	2M
11	175.36542	-73.78417	11			TY
12	175.86125	-78.07928	12	17	C	2M



Aladin v8.0

Location

Frame ICRS

DSS SDSS 2MASS GALEX PLANCK XMM Fermi Simbad NED +

DSS colored

13.43° x 12.48°

select cont pan pixel dist prop phot del draw tag filter cross x-y rgb assoc crop

Top Secret Candi

DSS colored

epoch size opac. zoom

Frame: ICRS

+90 +180 -180 -90

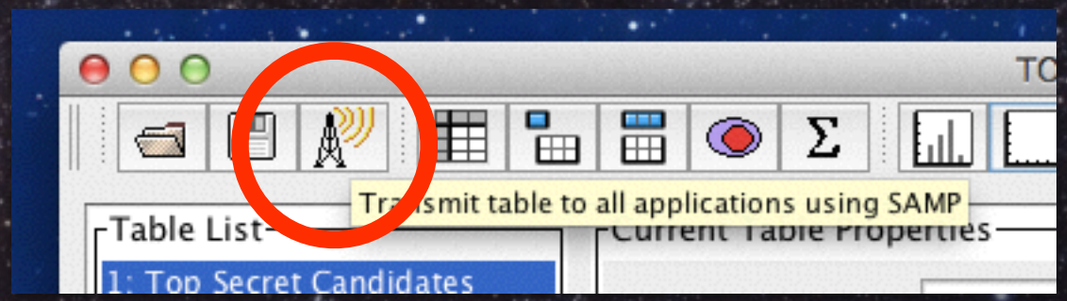
11:21:24.97 -80:10:10.4

13.43° x 12.48°

Search

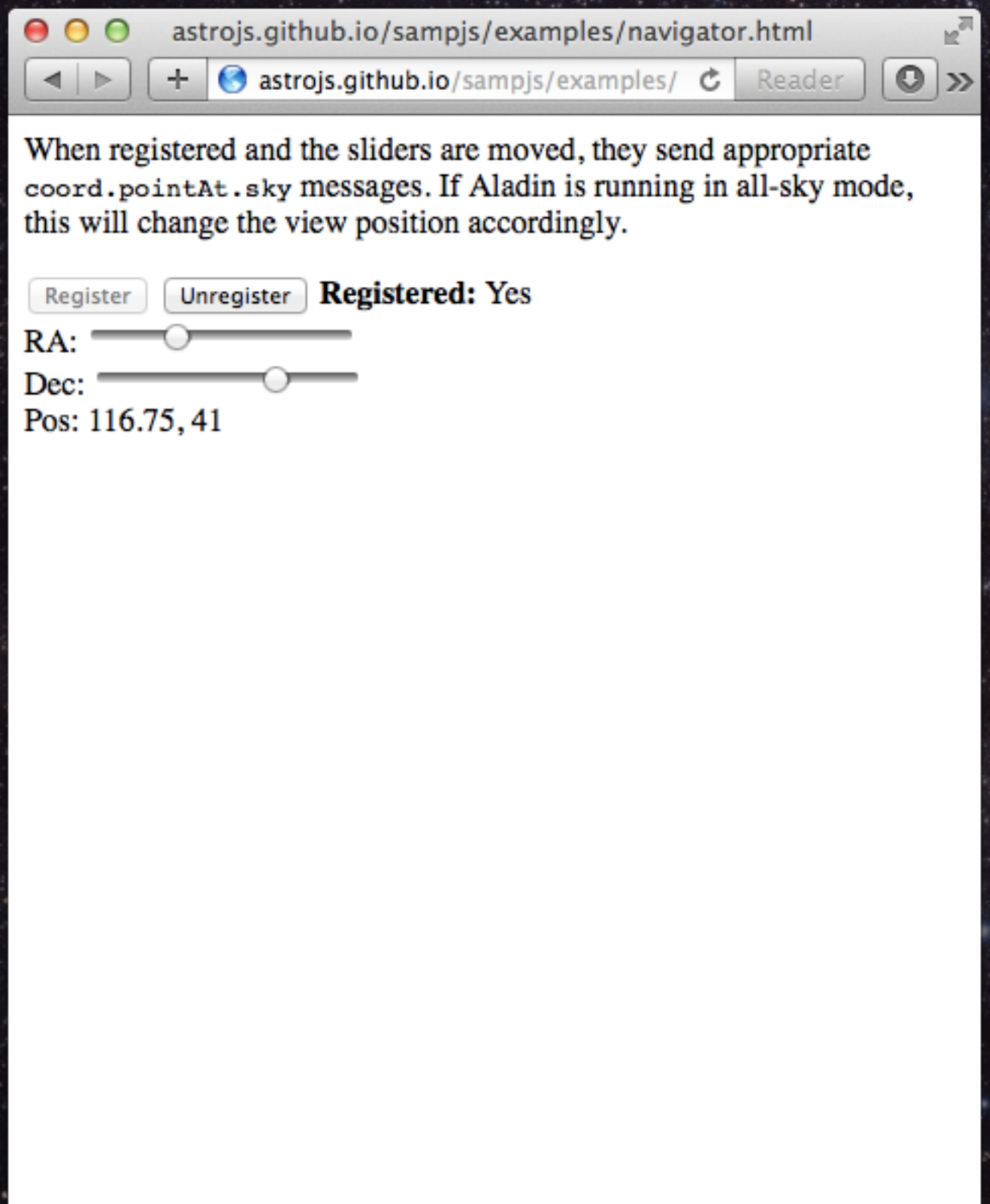
RAJ2000	DEJ2000	SpT	E(B...	HRV	E...	EWHa	pmRA	pmDE	Imag	Jmag
170.731667	-79.412167	M1.5	0.0	2.7	130	-2.0	-30.6	-16.8	11.62	10.52

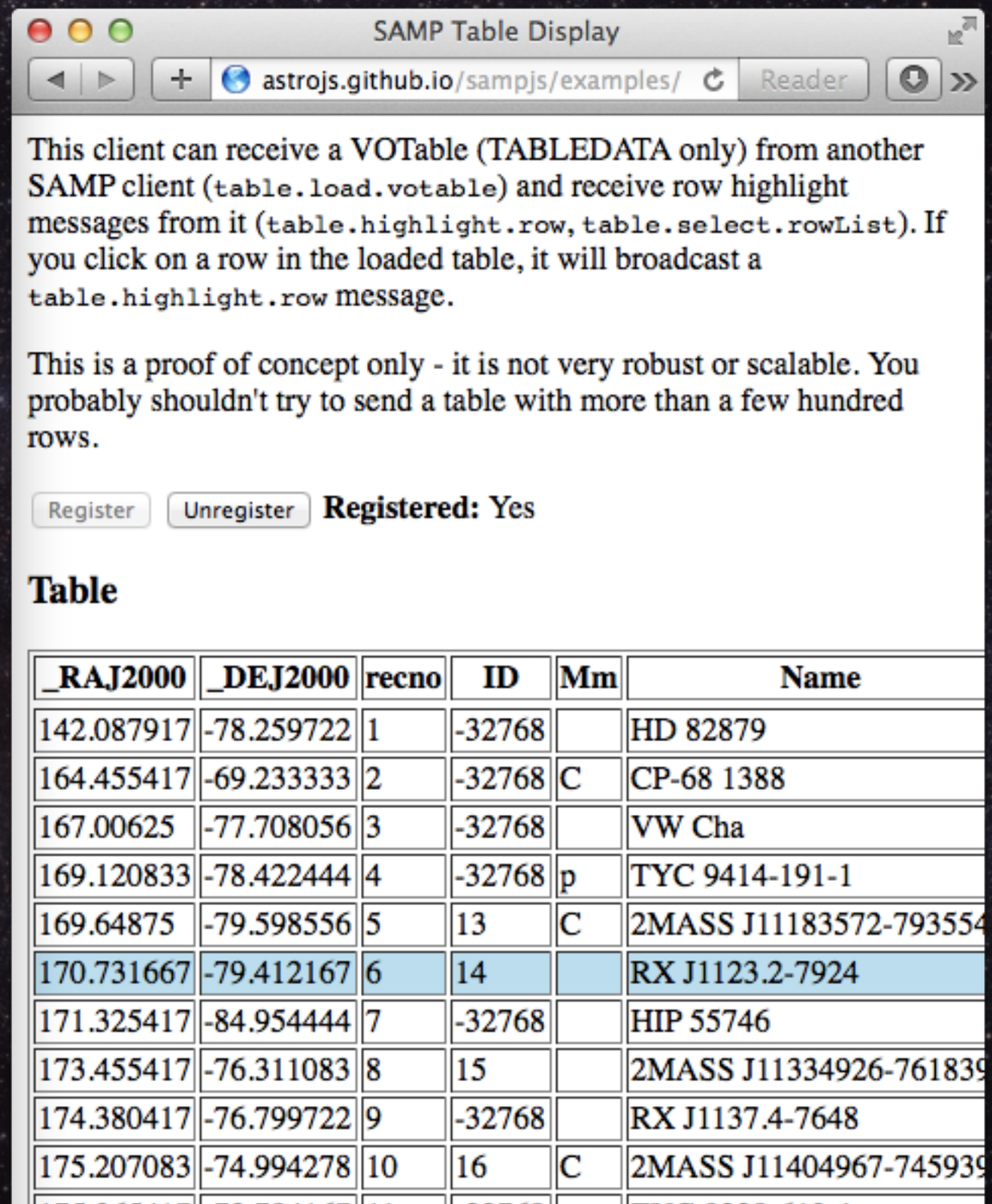
1 sel / 52 src 83fps / 87Mb



sampjs

http://astrojs.github.io/sampjs/





SUMMARY

- **Working knowledge of the VO vital** in the age of tera/peta/exabyte surveys
- Variety of ways to access SkyMapper data in ASVO:
  - **Desktop** (web, VO clients)
  - **Programmatically** (share your workflows!)
  - *Later:* **In the cloud** (custom, BYO software)
- ASVO SkyMapper Node live in **mid-2014**
  (ready for DR1 in 2015)

ASVO
ALL SKY VIRTUAL OBSERVATORY

SUMMARY

- **More information:**

  All Sky Virtual Observatory: **www.asvo.org.au**

  International VO Alliance: **www.ivoa.net**

  TOPCAT: **www.star.bris.ac.uk/~mbt/topcat/**

  ADQL Tutorials: **is.gd/ADQLTutorial(2)**

Questions, comments, suggestions?