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
• 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!
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
THE VIRTUAL OBSERVATORY

LEVEL 2

USER LAYER

Browser Based Apps
Desktop Apps
Script Based Apps

USING

VO Query Languages
VO CORE

Semantics
Formats
Resource Metadata

SAMP
SSO
CDP
WS BP
SIAP
SCS
SSAP
TAP
SLAP
SEAP

VOResource
VODataService
ApplicationRegExt
StandardRegExt
SimpleDALRegExt

ADQL
PQL
UCD
Vocabulary

CharDM
SpectrumDM
ObsCoreDM
ObsProvDM
SSLDM
SimDM
SimDAL

STC
Unit

VOEvent
VOQuery

Utypes
Units

VOSTable

Resource Identifier

Resource Metadata

Storage

Data and Metadata Collection

Computation

CDP

WS BP

SAMP

SSO

USERS

COMPUTERS

LEVEL 2

Mark Taylor, Exploring Gaia data with TOPCAT and the VO, GREAT-ESF GaiaBDs, Torino, 26 March 2014
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
Initial ASVO datasets chosen in consultation with AAL’s eResearch Advisory Committee and the community:


Future possibilities:
• More functionality for TAO and SkyMapper Nodes
• Data access services for other Australian datasets, e.g. next-gen radio telescopes
How is this relevant to science with SkyMapper?

- VO tries to make science 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
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 $\sim 10^2$ cols $\rightarrow$ **50-100 TB database**
  
  c.f. SDSS-DR7: 18 TB ($4\times10^8$ 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)
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)
WEB ACCESS: CONE SEARCH

The radial search will return all objects in the SkyMapper catalogue that fall within the search radius of the specified right ascension and declination. All values are in decimal degrees.

The system currently supports a maximum search radius of 10°.

The web interface is limited to displaying the first 1000 results of a query. There is a larger upper limit for results downloaded via the TAP service.

SkyMapper survey:
Main Survey

Right ascension (deg):
12:34:56.78

Declination (deg):
-01:23:45.67

Search radius (deg):
2.0

U min:

U max:
WEB ACCESS: IMAGES

Search

Radial  Rectangular  Raw Image  Bulk Catalogue  Bulk Image

The image search will return SkyMapper images centered on the specified right ascension and declination. All values are in decimal degrees.

Right ascension (deg):  Declination (deg):

Search SkyMapper

Useful Links
Astronomy Australia Ltd.
ANU
MSO
NCI
NeCTAR

Developed by Intersect 2013.
WEB ACCESS: CROSS MATCH

The bulk catalogue search will return all objects in the SkyMapper catalogue that fall within the search radius of the specified right ascension and declination points found in the uploaded csv file (example). All values are in decimal degrees.

SkyMapper survey:

File (csv):

Search radius (deg):

Download format:

Search SkyMapper

Asynchronous Query
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
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**
ADQL IN ONE SLIDE

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

```sql
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
DEMO

ADQL ACCESS TO SKYMAPPER USING TOPCAT
TAP URL (in registry mid-2014):
http://asvo.nci.org.au:8080/skymapperpublic-tap/tap/
SELECT * FROM public.ms_distilled WHERE class_star > 0.9 AND transient = 'true'
Cone Search

\[ 1 = \text{CONTAINS}(\text{POINT('ICRS', mean\_ra, mean\_dcl), CIRCLE('ICRS', 0.0, 0.0, 0.0, 0.5))} \]
Simple Joins

```sql
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

```sql
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)
NEW: CDS X-MATCH SERVICE

cdsxmatch.u-strasbg.fr/xmatch
DEMO
QUERYING THE GAVO DATA CENTRE
Table upload and join

JOIN TAP_UPLOAD.t1 AS mytable

SELECT * 
FROM ppmxl.main AS p 
JOIN TAP_UPLOAD.t1 AS epscha 
ON 1=CONTAINS(POINT('ICRS', p.raj2000, p.dej2000), 
CIRCLE('ICRS', epscha.RAJ2000, epscha.DEJ2000, 5./3600.)) 
WHERE epscha.SpT like 'M%' AND p.nobs>2
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](http://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)
DEMO

INTEROPERABILITY WITH SAMP
When registered and the sliders are moved, they send appropriate `coord.pointAt.sky` messages. If Aladin is running in all-sky mode, this will change the view position accordingly.

```
RA:  Unregister  Register: Yes
Dec:  
Pos: 116.75, 41
```

This client can receive a VOTable (TABLEDATA only) from another SAMP client (`table.load.votable`) and receive row highlight messages from it (`table.highlight.row, table.select.rowList`). If you click on a row in the loaded table, it will broadcast a `table.highlight.row` message.

This is a proof of concept only - it is not very robust or scalable. You probably shouldn't try to send a table with more than a few hundred rows.

```
Table

<table>
<thead>
<tr>
<th>RAJ2000</th>
<th>DEJ2000</th>
<th>recono</th>
<th>ID</th>
<th>Mm</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>42.087917</td>
<td>-78.259722</td>
<td>1</td>
<td>-32768</td>
<td></td>
<td>HD 82879</td>
</tr>
<tr>
<td>164.455417</td>
<td>-69.233333</td>
<td>2</td>
<td>-32768</td>
<td>C</td>
<td>CP-68 1388</td>
</tr>
<tr>
<td>167.00625</td>
<td>-77.708056</td>
<td>3</td>
<td>-32768</td>
<td></td>
<td>VW Cha</td>
</tr>
<tr>
<td>169.120833</td>
<td>-78.422444</td>
<td>4</td>
<td>-32768</td>
<td>p</td>
<td>TYC 9414-191-1</td>
</tr>
<tr>
<td>169.64875</td>
<td>-79.598556</td>
<td>5</td>
<td>13</td>
<td>C</td>
<td>2MASS J11183572-793554</td>
</tr>
<tr>
<td>170.731667</td>
<td>-79.412167</td>
<td>6</td>
<td>14</td>
<td></td>
<td>RX J1123.2-7924</td>
</tr>
<tr>
<td>171.325417</td>
<td>-84.954444</td>
<td>7</td>
<td>-32768</td>
<td></td>
<td>HIP 55746</td>
</tr>
<tr>
<td>173.455417</td>
<td>-76.311083</td>
<td>8</td>
<td>15</td>
<td></td>
<td>2MASS J11334926-761839</td>
</tr>
<tr>
<td>174.380417</td>
<td>-76.799722</td>
<td>9</td>
<td>-32768</td>
<td></td>
<td>RX J1137.4-7648</td>
</tr>
<tr>
<td>175.207083</td>
<td>-74.994278</td>
<td>10</td>
<td>16</td>
<td>C</td>
<td>2MASS J11404967-745939</td>
</tr>
<tr>
<td>175.365417</td>
<td>-73.874167</td>
<td>11</td>
<td>-32768</td>
<td></td>
<td>TYC 9238-612-1</td>
</tr>
</tbody>
</table>
```
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)
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?