

Exploring Gaia data with TOPCAT and the Virtual Observatory

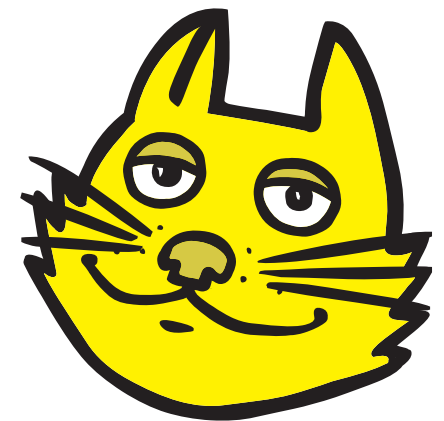
Mark Taylor (University of Bristol)

Gaia and the Unseen Brown Dwarf Question

GREAT-ESF Workshop

Torino University

26 March 2014



\$Id: tcvo.tex,v 1.23 2014/03/25 19:33:53 mbt Exp \$

Outline

Before coffee (9:00–10:00): Introduction (Mark Taylor)

- Gaia catalogue
- Virtual Observatory
- TOPCAT
 - ▷ Capabilities
 - ▷ Visualisation
 - ▷ Crossmatching
- SAMP
- STILTS
- Demo

After coffee (10:30–12:30)

- TOPCAT/VO hands-on ([White Room](#)):
 - ▷ Exercise: TOPCAT visualisation and crossmatching (Mark Taylor)
 - ▷ Exercise: TAP/ADQL (Simon Murphy)
- Brown Dwarf Ages ([Aula Magna](#)):
 - ▷ *Jackie Faherty et al.*

Catalogue in DPAC

Gaia DPAC

- CU1: System Architecture
- CU2: Simulations
- CU3: Core Processing
- CU4: Object Processing
- CU5: Photometric Processing
- CU6: Spectroscopic Processing
- CU7: Variability Processing
- CU8: Astrophysical Parameters
- CU9: Archive & Catalogue

CU9: Archive and Catalogue

- CU9 Work Packages:
 - 910: Management
 - 920: Documentation
 - 930: Archive architecture
 - 940: Validation
 - 950: Operations
 - 960: Education and Outreach
 - 970: Science-enabling Applications
 - 980: Visualisation
- CU9 Participation:
 - 52 institutes, 130 people
- CU9 Funding
 - National funding bodies
 - Institutions
 - GENIUS (EU FP7)

Gaia Data Access

Data Services (a selection):

- GACS: Gaia Archive Core Systems (ESAC)
 - ▷ Primary access to Gaia catalogue
 - ▷ TAP (TAP+?) service
 - ▷ Persistent user-uploaded tables
 - ▷ Indexes to other large surveys
 - ▷ Hadoop cluster for advanced operations (near-data map/reduce)
- GAVO TAP service (ARI Heidelberg)
 - ▷ Selection of other large and small datasets in same database
- WFAU TAP service (Edinburgh)
 - ▷ Federated TAP services
- VizieR (CDS)
 - ▷ TAPVizieR
 - ▷ VizieR web page access and CDS portal
 - ▷ CDS XMatch service

Gaia Catalogue Release Scenario

Provisional release *scenario* (*schedule* later):

Now: GUMS-10 Simulation (MW, LMC, SMC, GAL, QSO, SN)

L+22 months: Single-star α , δ , G magnitudes; Hundred Thousand Proper Motions

L+28 months: Single-star 5-param solutions; Integrated BP/RP photometry; Mean radial vels

L+40 months: Orbital solutions for some binaries; Object classifications; BP/RP/RVS spectra

L+65 months: Variable star classifications; Epoch photometry; Solar system; Non-single stars

EOM+3 years (Final Release): Full astro, photo and, RV catalogues; All variable and non-single star solutions; Exo-planets; Epoch and transit data; Ground-based observations

More detail:

- <http://www.cosmos.esa.int/web/gaia/release>
- GAIA-CG-PL-ESA-TJP-011-01

Virtual Observatory

What is the Virtual Observatory (VO)?

- *“All astro archives in your computer”*
- A set of protocols that allows software clients to talk to external data services in a uniform way

Why is it relevant to getting (BD) science from Gaia data?

1. Gaia catalogue will be made available using VO protocols (TAP)
2. *Other* data services are available using VO protocols (surveys, images, spectra, followup observations, ...)
⇒ multi-wavelength, multi-dataset science

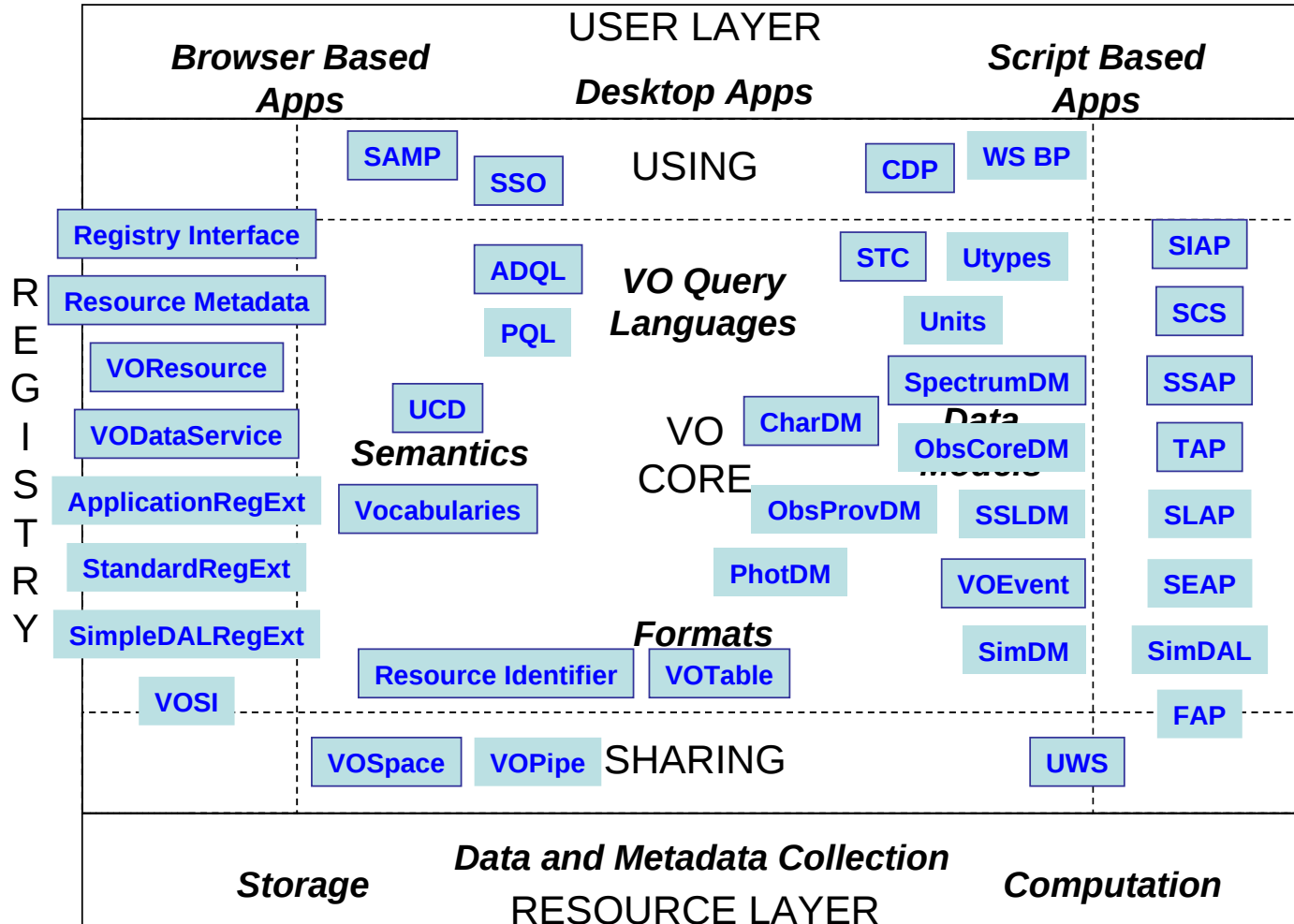
Virtual Observatory Architecture

LEVEL 2

USERS



COMPUTERS

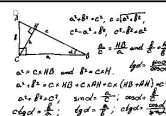


DATA PROTOCOLS

20101004
IVOA Architecture



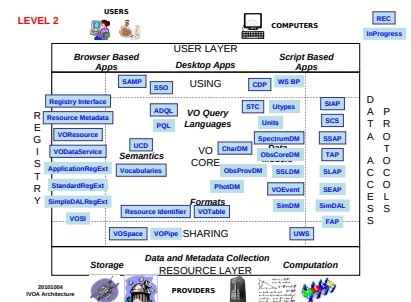
PROVIDERS



Virtual Observatory Protocols

Most important VO protocols:

- Directory service:
 - ▷ **Registry** — Locate data archives/services
- Positional data access services (by RA/Dec + radius):
 - ▷ **Simple Cone Search** — Rows from catalogue
 - ▷ **Simple Image Access (SIA)** — Images from archive
 - ▷ **Simple Spectral Access (SSA)** — Spectra from archive
- General data access service:
 - ▷ **Table Access Protocol (TAP)** — SQL-like query of database
- Application communications:
 - ▷ **Simple Application Messaging Protocol (SAMP)** — exchange data/control between local applications



Not quite the VO

- Vizier
- CDS Xmatch service

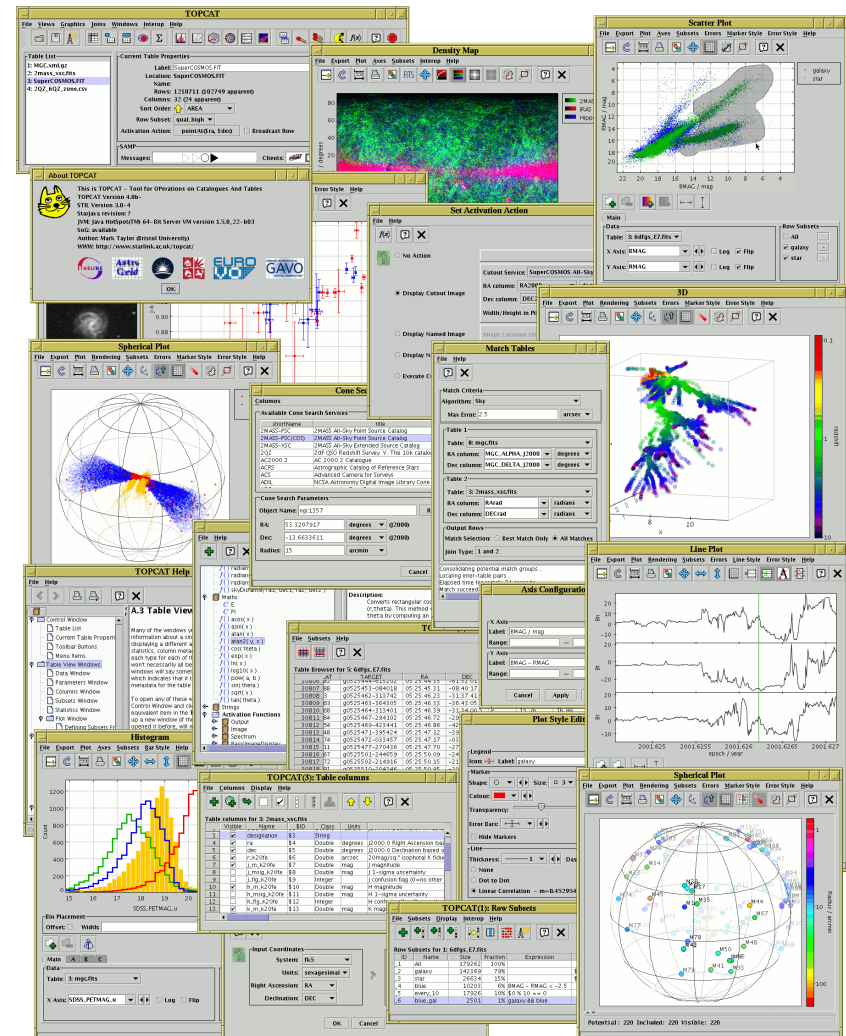
TOPCAT Overview

TOPCAT = Tool for OPerations on Catalogues And Tables

"Does what you want with tables"

Aims:

- Easy to use
- Easy to learn
- Simple things obvious, complicated things well-documented
- Easy to investigate data — good for interactive exploration
- Easy to install and run (pure Java — one download file, no library issues)
- Fast
- Handle (fairly) large tables (10⁶ rows × 10² cols easily, maybe more)
- User-driven development



TOPCAT Capabilities

It can do:

- Read/write tables in multiple formats
- View/edit data
- View/edit metadata
- Calculations and statistics
- Visualisation
- Make/combine/display row selections in various ways (linked views)
- Crossmatching — efficient and very flexible
- Access external data services (VO and others)
- Talk to other astro tools (SAMP)



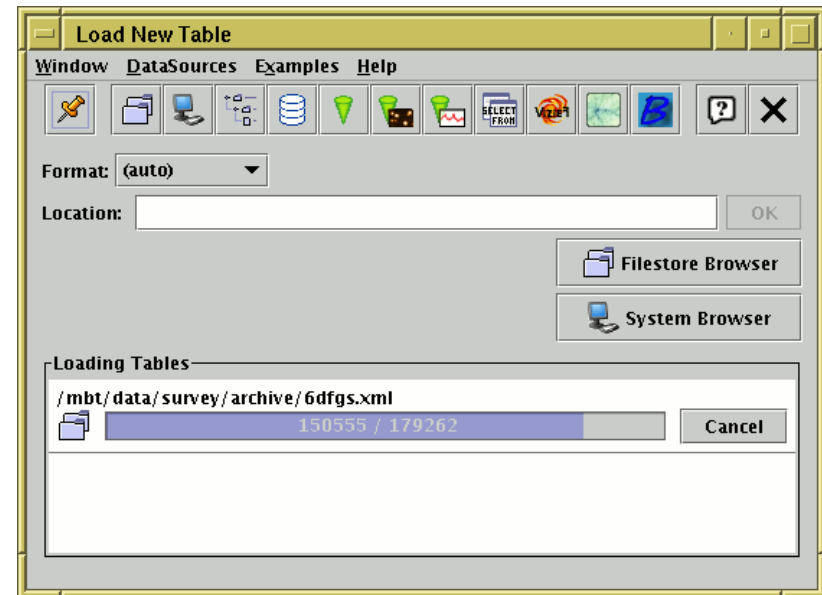
TOPCAT: Input/Output

- Table format support:

- Table file formats:
 - ▷ FITS binary and FITS ASCII tables
 - ▷ ASCII (*but not all ASCII*)
 - ▷ CSV
 - ▷ VOTable
 - ▷ L^AT_EX (*output only*)
 - ▷ HTML (*output only*)
 - ▷ a few others; extensible
- ... making conversion between any of these trivial

- Input sources:

- Local disk (custom, system or tree browser)
- Virtual Observatory services (Cone, TAP, SIA, SSA)
- SAMP
- VizieR
- Millennium Simulation
- SQL database
- ... a few others; extensible



TOPCAT: Table Data and Metadata

TOPCAT(1): Table Parameters

Window Parameters Display Help

Table Parameters for 1: 6dfgs_mini.xml.bz2

Name	Value	Units	UCD	Description
Name	6dfgs_E7_subset			Table name
URL	jar:file:/data/andromeda1/starjava/java/li...			URL of original table
Column Count	17			Number of columns
Row Count	875			Number of rows
Description	6dFGS master config file (version E7 March...			
Original Source	http://www-wfau.roe.ac.uk/6dFGS/6dfgs...			URL of data file used to
Credits	Column explanations provided by Mike Re...			
Conversion	Converted from 6dfgs_E7.fld.gz by Mark T...			
RESOLUTION	15	arcsec	stat.error;pos.eq.ra	Nominal positional error

Name: Description
 Class: String
 Shape:
 Units:
 Description:
 UCD:
 Value: 6dFGS master config file (version E7 March 2004) - DEMO SUBSET. These data are taken from the 6dF Galaxy Redshift Survey Database, see astro-ph/0505068. Kindly provided by Mike Read, ROE. These data are for EXAMPLE PURPOSES ONLY, intended for demonstrations of some of TOPCAT's properties. For science use, please consult the

Table Metadata view

TOPCAT(1): Table Browser

Window Subsets Help

Table Browser for 1: dr5qso.fits

	SDSSName	RA	DEC	z	psfmag_u	psfmagerr
21412	092322.64+020135.5	140.84436	2.02655	0.3831	20.039	0.044
21413	092322.67+282526.5	140.84449	28.42405	0.3183	18.928	0.022
21414	092322.86+033821.5	140.84526	3.63933	3.006	21.529	0.127
21415	092323.01+461835.3	140.84588	46.30982	1.608	19.241	0.035
21416	092323.65+580256.0	140.84855	58.0489	0.7481	19.289	0.025
21417	092323.92+610154.0	140.84969	61.03167	1.5332	19.536	0.039
21418	092324.25+382812.8	140.85104	38.47024	0.788	19.141	0.026
21419	092324.47+533005.4	140.85197	53.50152	0.8781	19.136	0.038
21420	092324.49+034901.7	140.85207	3.81716	0.8634	18.8	0.03
21421	092325.25+453222.1	140.85521	45.5395	3.4523	20.338	0.053
21422	092326.45+254023.6	140.86021	25.67324	1.2275	19.276	0.025
21423	092326.53+264223.3	140.86055	26.7065	0.7604	20.642	0.074
21424	092326.86+543824.7	140.86192	54.64021	0.4774	18.825	0.026
21425	092326.88+294641.1	140.86201	29.77800	1.8742	20.002	0.065

Table data view

TOPCAT(3): Table columns

Window Columns Display Help

Table columns for 3: 2mass_xscfits

Visible	Name	ID	Class	Units	
<input checked="" type="checkbox"/>	designation	\$3	String		
<input checked="" type="checkbox"/>	ra	\$4	Double	degrees	J2000.0 Right Ascension base
<input checked="" type="checkbox"/>	dec	\$5	Double	degrees	J2000.0 Declination based o
<input checked="" type="checkbox"/>	r_k20fe	\$6	Double	arcsec	20mag/sq." isophotal K fidu
<input checked="" type="checkbox"/>	j_m_k20fe	\$7	Double	mag	J magnitude
<input type="checkbox"/>	j_msig_k20fe	\$8	Double	mag	J 1-sigma uncertainty
<input type="checkbox"/>	j_flg_k20fe	\$9	Integer		J confusion flag (0=no other
<input checked="" type="checkbox"/>	h_m_k20fe	\$10	Double	mag	H magnitude
<input type="checkbox"/>	h_msig_k20fe	\$11	Double	mag	H 1-sigma uncertainty
<input type="checkbox"/>	h_flg_k20fe	\$12	Integer		H confusion flag (0=no other
<input checked="" type="checkbox"/>	k_m_k20fe	\$13	Double	mag	K magnitude

Column Metadata view

Row Selections

Different ways to make single or multiple row selections:

- Select point(s) graphically from a plot
- Select row(s) from the table view
- Use an algebraic expression
- Combine existing subsets
- Receive from an external application (SAMP)

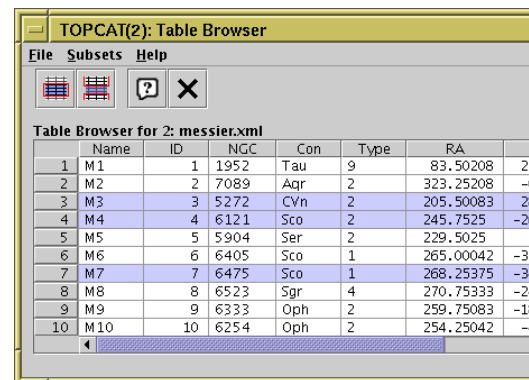
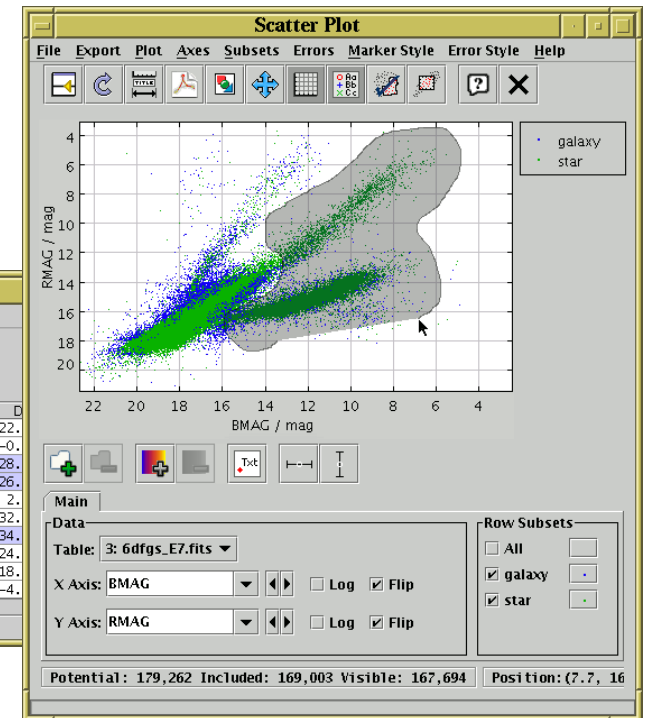


Table Browser for 2: messier.xml

	Name	ID	NGC	Con	Type	RA	D
1	M1	1	1952	Tau	9	83.50208	22.
2	M2	2	7089	Aqr	2	323.25208	-0.
3	M3	3	5272	CVn	2	205.50083	28.
4	M4	4	6121	Sco	2	245.7525	-26.
5	M5	5	5904	Ser	2	229.5025	2.
6	M6	6	6405	Sco	1	265.00042	-32.
7	M7	7	6475	Sco	1	268.25375	-34.
8	M8	8	6523	Sgr	4	270.75333	-24.
9	M9	9	6333	Oph	2	259.75083	-18.
10	M10	10	6254	Oph	2	254.25042	-4.



Linked views mean a selection made one way is visible in other ways

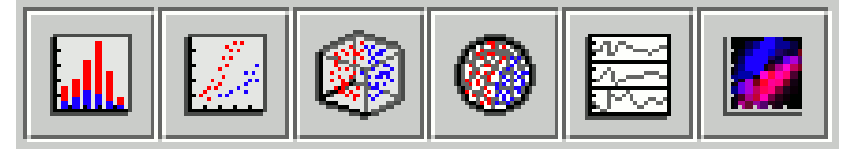
- Perform crossmatch only on items in red giant branch
- Where on the sky is this colour cut?
- Spot outliers
- Identify objects on ds9 image display

Calculations

- Expression language used for creating columns, defining selections, specifying axes etc:
 - Straightforward arithmetic syntax (C-like)
 - Use column names like variables
 - Standard arithmetic operators (+, -, /, *)
 - Standard mathematical functions (`abs`, `max`, `round`, `sin`, `cos`, `pow`, ...)
 - Sky coordinates (degrees, sexagesimal, sky distances)
 - Cosmological distances (redshift, luminosity dist, lookback time, ...)
 - Fluxes (Johnson AB Magnitudes, Jansky)
 - Time conversions (ISO8601, MJD, Julian, Besselian)
 - ... and more (and it's extensible)
- Examples:
 - `mag_u - mag_g`
 - `janskyToAb(flux)`
 - `skyDistanceDegrees(ra, dec, 14.1, -72.9) < 1.2`

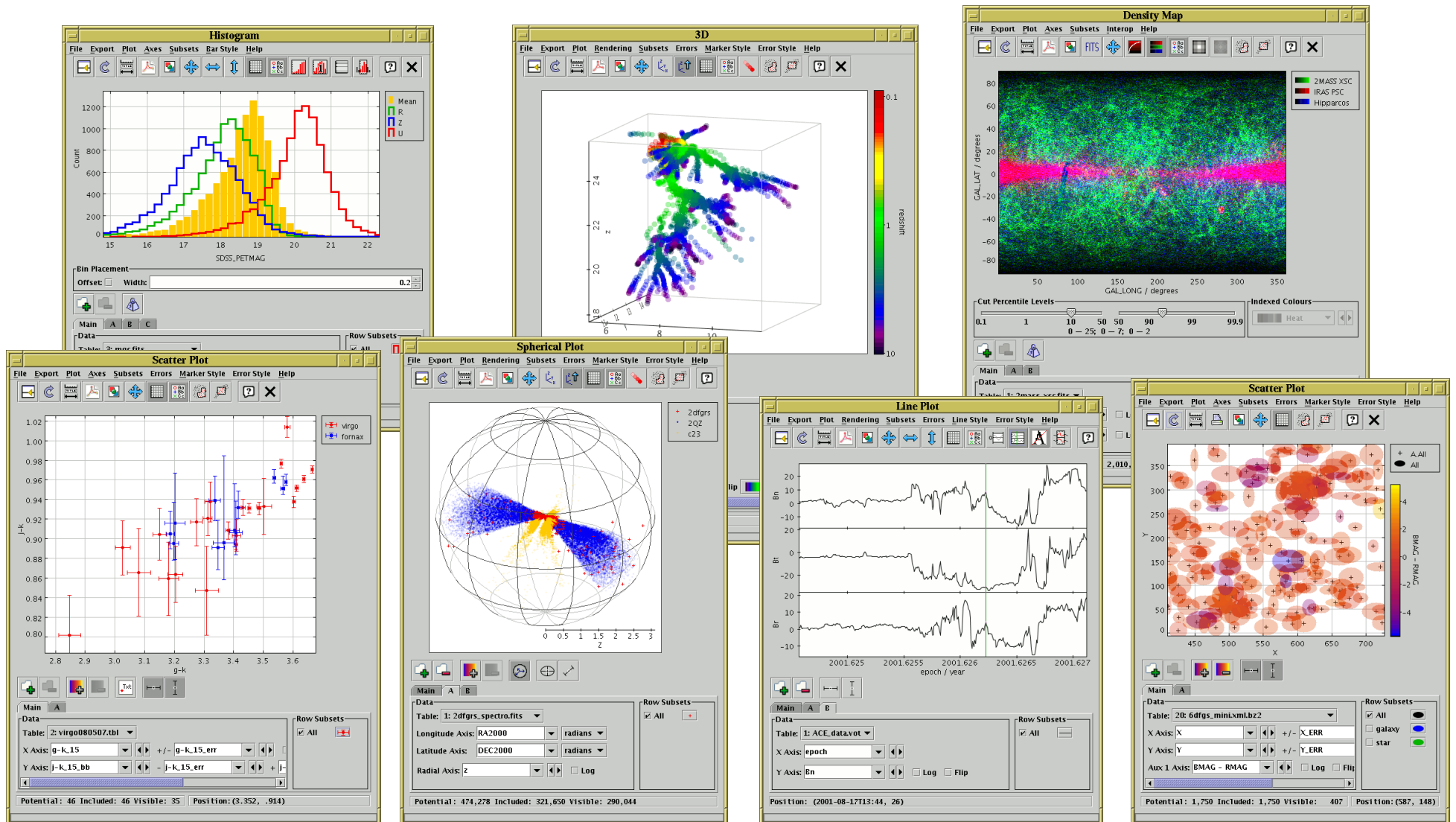
TOPCAT: v3 Visualisation

Established functions (*now slightly deprecated*)



- Histograms, 2-d and 3-d scatter plots, density map, stacked line plots
- Error bars
- Variable transparency
- Highly configurable
- Interactive
- Code points by colour
- Text labels
- Large datasets
- ...

TOPCAT: v3 Visualisation



TOPCAT v4 Visualisation

v4.0b March 2013, v4.1 March 2014

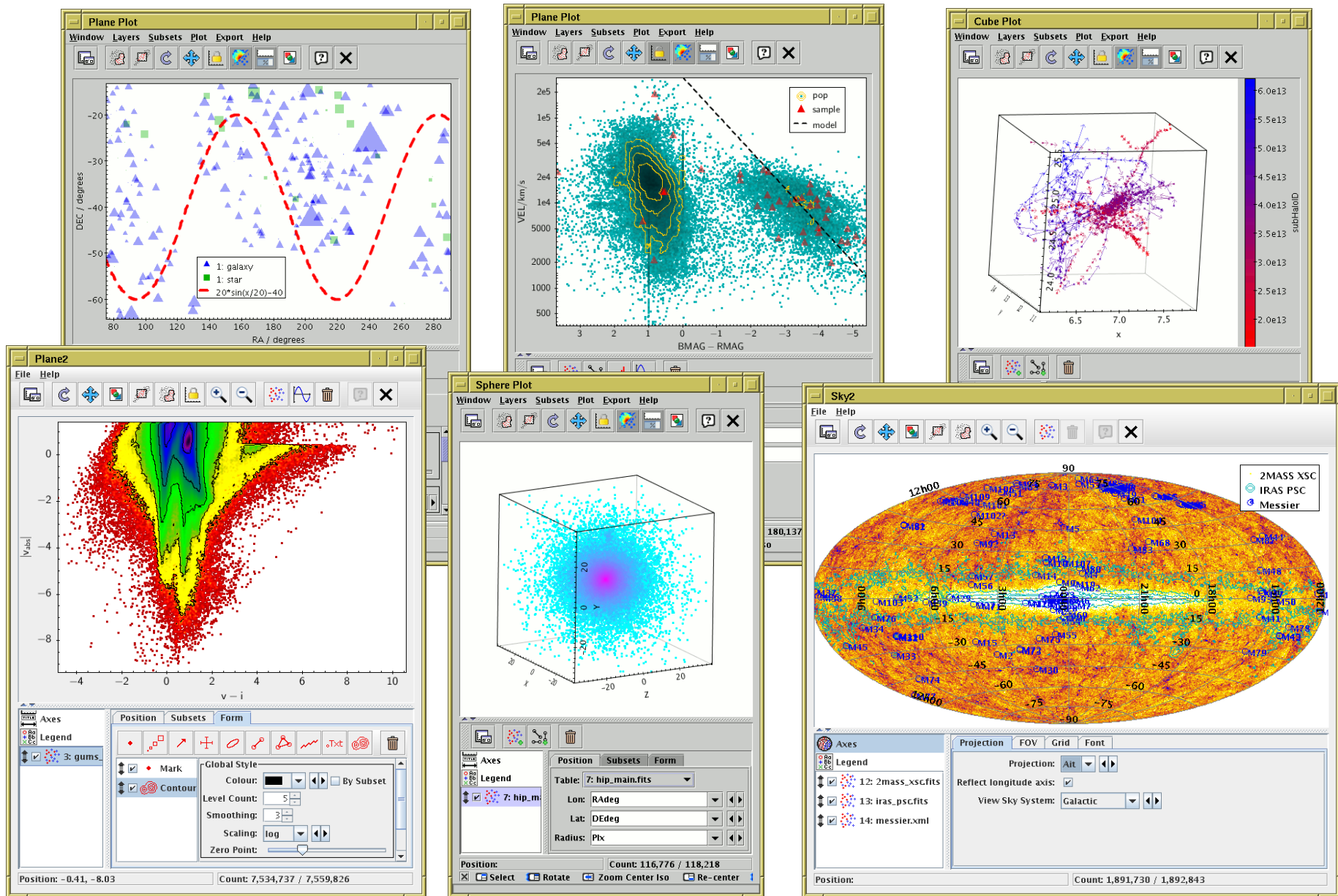
Most of the version 3 functions **plus**:



- Different user interface
- Stack different types of plots in layers over each other
- New options: vectors, density contours, code by marker size, ellipses, ...
- Pair links (crossmatch results)
- Hybrid scatter plot/density map in 2d and 3d
- Sky coordinates
- Better navigation (especially 3d)
- More responsive
- Better axis labelling, including \LaTeX
- Better support for large data sets
- Many more configuration options
- Analytic function plotting
- ...

Version 3 windows are still available (**Graphics** menu)

TOPCAT v4 Visualisation




Crossmatching

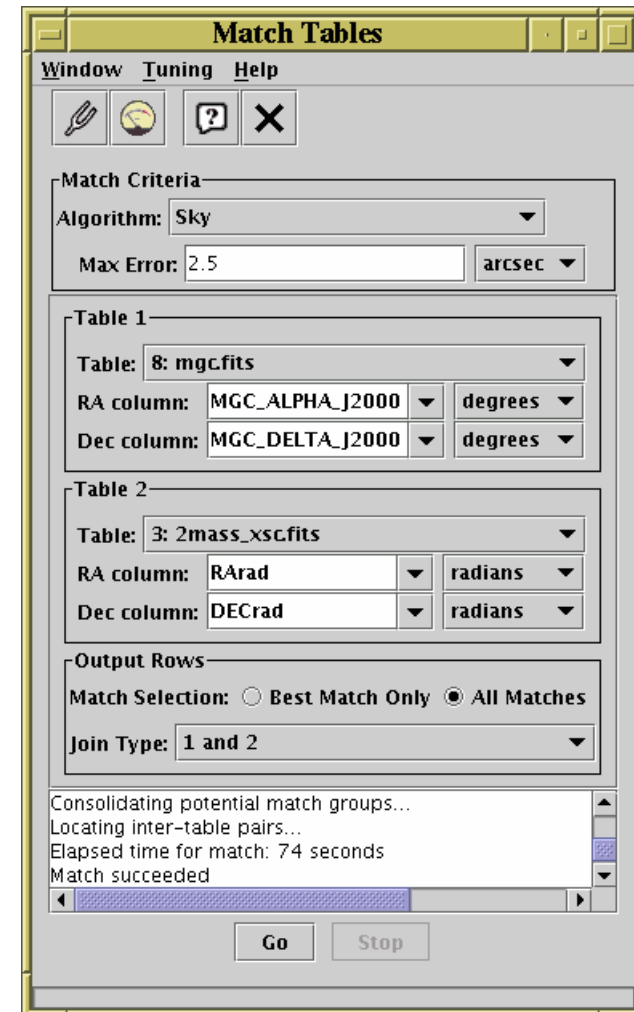
Various options for crossmatching tables

- Internal
 - ▷ Download all tables, use TOPCAT crossmatch dialogues
 - ▷ Flexible, easy, usually fast
 - ▷ Good up to a few $\times 10^6$ row
- TAP
 - ▷ TOPCAT TAP window, maybe upload local table; other TAP clients are available
 - ▷ Flexible, requires some knowledge
 - ▷ Good for large external catalogues, plus maybe medium local one
- Multi-cone
 - ▷ One cone-search query for each row of local table, use TOPCAT multi-cone window
 - ▷ Slow, inflexible, only by sky position
 - ▷ Only good for small local catalogue, large external catalogue
- CDS XMatch (<http://cdsxmatch.u-strasbg.fr/>)
 - ▷ Web form, TOPCAT interface one day
 - ▷ Very fast, not very flexible, only by sky position
 - ▷ Works with huge (10^9 row) catalogues, any from Vizier or local uploaded

TOPCAT Crossmatch Window

Crossmatch tables already loaded in TOPCAT:

- Pair match, Intra-table match, 3-, 4-, 5-table ...
- Flexible match criteria:
 - ▷ RA, Dec
 - ▷ RA, Dec, radius (or redshift, or other coord)
 - ▷ 2D or 3D (or more) Cartesian positions
 - ▷ ... with/without errors (circular, elliptical)
 - ▷ Exact (e.g. object ID)
 - ▷ Combinations of the above ...
- Retain closest only or all matches
- Output matched rows, unmatched rows, union, XOR, ...
- Efficient algorithm: usually < a minute or two
- *New in v4:*  visualise result



TOPCAT TAP Window

TAP (Table Access Protocol)

- Select TAP service (registry query)
- Browse DB metadata (table names and columns)
- Enter query in ADQL; optionally specify upload table from TOPCAT
- Service executes query, TOPCAT loads result

ADQL (Astronomical Data Query Language)

- A dialect of SQL, includes some geometry functions
- Syntax not always memorable
- [Examples](#) button is here to help!

SELECT

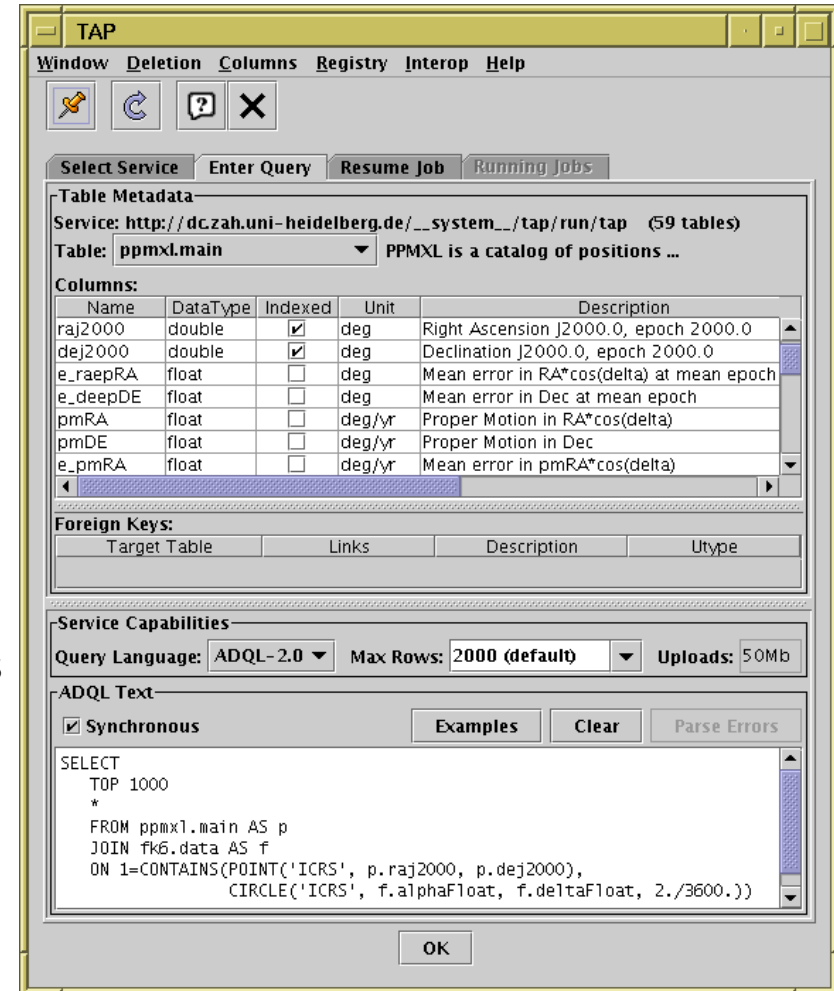
```
TOP 100000
```

```
db.ipix, db.raj2000, db.dej2000, tc.name, tc.alpha, tc.delta
```

```
FROM ppmxl.main AS db
```



```
JOIN TAP_UPLOAD.t3 AS tc
```

```
ON 1=CONTAINS(POINT('ICRS', db.raj2000, db.dej2000),  
              CIRCLE('ICRS', tc.RA2000, tc.DEC2000, 5./3600.))
```



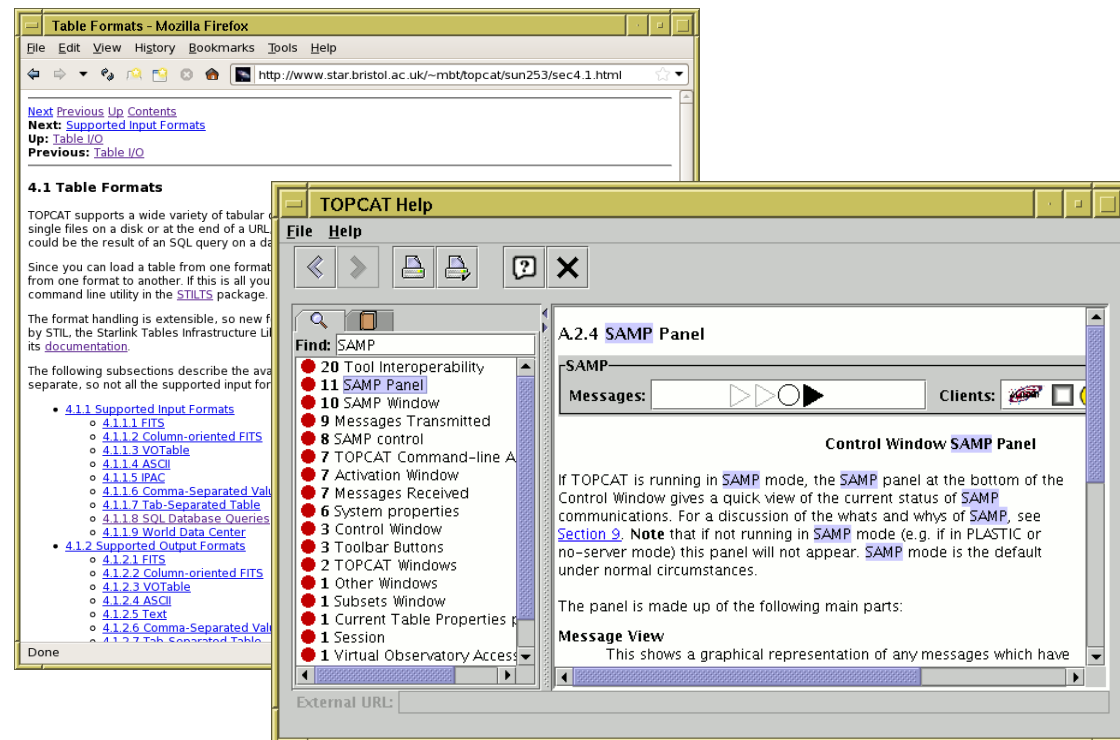
TOPCAT: Help!

- Full tutorial and reference documentation:

- ▶ HTML/PDF manual on web page <http://www.starlink.ac.uk/topcat/> (or Google it)
- ▶ **Help for Window** button  on every window
- ▶ Help browser includes search tool
- ▶ More options in Help Menu (including **Help for Window in Browser** item )
- ▶ Or print out the 350-page manual

- Support by mail:

- ▶ on list: topcat-user@bristol.ac.uk
- ▶ in person: m.b.taylor@bristol.ac.uk



SAMP

Simple Application Messaging Protocol

Tools can exchange data

- table
- row selection
- FITS image
- spectrum
- sky position

Wide support

- Desktop tools: TOPCAT, ds9, Aladin, SPLAT, MS WWT, ASPRO2, HIPE, ...
- Web pages: Vizier, MAST, Xamin, ...
- Languages: Python, JavaScript, C, Java, ...

The image displays three overlapping windows from the SAMP ecosystem. The top window is 'Plane Plot', showing a scatter plot of points in red and blue on a coordinate system with 'pmDE / mas/yr' on the y-axis. The middle window is 'Aladin v8.0', showing a star field with a grid and various toolbars. The bottom window is 'VizieR - Mozilla Firefox', showing a search interface for 'pleiades' and a table of results.

Full	<i>r</i>	RAJ2000	DEJ2000	HIP	RAhms	DEdms	Vmag	RA(IC)	deg
1	0.111128	56.871152	24.105137	1770203 47 29.06	+24 06 18.9	2.85	56.8711		
2	0.204985	56.872722	24.288344	1770403 47 29.44	+24 17 18.4	6.83	56.8726		
3	0.228006	56.581558	23.948358	1760803 46 19.56	+23 56 54.5	4.14	56.5815		
4	0.323250	56.837371	23.803347	1769203 47 20.96	+23 48 12.4	6.99	56.8373		
5	0.366923	56.456694	24.367749	1757303 45 49.59	+24 22 04.3	3.87	56.4566		
6	0.395330	56.821427	23.726751	1768403 47 17.16	+23 43 36.5	6.98	56.8215		
7	0.403541	56.747488	24.520124	1766403 46 59.38	+24 31 12.8	6.83	56.7474		
8	0.411781	57.125397	24.345524	1779103 48 30.08	+24 20 44.3	6.96	57.1253		
9	0.464967	56.512085	24.527898	1758803 46 02.89	+24 31 40.8	6.43	56.5120		
10	0.484807	56.218905	24.113339	1749903 44 52.52	+24 06 48.4	3.72	56.2188		
11	0.497514	57.290594	24.053415	1784703 49 09.73	+24 03 12.7	3.62	57.2905		
12	0.499333	57.296734	24.136712	1785103 49 11.20	+24 08 12.6	5.05	57.2966		
13	0.503665	56.476985	24.554511	1757903 45 54.46	+24 33 16.6	5.76	56.4769		
14	0.515151	57.237254	23.857135	1783203 48 56.91	+23 51 26.2	6.47	57.2371		

STILTS

STIL Tool Set (STIL = Starlink Tables Infrastructure Library)

- Has pretty much the same capabilities as TOPCAT
- but works from the command line (also **JyStilts** from Jython)

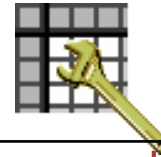
TOPCAT



GUI

Interactive
Easy to use
Good for data exploration
Exploratory phase
few $\times 10^6$ rows

STILTS



Command line

Scriptable
Reproducible
Good for batch/programmed use
Production phase
Unlimited size (for most things)

Typical usage:

- start off with TOPCAT
- maybe move on to STILTS for more specialised requirements

Hands On Examples

After the break (10:30–12:30), **White Room**

- Rough schedule:
 - 10:30** TOPCAT visualisation (Mark Taylor)
 - 11:00** TOPCAT crossmatching (Mark Taylor)
 - 11:30** TAP/ADQL (Simon Murphy)
- Materials (script + data files):
 - ▷ Online:
http://andromeda.star.bris.ac.uk/topcat_gaiabds/tcex/
 - ▷ Download tarball (5 Mb):
http://andromeda.star.bris.ac.uk/topcat_gaiabds/tcex.tar.gz
 - ▷ ... or copy it from me on a USB stick

... or **Brown Dwarf Ages**, Aula Magna