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)?
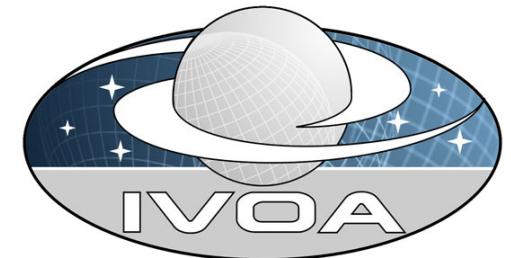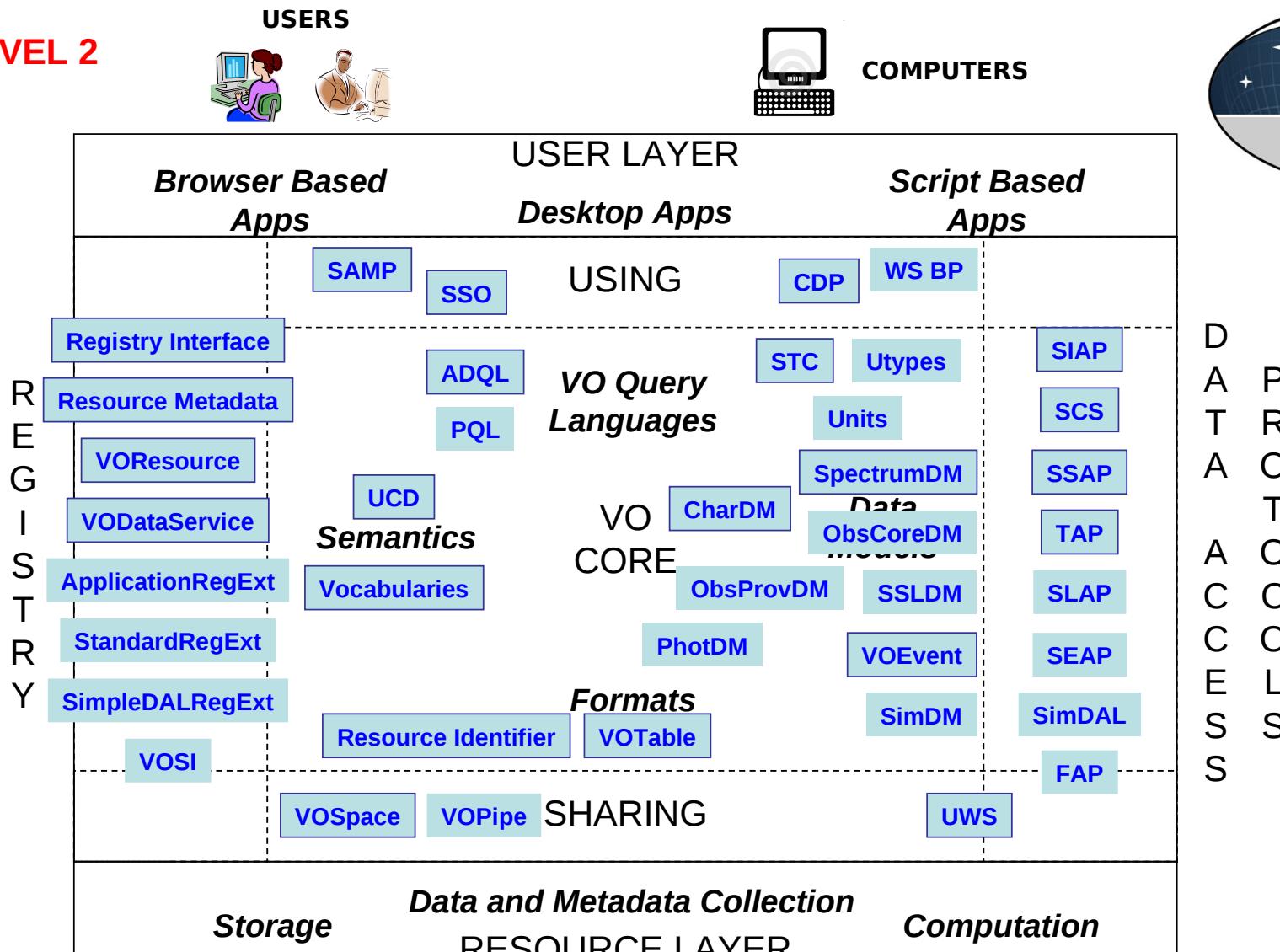
– *"All astro archives in your computer"*

– 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, ...)
   $\Rightarrow$ multi-wavelength, multi-dataset science

Virtual Observatory Architecture

LEVEL 2

USERS

COMPUTERS

USER LAYER

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

USING

| SAMP | SSO | | CDP | WS BP |

REGISTRY

Registry Interface

Resource Metadata

VOResource

VODataService

ApplicationRegExt

StandardRegExt

SimpleDALRegExt

VOSI

| ADQL | | STC | Utypes |
| PQL | **VO Query Languages** | Units |

UCD

Semantics

Vocabularies

VO CORE

CharDM

Data Models

SpectrumDM

ObsCoreDM

ObsProvDM   SSLDM

PhotDM   VOEvent

SimDM

Formats

Resource Identifier   VOTable

DATA ACCESS PROTOCOLS

SIAP

SCS

SSAP

TAP

SLAP

SEAP

SimDAL

FAP

SHARING

VOSpace   VOPipe   UWS

Data and Metadata Collection

Storage   *Computation*

RESOURCE LAYER

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
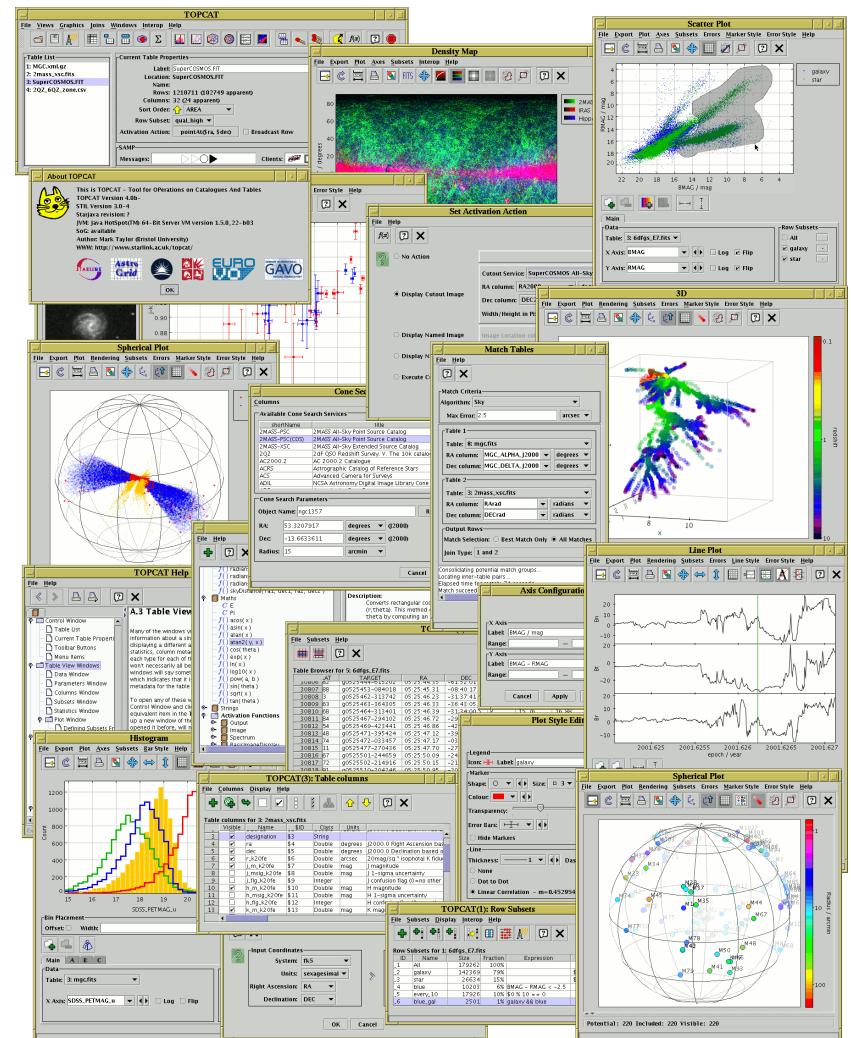
TOPCAT Overview

TOPCAT = Tool for OPerations on Catalogues And Tables

"Does what you want with tables"

Aims:
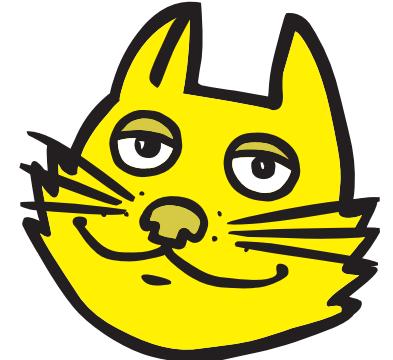
- **Easy** to use

- 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
  (10^6 rows $\times$ 10^2 cols easily, maybe more)

- User-driven development

TOPCAT Capabilities

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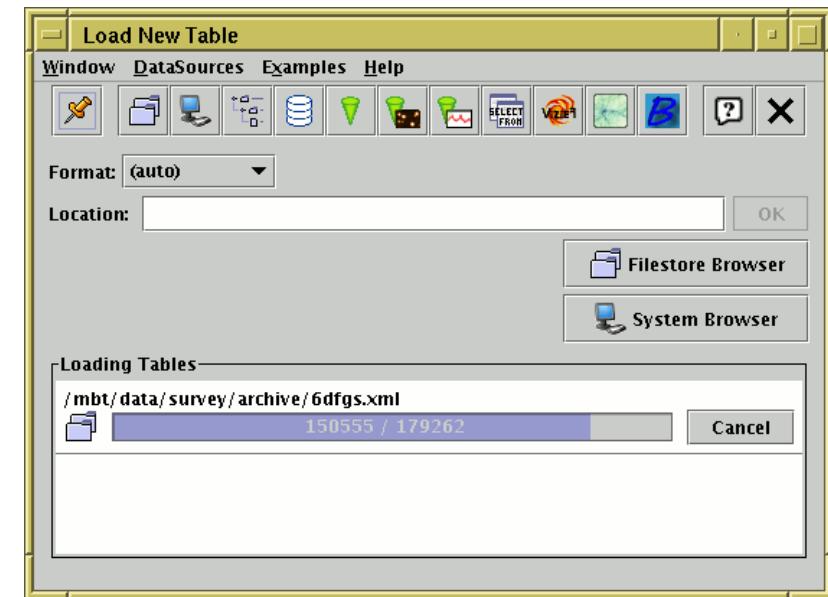
- Read/write tables in multiple formats
- View/edit data
- View/edit metadata
- Calculations and statistics
- Visualisation
- Make/combine/display row selections in various ways (linked views)
- Crossmatching — efficient and very flexible
- Access external data services (VO and others)
- Talk to other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

Table data view

Table Metadata view

Column Metadata view

Row Selections
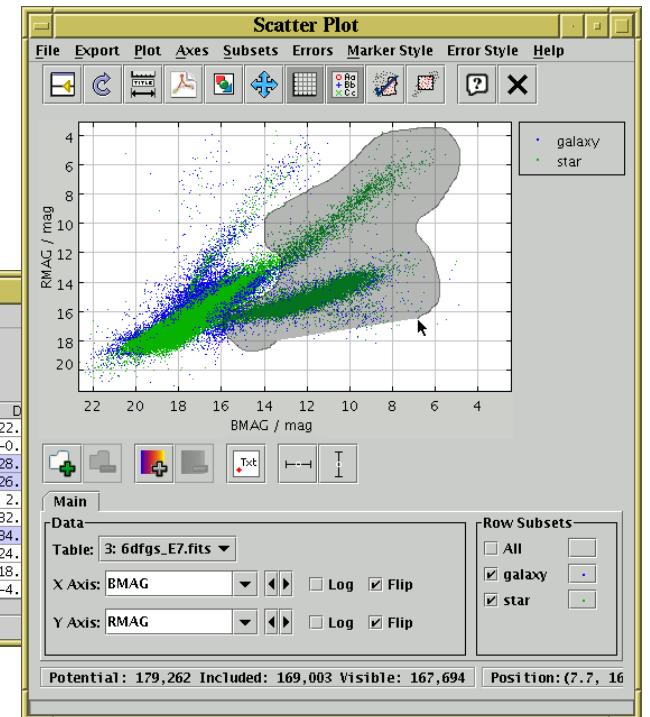
Different ways to make single or multiple row selections:

- 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)

Linked views mean a selection made one way is visible in other ways
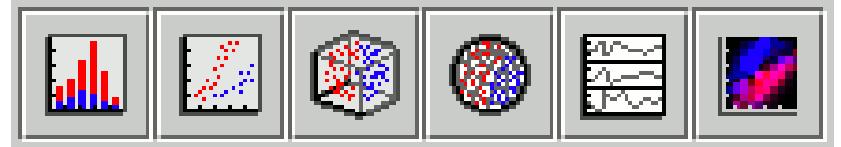
- Perform crossmatch only on items in red giant branch
- Where on the sky is this colour cut?
- Spot outliers
- Identify objects on ds9 image display

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`
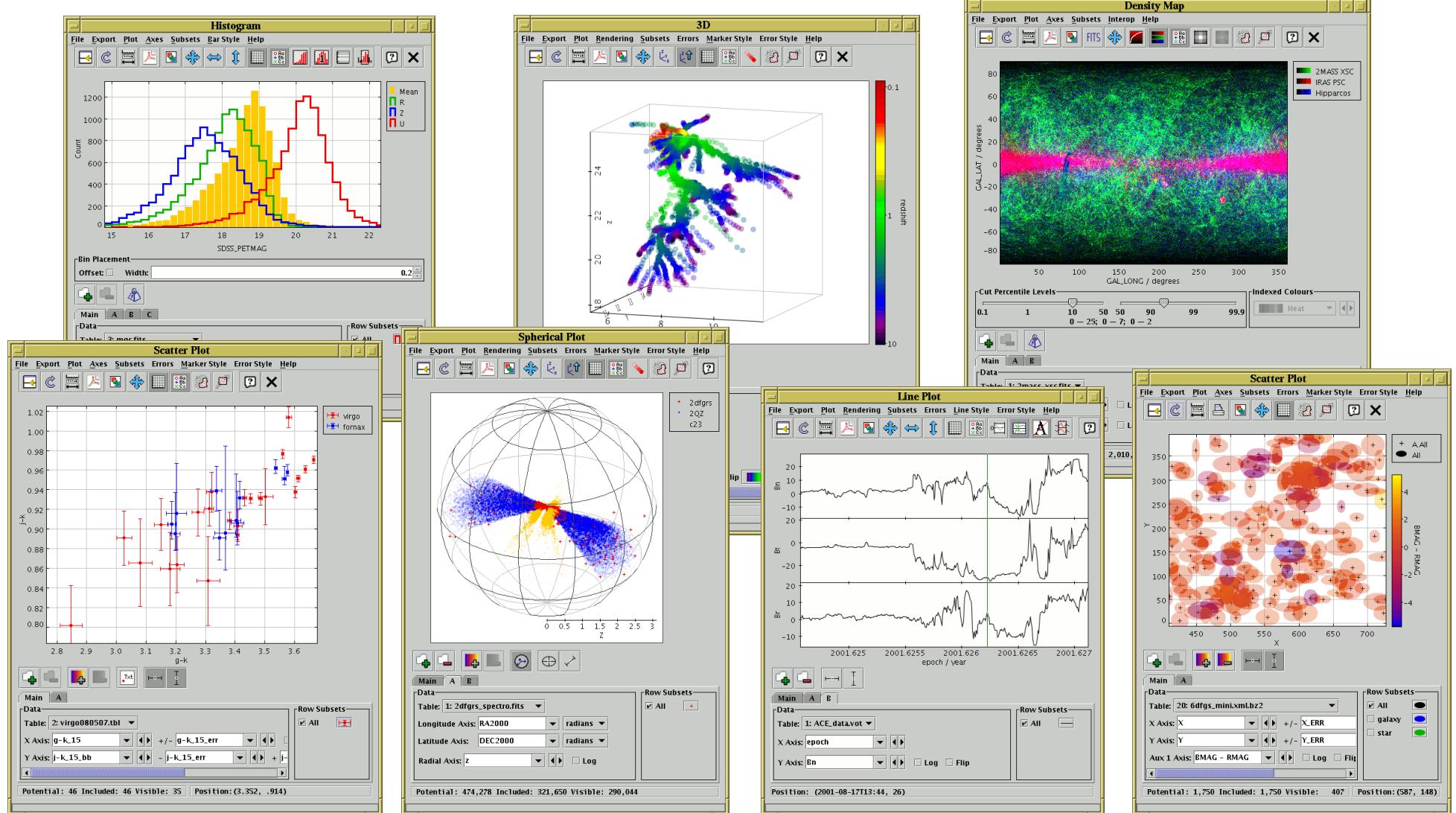
TOPCAT: v3 Visualisation

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
- ...
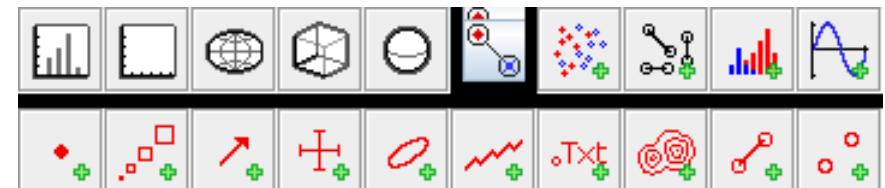
TOPCAT: v3 Visualisation

# TOPCAT v4 Visualisation

v4.0b March 2013, v4.1 March 2014
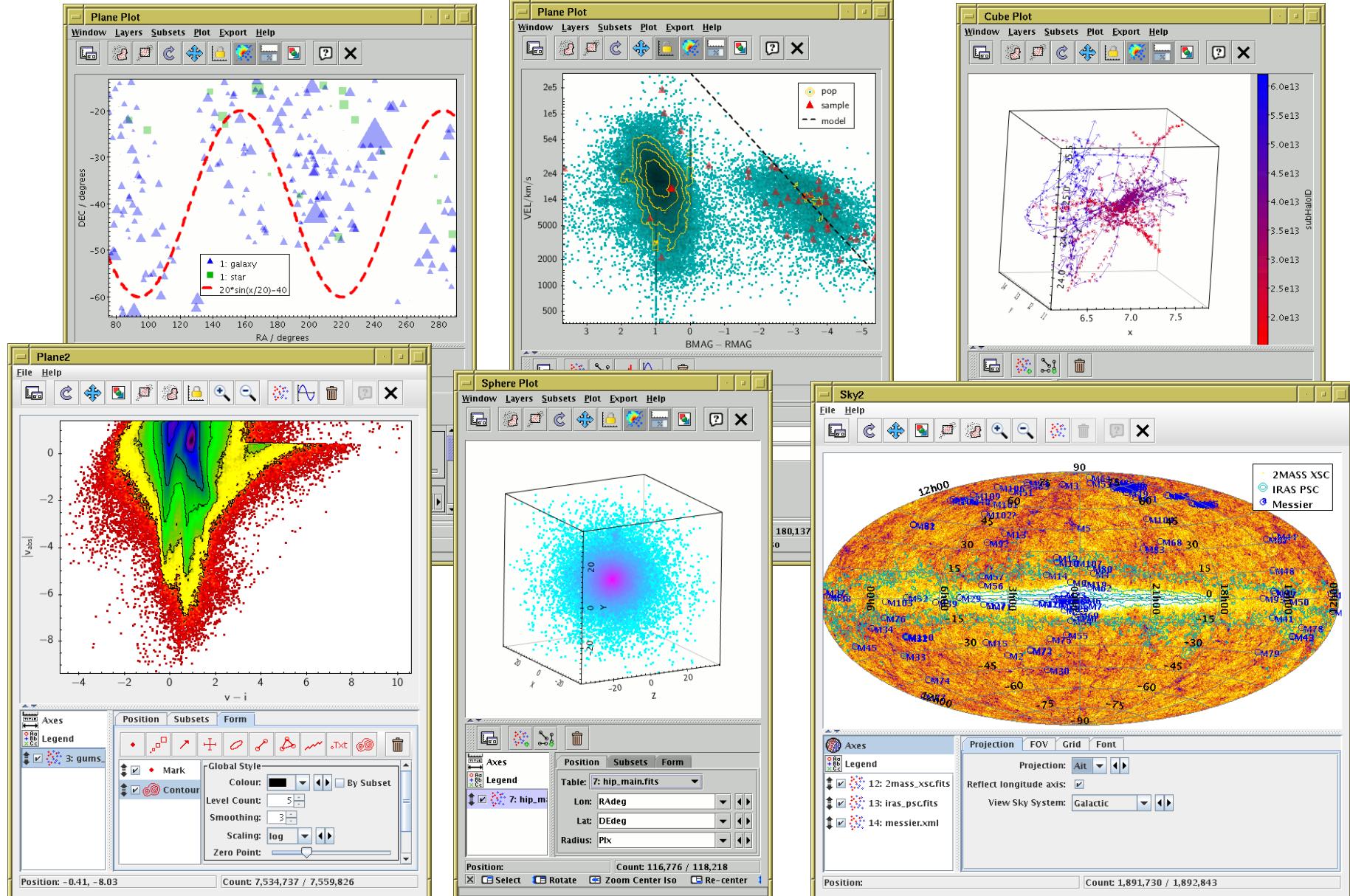
Most of the version 3 functions **plus**:

- 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
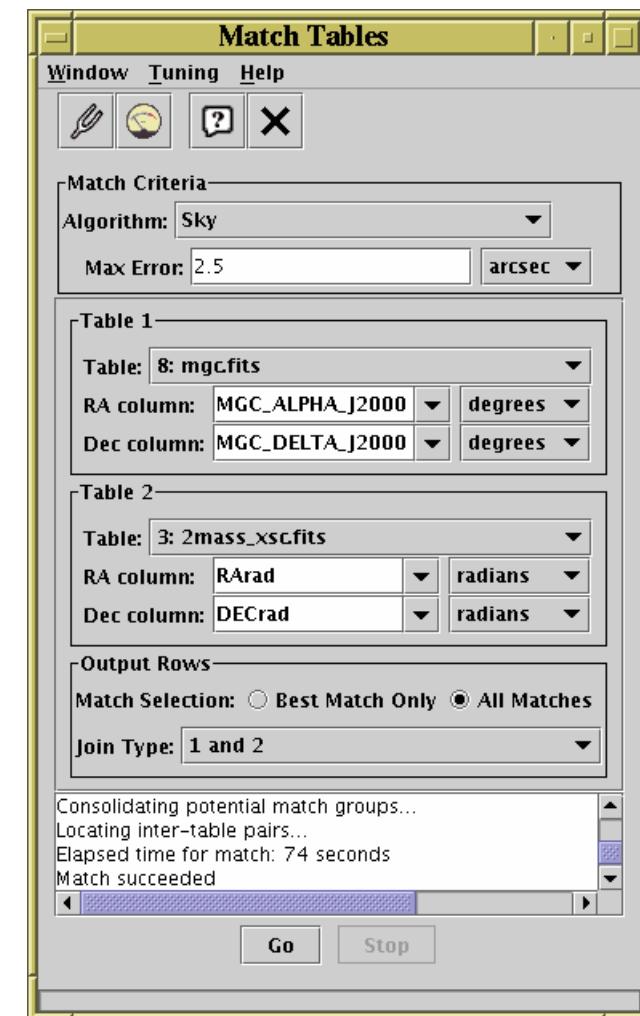
Various options for crossmatching tables

- 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:* Plot Result visualise result

TOPCAT TAP Window
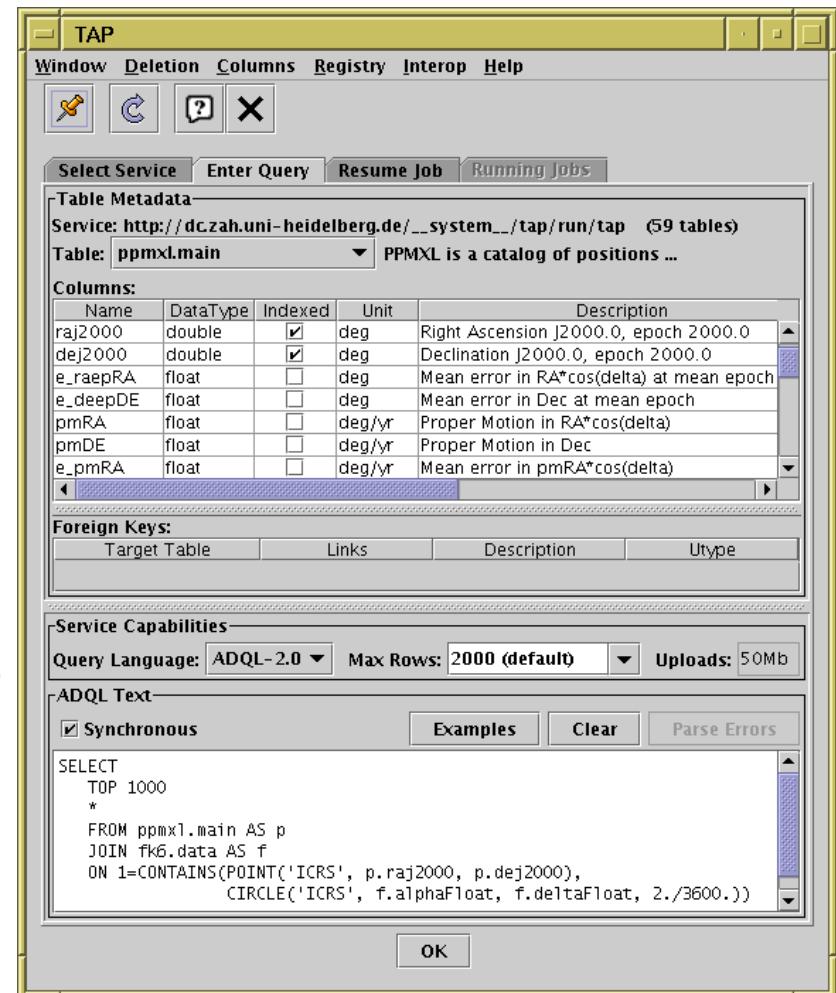
TAP (Table Access Protocol)

- 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.))
```
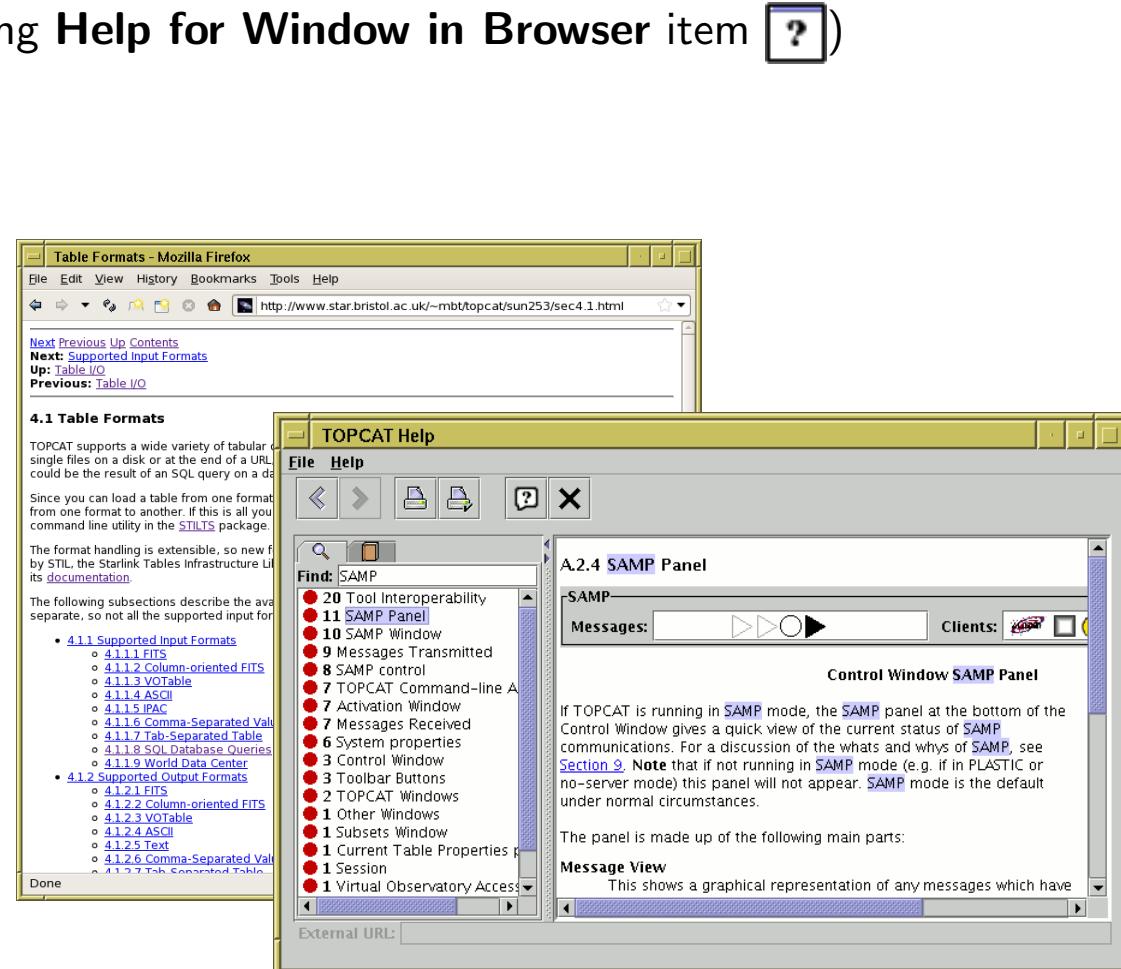
TOPCAT: Help!

- 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
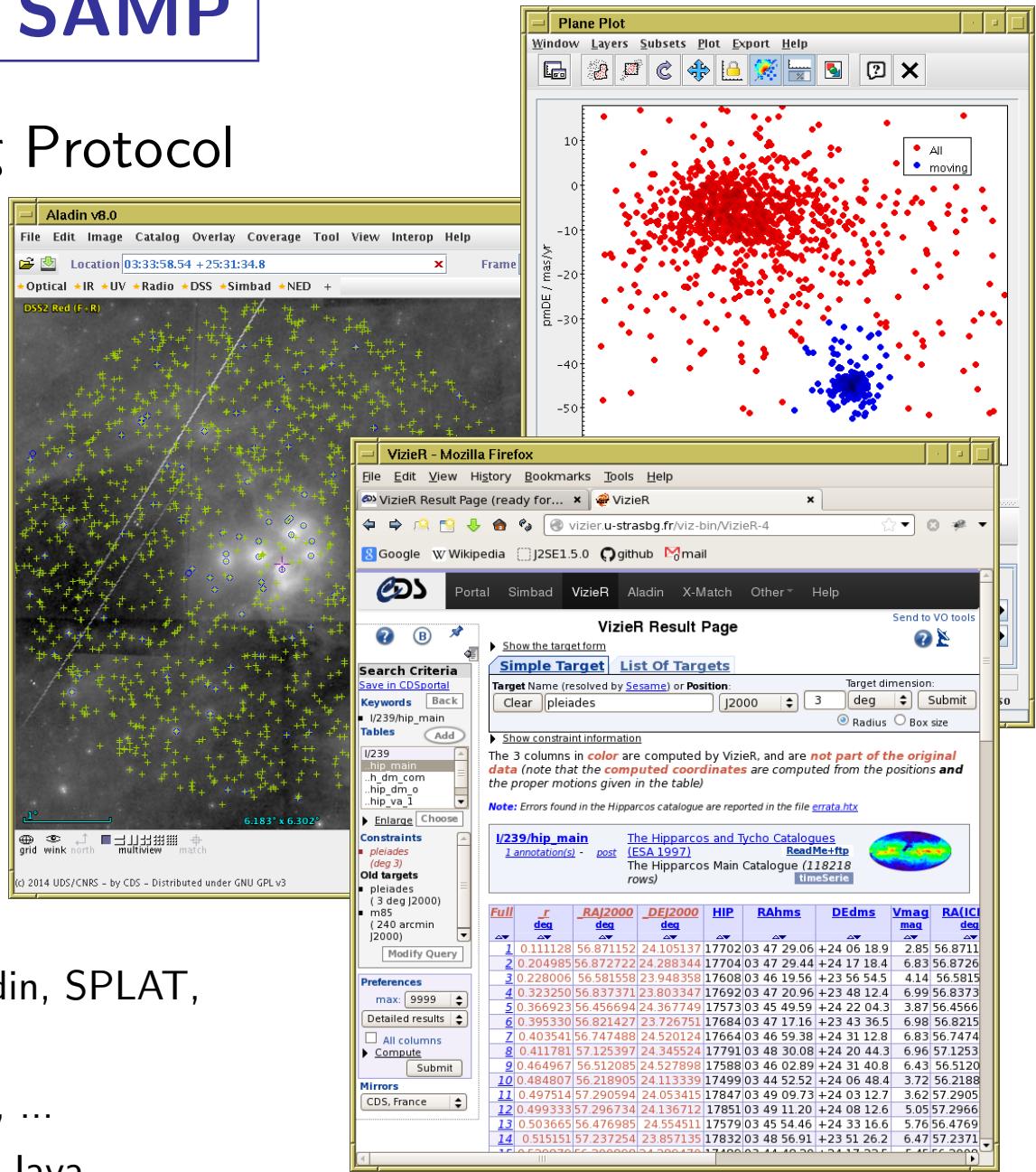
Simple Application Messaging Protocol

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, ...

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	**STILTS**
GUI	Command line
Interactive	Scriptable
Easy to use	Reproducible
Good for data exploration	Good for batch/programmed use
Exploratory phase	Production phase
few$\times 10^6$ rows	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