The Hitchhiker's Guide to Astronomical Databases

Author: Emil Puha[1]

(1) Department of Astronomy, Physics of the Earth, and Meteorology, FMPI, Comenius University, Mlynská dolina, 842 48 Bratislava

Overview

- **Simbad and Vizier catalogs**
  - Basic handling, searching the catalogs
  - Export of tables
- **TopCat**
  - Basic handling
  - Processing tables and creating plots
  - Cross-matching tables
  - TAP queries
- **Useful Python libraries**
  - Astroquery
  - PyStilts
  - Pandas

Simbad and Vizier catalogs

Strasbourg astronomical Data Center (CDS)

- Strasbourg astronomical Data Center (CDS) is dedicated to the collection and worldwide distribution of astronomical data and related information
- The CDS mission is to:
  - Collect useful information concerning astronomical objects that is available in computerized form
  - Upgrade these data by critical evaluations and comparisons
  - Distribute the results to the astronomical community
  - Conduct research, using these data

CDS

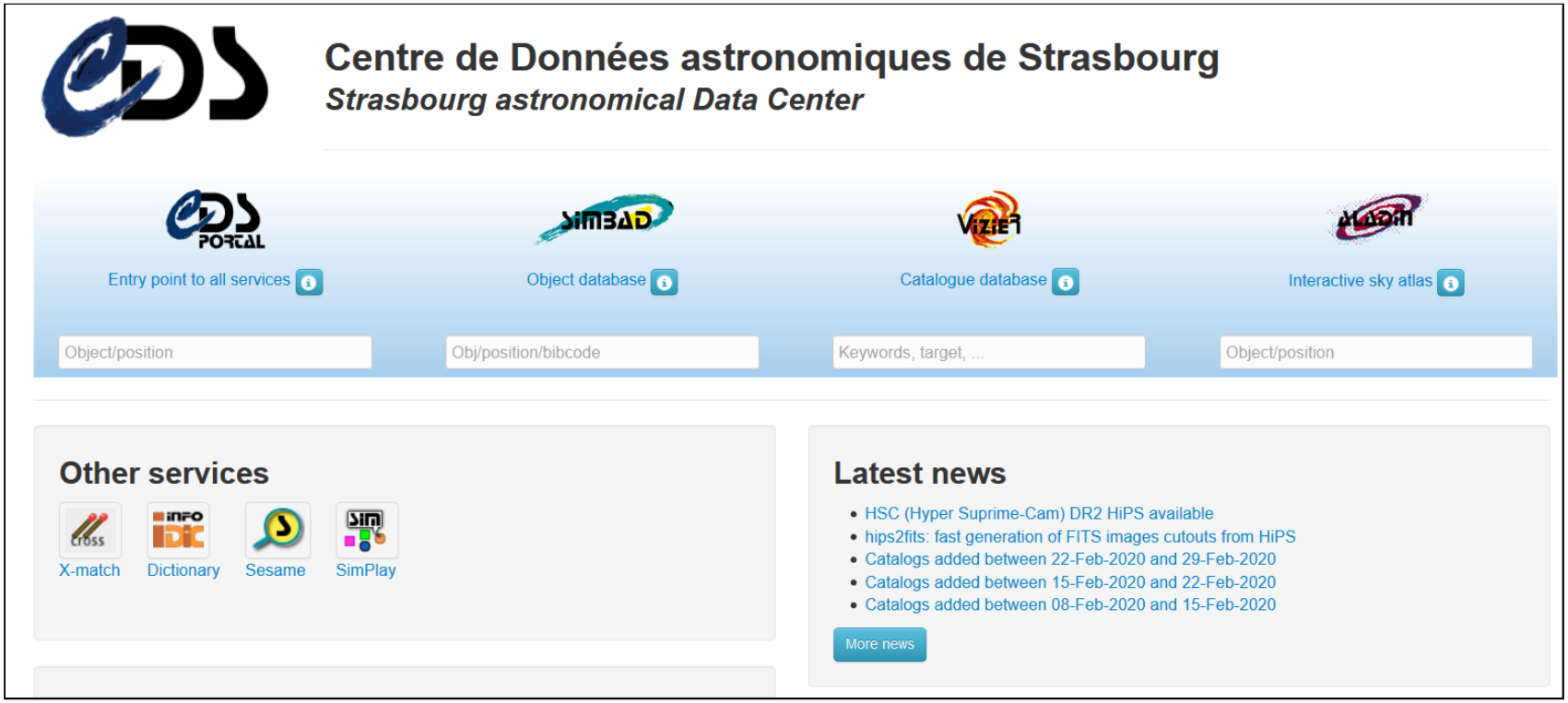


Figure: The home page of CDS.

Strasbourg astronomical Data Center (CDS)

- The CDS hosts:
  - **SIMBAD** astronomical database, the world reference database for the identification of astronomical objects
  - **VizieR** catalogue service for the CDS reference collection of astronomical catalogues and tables published in academic journals
  - **Aladin** interactive software sky atlas for access, visualization and analysis of astronomical images, surveys, catalogues, databases and related data

Simbad

- **SIMBAD** (the Set of Identifications, Measurements and Bibliography for Astronomical Data) is an astronomical database of objects beyond the Solar System.
- **SIMBAD** was created by merging the Catalog of Stellar Identifications (CSI) and the Bibliographic Star Index as they existed at the Meudon Computer Centre until 1979, and then expanded by additional source data from other catalogues and the academic literature.

Simbad

- **SIMBAD** (the Set of Identifications, Measurements and Bibliography for Astronomical Data) is an astronomical database of objects beyond the Solar System.
- **SIMBAD** was created by merging the Catalog of Stellar Identifications (CSI) and the Bibliographic Star Index as they existed at the Meudon Computer Centre until 1979, and then expanded by additional source data from other catalogues and the academic literature.
- As of 2 March 2020, it contains information for 11,021,649 objects under 35,753,171 different names, with 368,154 bibliographical references and 20,713,563 bibliographic citations.

Simbad queries

- **SIMBAD** is easily accessible
- It allows queries by
  - Object identifiers
  - Coordinates
  - Bibliographical references
  - And various other criteria ...
- Useful is the query of a list of coordinates: you can upload your own ASCII table and query all its rows in **SIMBAD**

Simbad

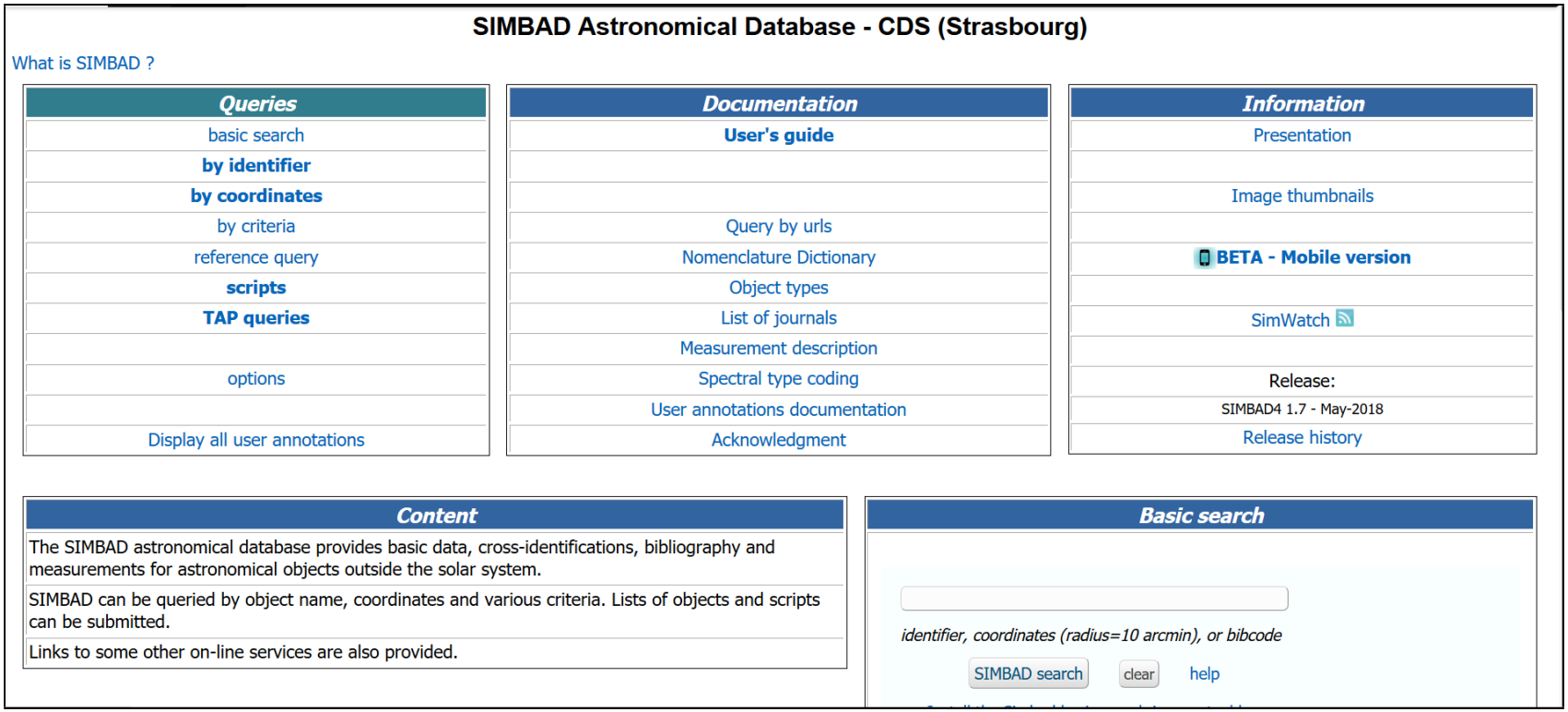


Figure: The home page of Simbad catalog.

Simbad - Coordinate query

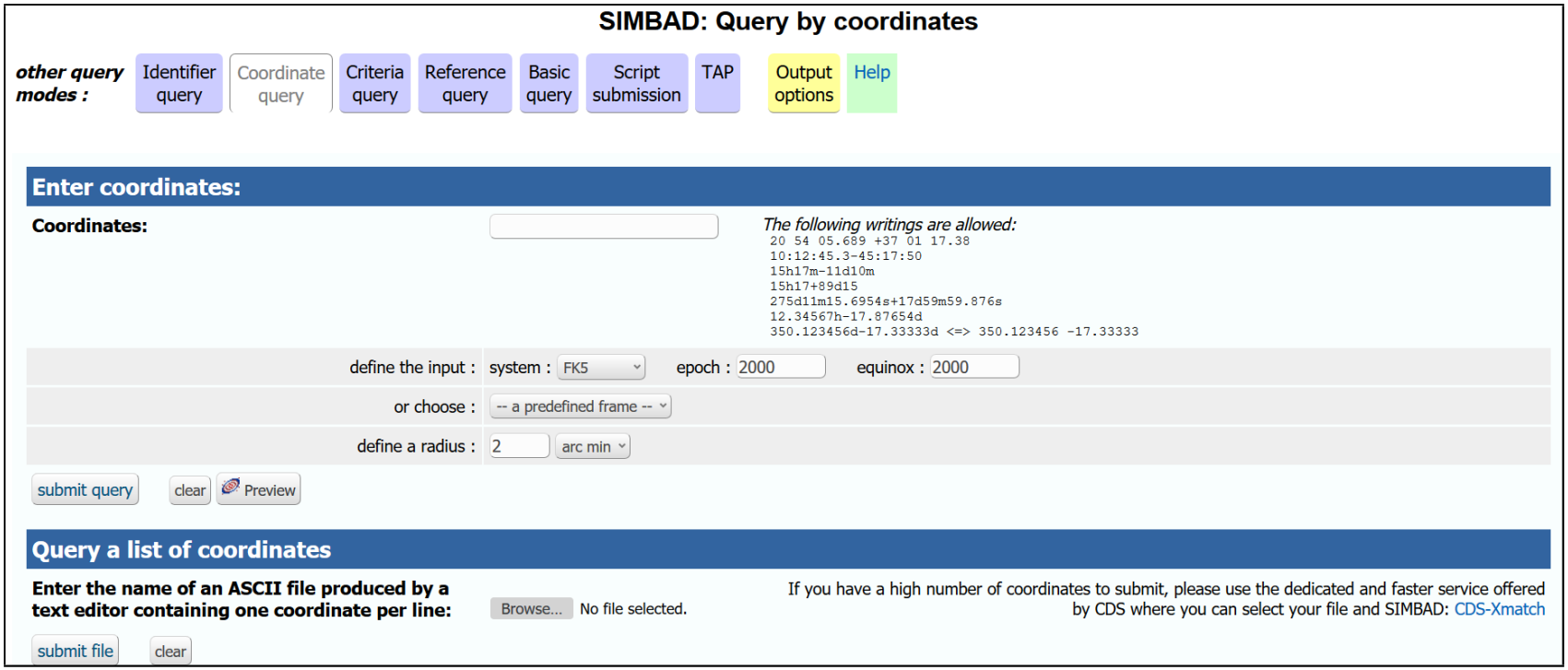


Figure: Simbad query by coordinates.

Simbad - Query from a list

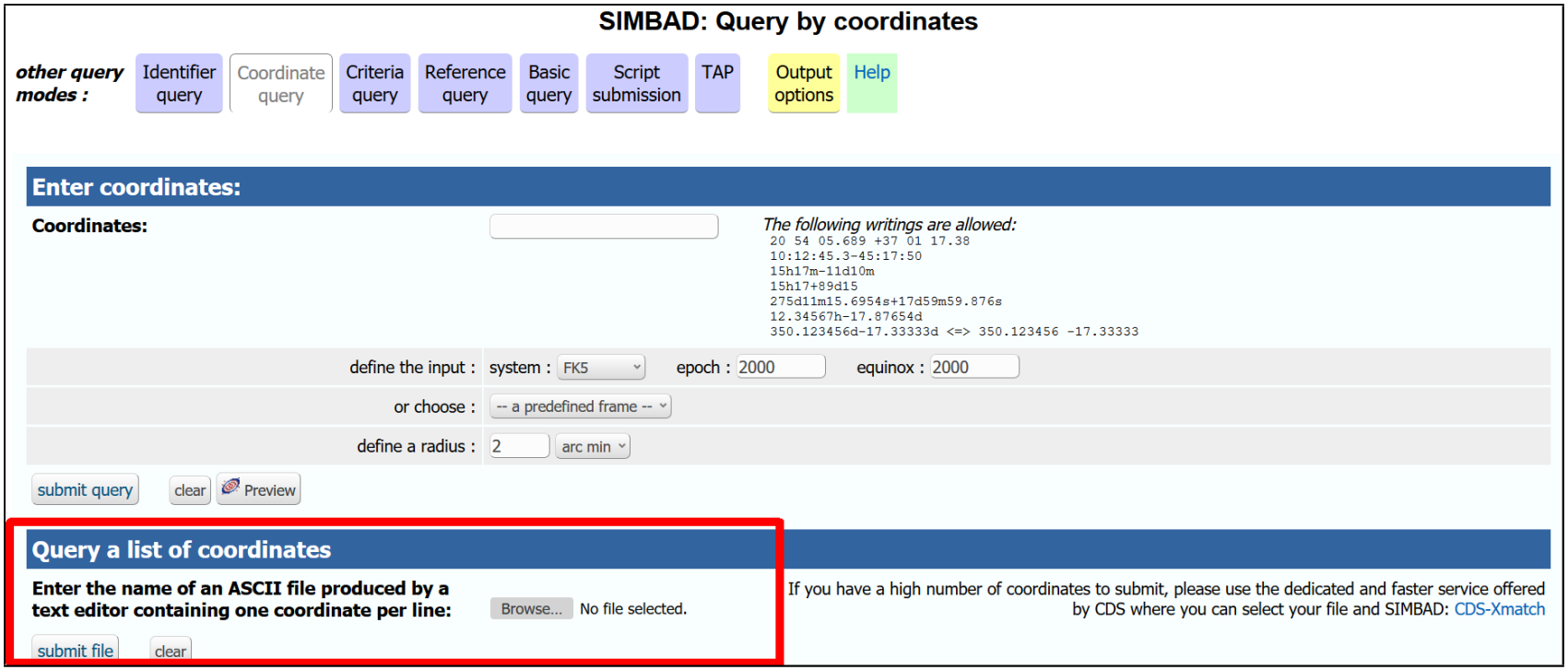


Figure: Simbad query from a list.

Vizier

- **VizieR** is an astronomical catalog service provided by Centre de données astronomiques de Strasbourg (CDS)
- Since 1996, **VizieR** has become a reference point for astronomers worldwide engaged in research, who come to access catalogued data regularly published in astronomical journals
- As of now it contains more than 19586 catalogues

Vizier

- The **VizieR** service allows you to search for specific objects or bibliographical references
- You can able browse various survey data releases such as Gaia, Hipparcos, Lamost, SLOAN ...

VizieR

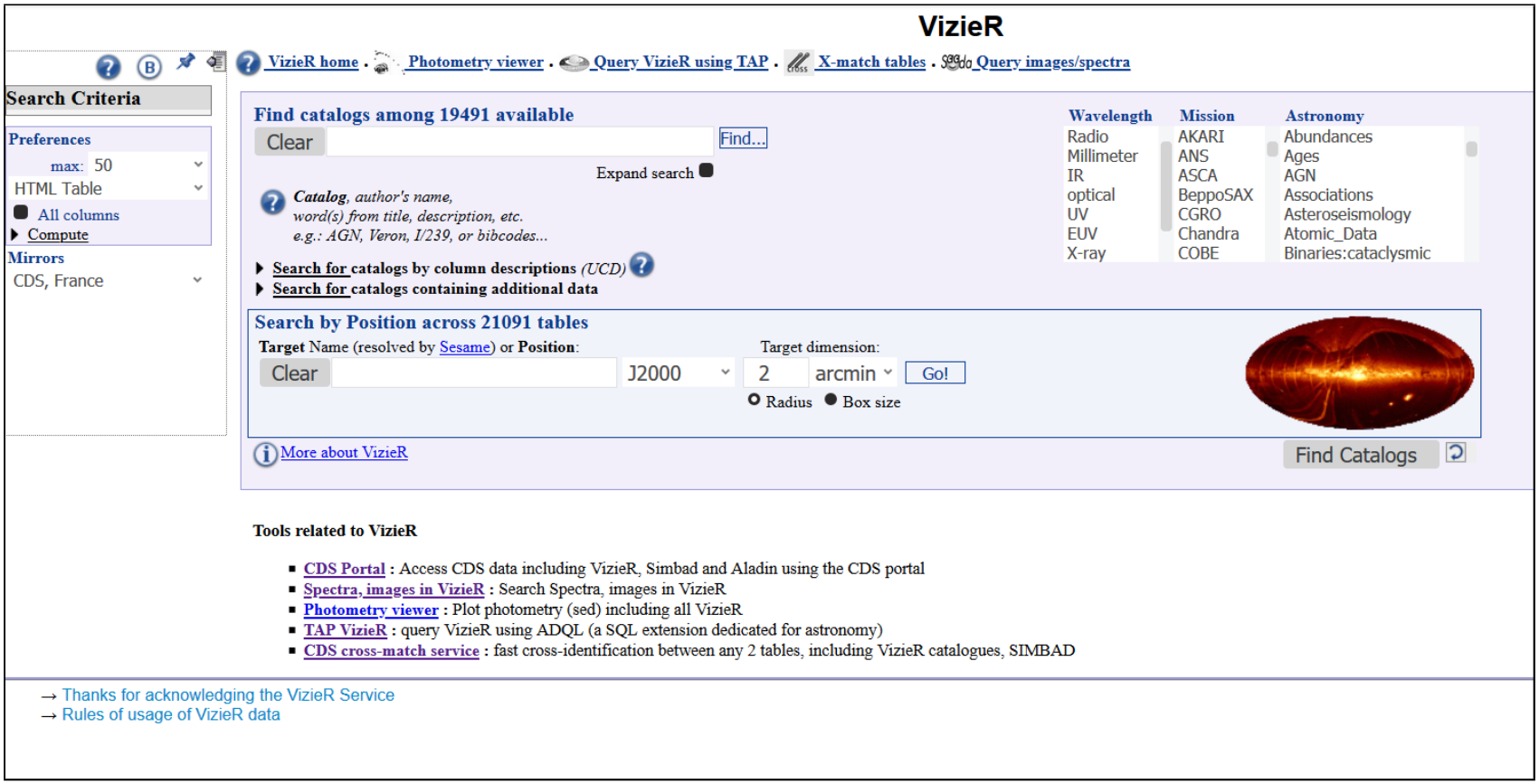


Figure: VizieR query.

VizieR

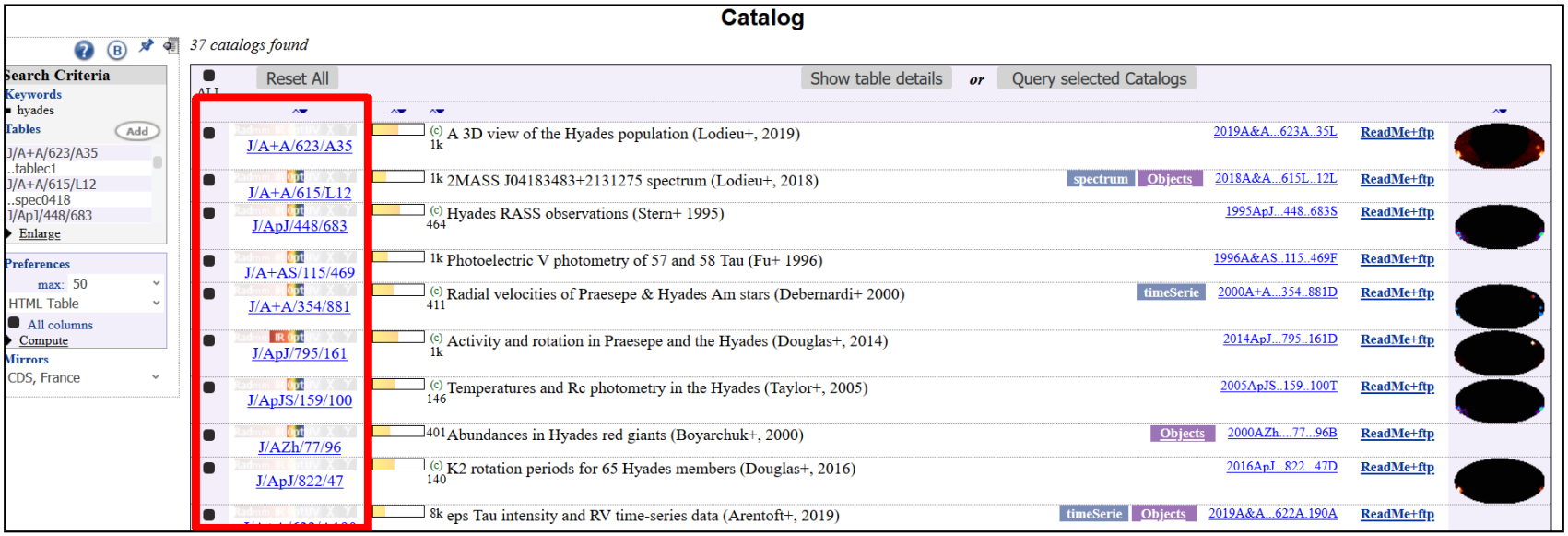


Figure: VizieR query of the Hyades cluster. The query returns all the available catalogs (and papers), which have data of the cluster. The table identifiers are highlighted.

VizieR from ADS

- You can also directly access the **VizieR** service from the ADS
- If a paper on ADS has some tables attached, these tables are often stored on Vizier and Simbad

VizieR from ADS

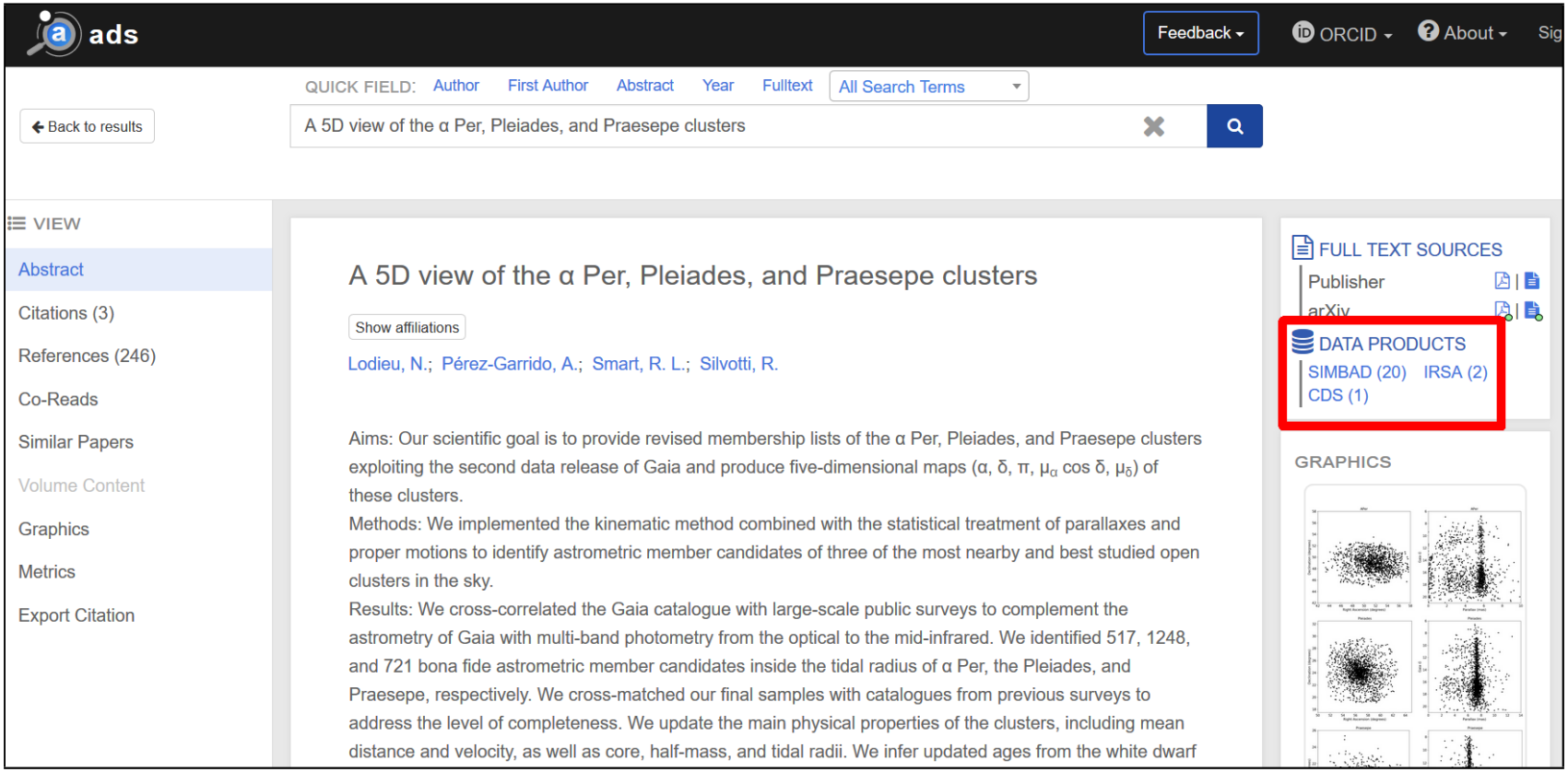


Figure: A link to Vizier, where the corresponding tables of the paper are stored.

VizieR from ADS

- Here you can see and browse all the tables from the paper
- You can query query each column of the table
- Or you can download the table in various formats - voTable, fits, ASCII, csv ...

VizieR from ADS

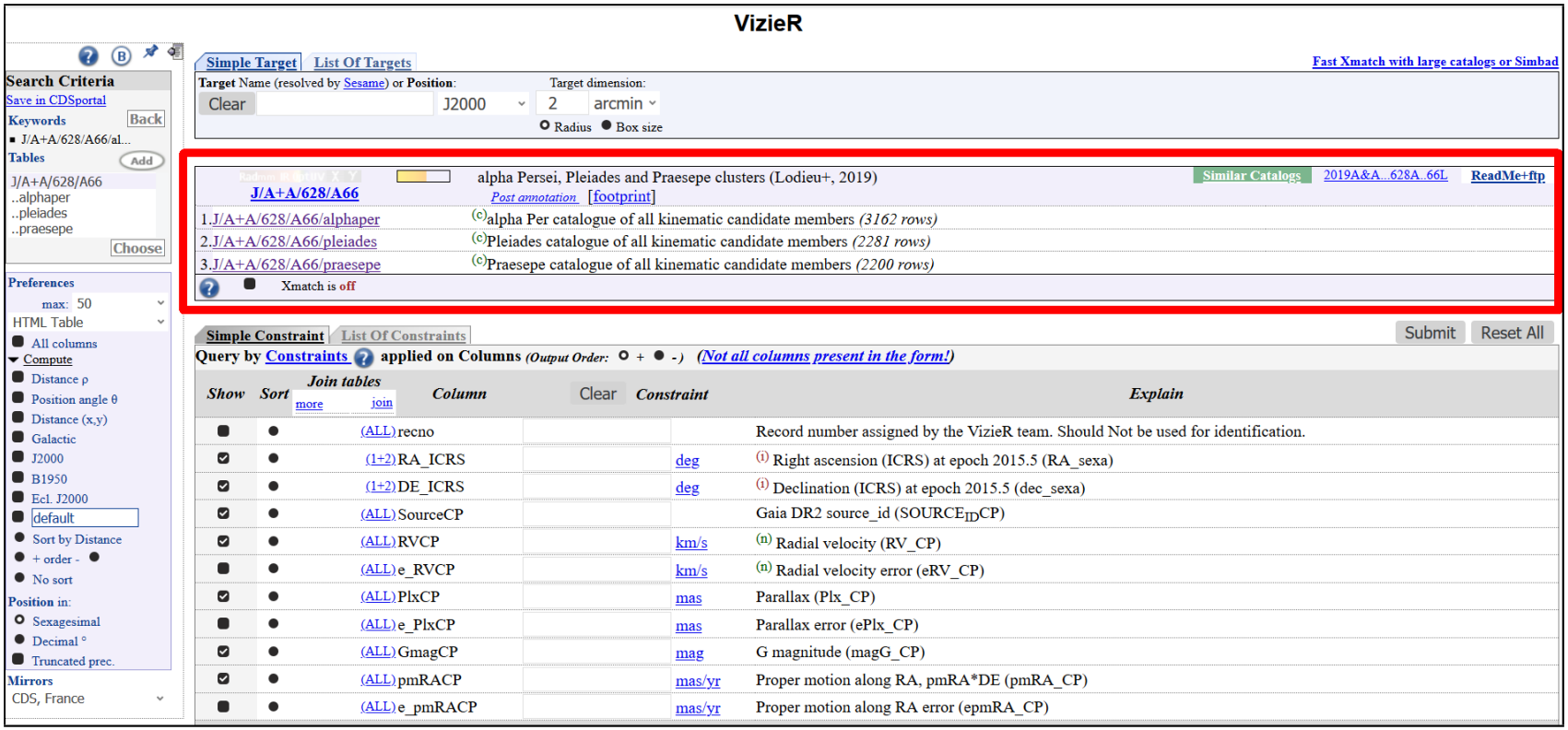


Figure: Tables of the selected paper on Vizier.

VizieR from ADS

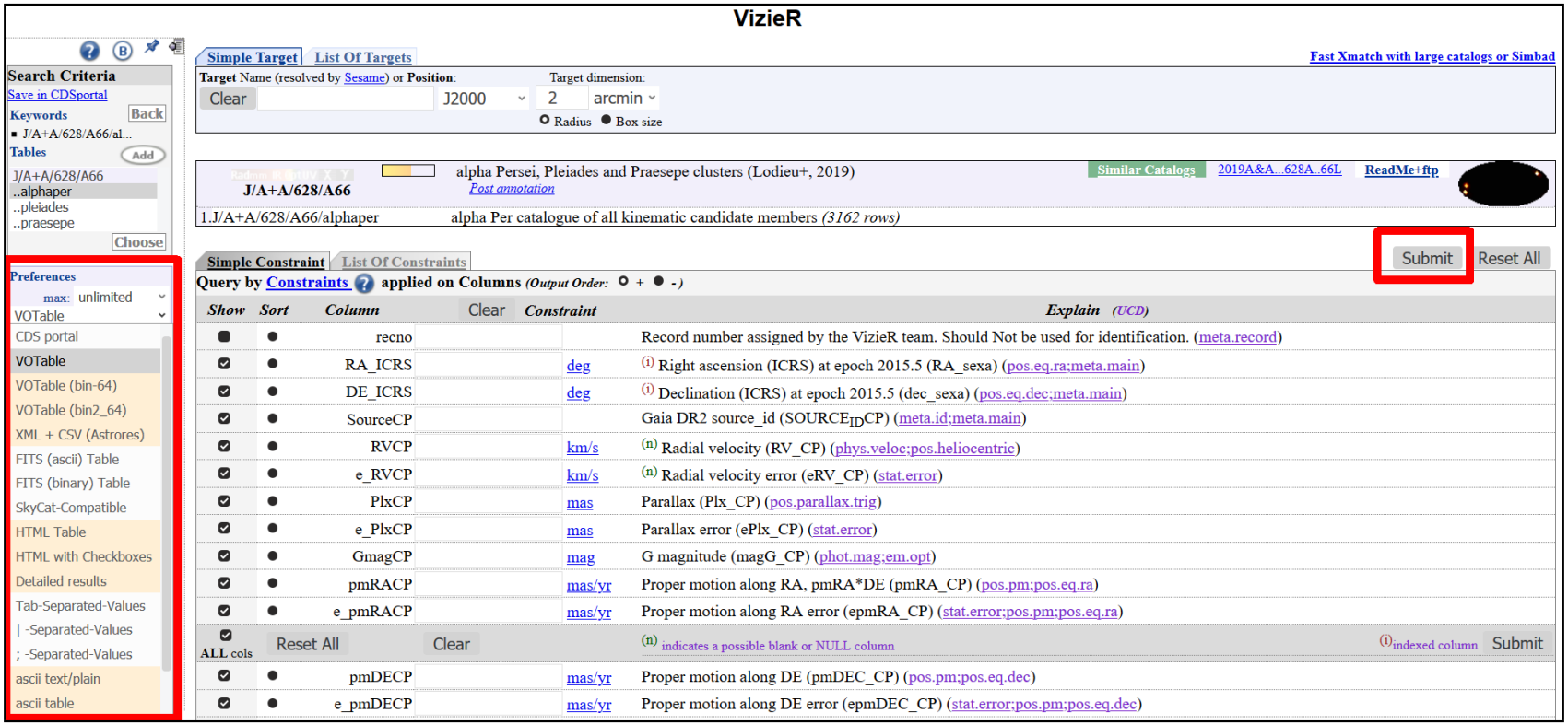


Figure: Download the table in various formats.

Topcat

Topcat

- **Topcat** is an interactive graphical viewer and editor for tabular data
- Its aim is to provide most of the facilities that astronomers need for analysis and manipulation of source catalogues and other tables, though it can be used for non-astronomical data as well
- It understands a number of different astronomically important formats (including FITS, VOTable and CDF) and more formats can be added

Topcat

- Fast access to large datasets (millions of rows/hundreds of columns)
- View/edit table data in a scrollable browser
- Insert 'synthetic' columns defined by algebraic expression
- Sort rows on the values in a given column
- Define row subsets in various ways
- Plot types are histogram, plane, sky, cube, sphere, time
- Features include variable transparency, error bars, point labelling, colour axes, all-sky plots, configurable density shading, vectors, ellipses, polygons, contours, density maps, KDEs, analytic functions, plain text/LaTeX axis annotation,
- Calculate statistics on each column for some or all rows
- Perform flexible and fast matching of rows in the same or different tables
- Acquire tables from a file, URL, or SQL query
- Communicate with external VO and non-VO data services, including TAP, cone search, VizieR, CDS X-Match, CDS Hips2fits, SIA, SSA or DataLink
- Write modified tables out in original or different format to file or an SQL database

Topcat

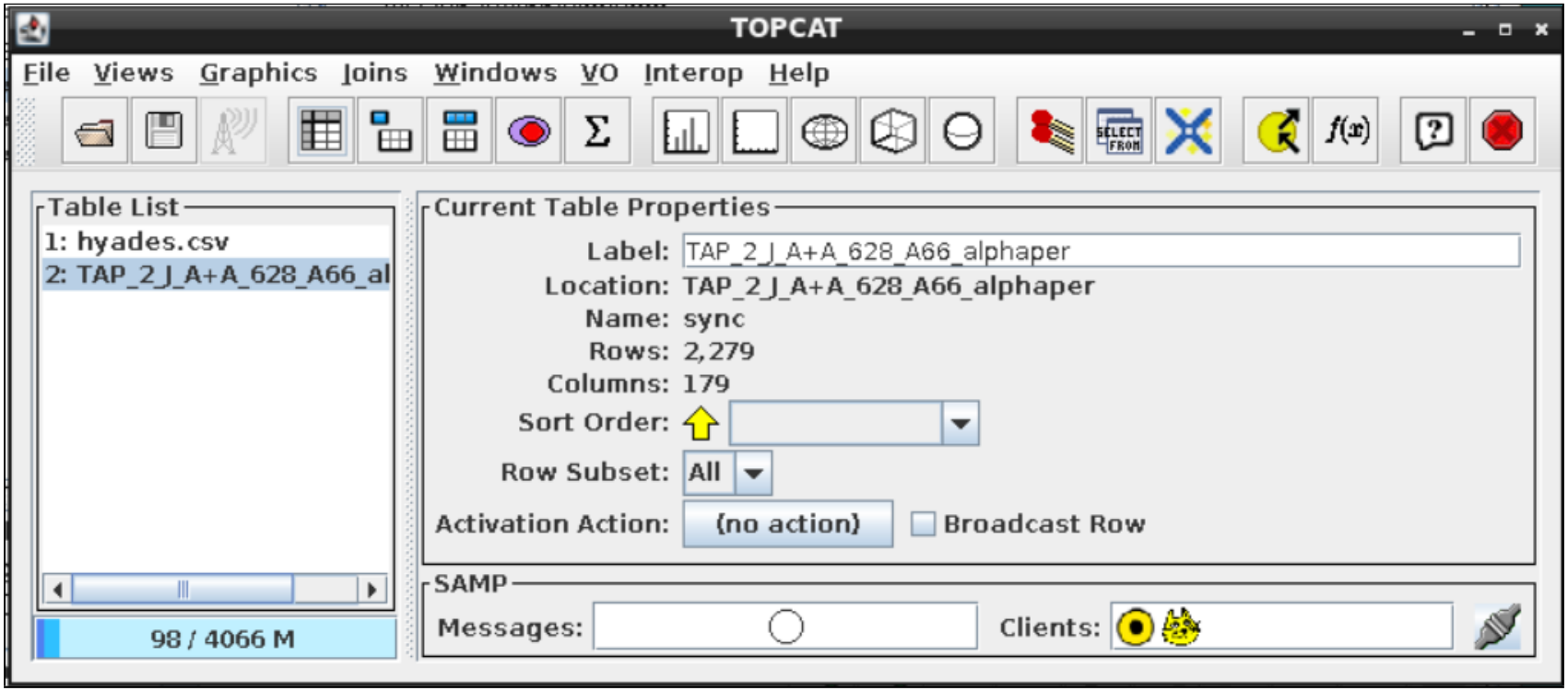


Figure: Topcat - main window.

Topcat features

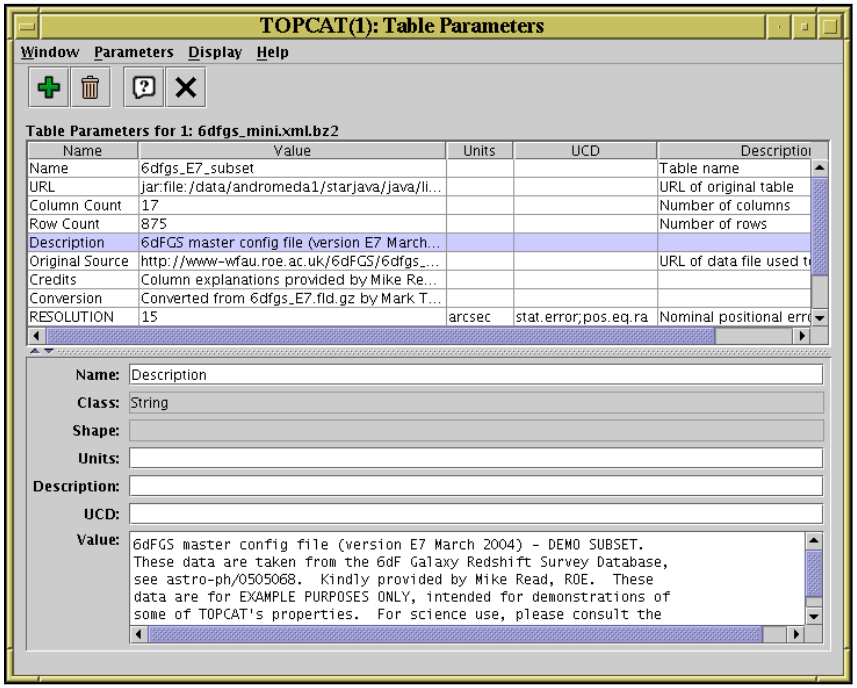
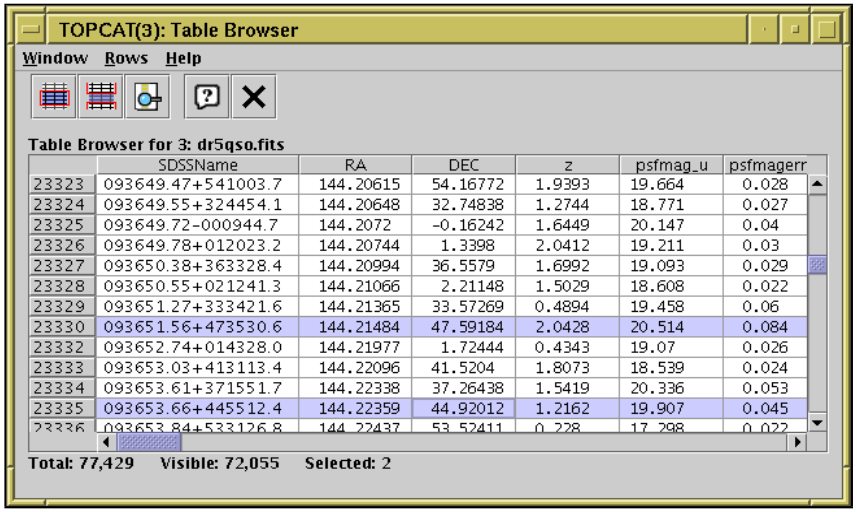


Figure: Topcat - table parameters.



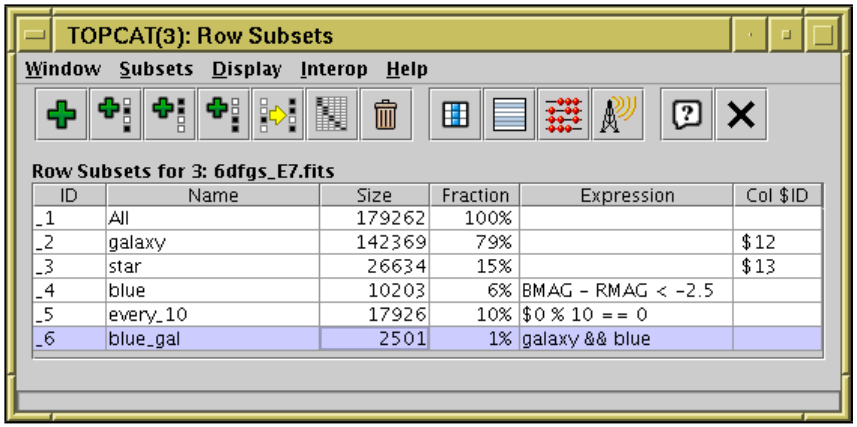
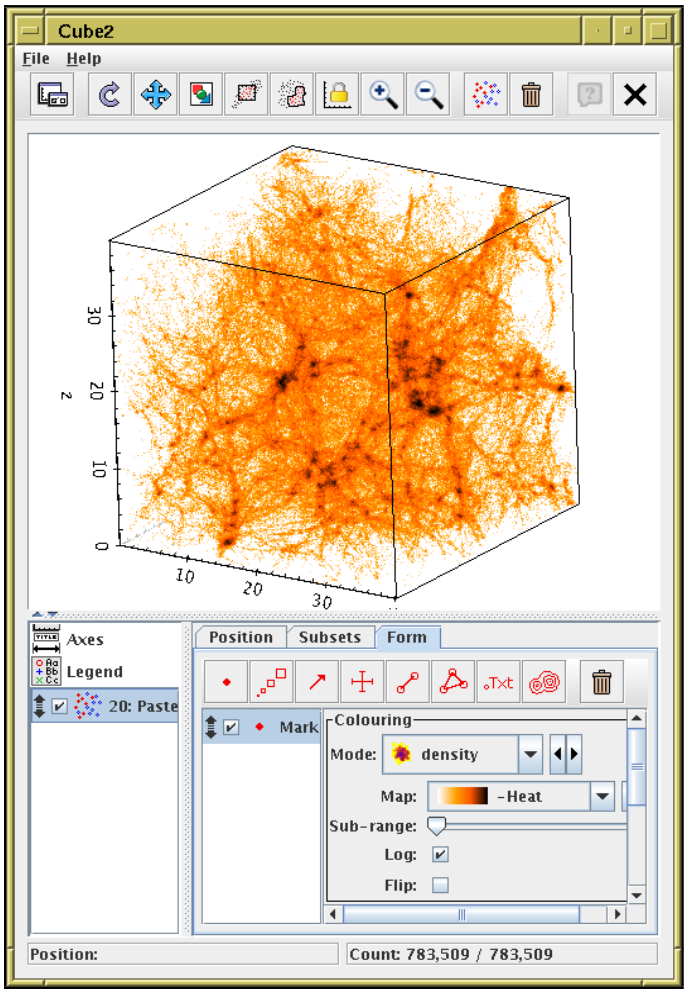


Figure: Topcat - table view and subsets.

Topcat plots



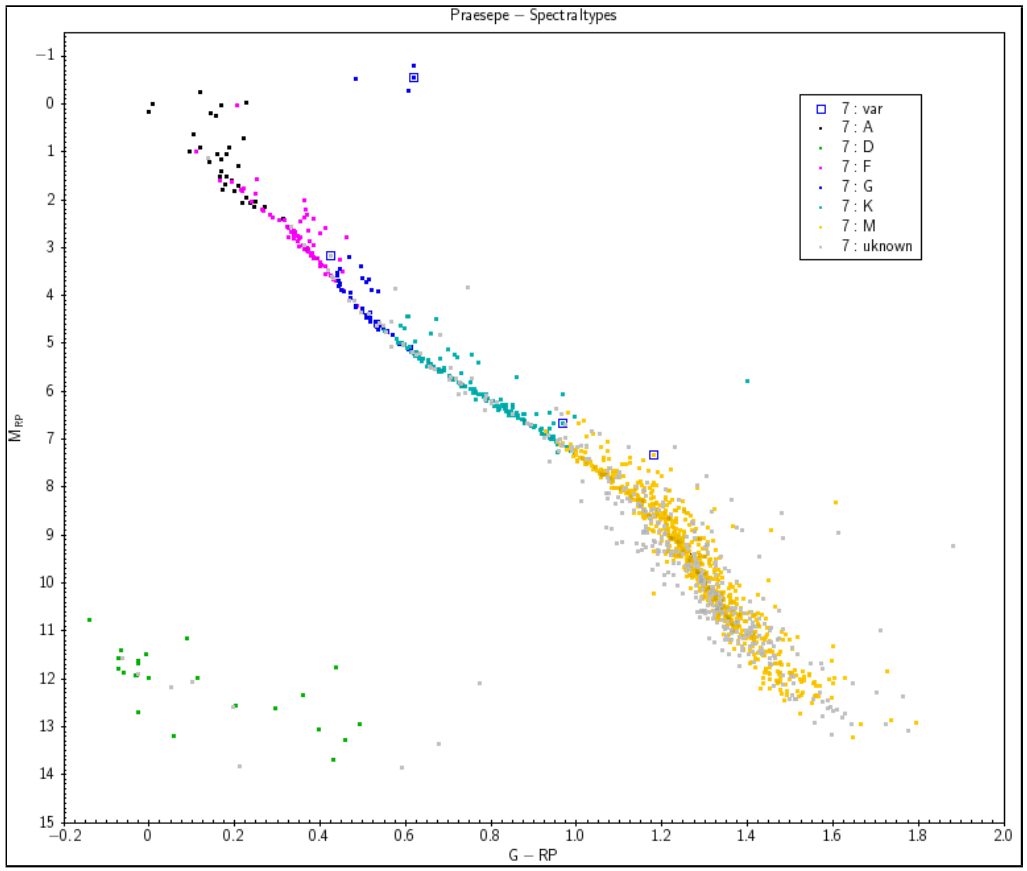
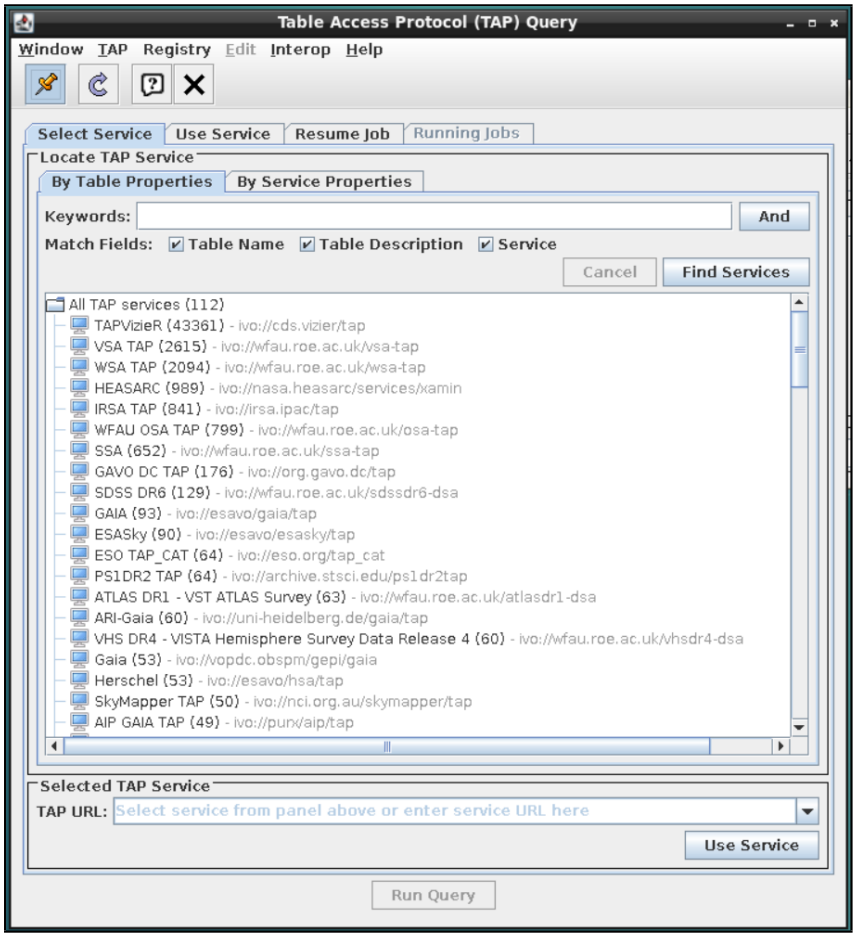


Figure: Various plots produced with Topcat.

Acces catalogs from Topcat - TAP queries

- You can use **Topcat** to directly access and query catalog, which are available online
- The queried are then downloaded to Topcat as a new table
- This feature is called **Table Access Protocol (TAP)** query

Topcat TAP



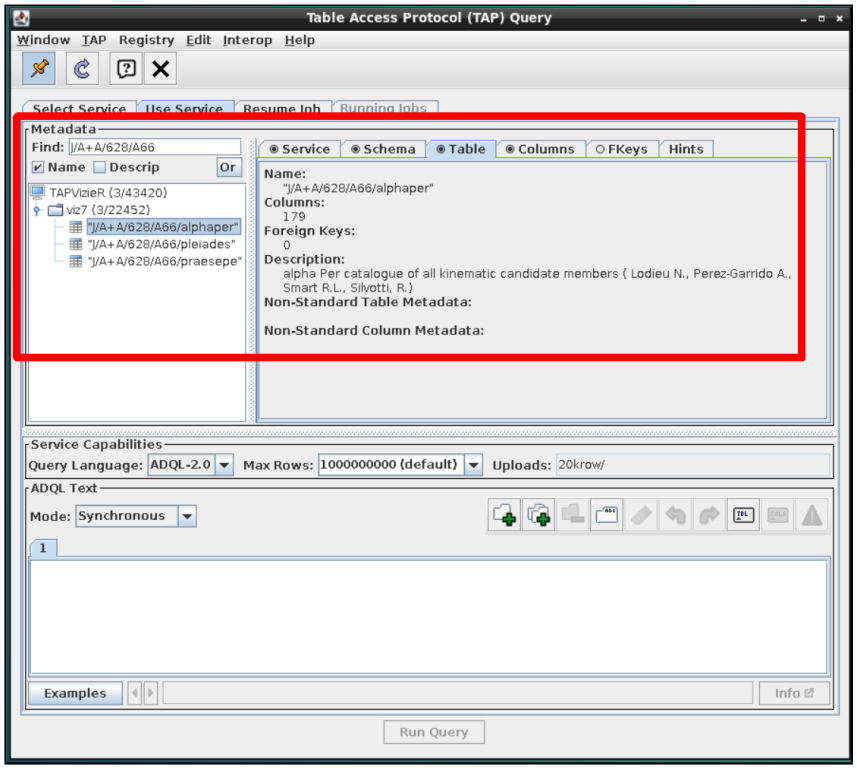
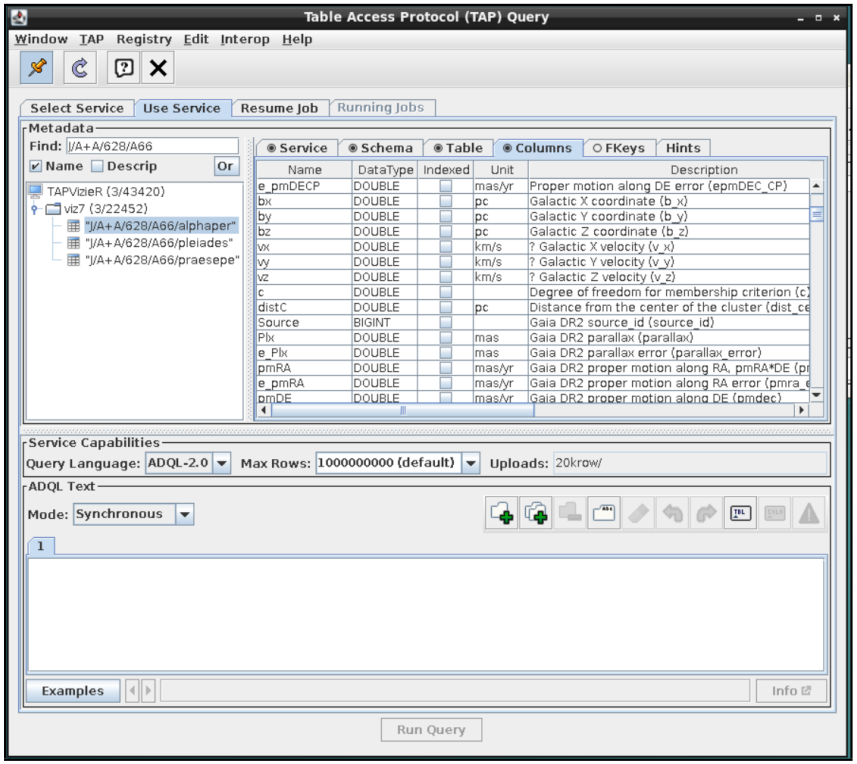


Figure: Available services and a selected entry from VizieR.

Topcat TAP



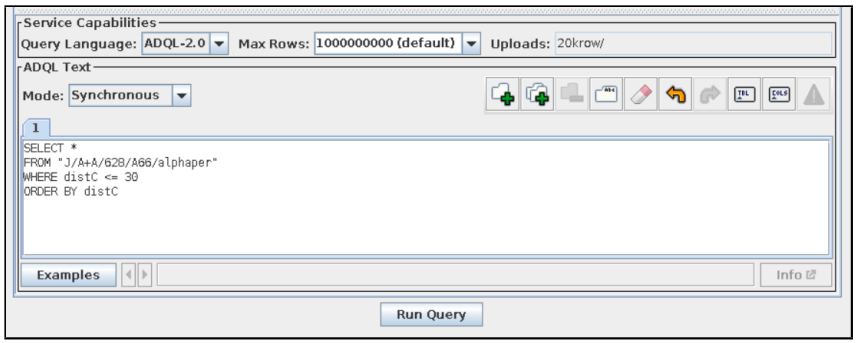
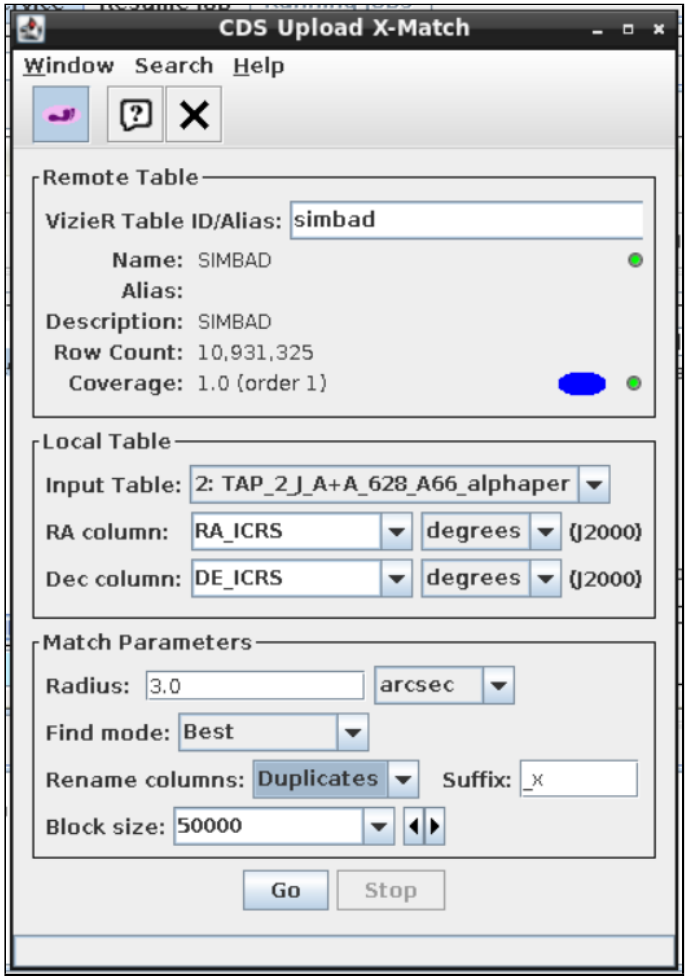


Figure: Select data from the table by selecting columns or by an SQL (ADQL) query.

Cross-matching tables with Topcat

- Topcat is able to cross-match tables
- You can crossmatch online tables or your local table with online tables (for example with Simbad)
- Usually, cross-matching is done through coordinates of the objects

Cross-matching tables with Topcat



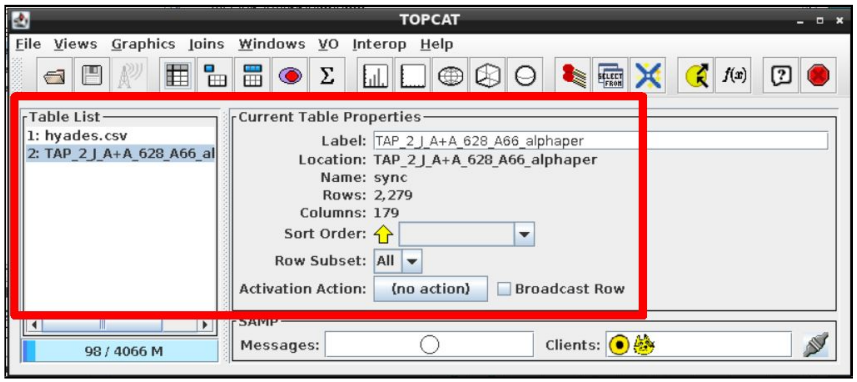


Figure: Cross-matching a local table through coordinates with the Simbad catalog.

Python libraries

Accessing databases and manipulating tables with Python

- Topcat is a highly efficient and very fast tool for accessing databases, manipulating tables and to create plots
- However, you might want to do all these tasks from your own algorithm
- Fortunately, Python has a library for everything ...
- Useful libraries to access and manipulate databases are:
  - *astroquery*
  - *pystilts*
  - *pandas*

astroquery

- *astroquery* is a set of tools for querying astronomical web forms and databases
- The list of available databases among many others includes"
  - ALMA Queries (astroquery.alma)
  - Gaia TAP (astroquery.gaia)
  - IRSA Dust Extinction Service Queries (astroquery.irsadust)
  - JPL Spectroscopy Queries (astroquery.jplspec)
  - Minor Planet Center Queries (astroquery.mpcastroquery.solarsystem.MPC)
  - NASA ADS Queries (astroquery.nasaads)
  - SIMBAD Queries (astroquery.simbad)
  - VizieR Queries (astroquery.vizier)
  - And many more ...

astroquery TAP query

```
>>> from astropy import coordinates
>>> import astropy.units as u
>>> # works only for ICRS coordinates:
>>> c = coordinates.SkyCoord("05h35m17.3s -05d23m28s", frame='icrs')
>>> r = 5 * u.arcminute
>>> result_table = Simbad.query_region(c, radius=r)
>>> result_table.pprint(show_unit=True, max_width=80, max_lines=5)
  MAIN_ID          RA           DEC       ... COO_WAVELENGTH     COO_BIBCODE
                "h:m:s"      "d:m:s"      ...
----------- ------------ ------------ ... -------------- -------------------
       M  42   05 35 17.3    -05 23 28 ...                1981MNRAS.194..693L
         ...          ...          ... ...            ...                 ...
V* V2114 Ori 05 35 01.671 -05 26 36.30 ...              I 2003yCat.2246....0C
```

Figure: Example of an *astroquery* coordinate query from the Simbad catalog.

pystilts

- STILTS is a set of command-line tools for processing tabular data. It has been designed for, but is not restricted to, use on astronomical data such as source catalogues.
- *pystilts* is the Python analogue of the STILTS
- Library *pystilts* allows you to execute TAP queries and cross-matching on online databases, similarly to Topcat

pystilts cross-matching

```
from pystilts import tapskymatch
from pystilts import keepcols, renamecols

filename = 'table1.fits'
outfile = 'table2.fits'
name = 'ppmxl'
tapurl = "http://dc.zah.uni-heidelberg.de/tap"
taptable = "ppmxl.main"
taplong = 'raj2000'
taplat = 'dej2000'
crossmatch = 3
nra1, ndec1 = 'ra', 'dec'
nra2, ndec2 = 'ra_table1', 'dec_table1'
# Run sky match via TAP
tapskymatch(tapurl=tapurl, taptable=taptable,
            taplon=taplong, taplat=taplat,
            infile=filename, inlon=nra1, inlat=ndec1,
            radius=crossmatch, outfile=outfile,
            icmd=keepcols([nra1, ndec1]), fixcols='all', suffixin='',
            suffixremote='_{0}'.format(name))
# Rename table1.fits ra and dec columns
renamecols(outfile, [nra1, ndec1], [nra2, ndec2])
```

Figure: Example of a *pystilts* cross-matching.

pandas

- *pandas* is a powerful Python data analysis toolkit
- It has immense features for manipulating tables, creating plots and histograms and analyzing data
- Once you downloaded your data with the *astroquery*or *pystilts*, *pandas* offers an efficient way of data manipulation

pandas - some important features

- *pandas* is in many ways similar to MS Excel as it offers some similar features:
  - Labeling and indexing data
  - Sorting and filtering data columns
  - Cross-matching tables (the equivalent of the Excel vlookup function)
  - Creating pivot tables
  - Grouping the data
  - Performing mathematical operations on the data
  - Visualizing data
  - And many more ...

A quick look to *pandas*

```
1  import pandas as pd
2  df = pd.DataFrame({'Num': [1, 2, 3, 4], 'Alp': ['a', 'b', 'c', 'd'],
3  'Dat': pd.date_range('2018-08-01', freq='d', periods=4),
4  'Cat': pd.Categorical(['a', 'b', 'c', 'c'])
5  }, index=pd.Index(range(4), name='ids'))
6  df
```

	Num	Alp	Dat	Cat
ids				
0	1	a	2018-08-01	a
1	2	b	2018-08-02	b
2	3	c	2018-08-03	c
3	4	d	2018-08-04	c

```
df.dtypes
```

```
Num              int64
Alp             object
Dat     datetime64[ns]
Cat           category
dtype: object
```

```
8  df.to_json('test.json', orient='table')
9  new_df = pd.read_json('test.json', orient='table')
10 new_df
```

Farewell

Summary

- The aim of this presentation was to introduce online accessible astronomy catalog
- The application Topcat is an effective and fast tool to access these catalogs and to process the data
- On the other hand, the processing can be done through Python

Thank you for your attention